The Philosophy of Science

An Encyclopedia

Sahotra Sarkar
Jessica Pfeifer
EDITORS

The Philosophy of Science

An Encyclopedia

EDITORIAL ADVISORY BOARD

Justin Garson
University of Texas at Austin

Paul Griffiths
University of Queensland

Cory Juhl
University of Texas at Austin

James Justus
University of Texas at Austin

Phillip Kitcher
Columbia University

Ian Nyberg
University of Texas at Austin

Anya Plutyinski
University of Utah

Sherrilyn Roush
Rice University

Laura Ruetsche
University of Pittsburgh

John Stachel
Boston University

William Wimsatt
University of Chicago

The Philosophy of Science

An Encyclopedia

Sahotra Sarkar
Jessica Pfeifer
EDITORS

EDITORIAL ASSISTANTS
Justin Garson, James Justus, and Ian Nyberg
University of Texas

Routledge
Taylor & Francis Group
New York London

Published in 2006 by
Routledge
Taylor & Francis Group
270 Madison Avenue
New York, NY 10016

Published in Great Britain by
Routledge
Taylor & Francis Group
2 Park Square
Milton Park, Abingdon
Oxon OX14 4RN

© 2006 by Taylor & Francis Group, LLC
Routledge is an imprint of Taylor & Francis Group

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 0-415-93927-5 (Hardcover)
International Standard Book Number-13: 978-0-415-93927-0 (Hardcover)
Library of Congress Card Number 2005044344

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

The philosophy of science : an encyclopedia / Sahotra Sarkar, Jessica Pfeifer, editors.
p. cm.
Includes bibliographical references and index.
ISBN 0-415-93927-5 (set : alk. paper)--ISBN 0-415-97709-6 (v. 1 : alk. paper) -- ISBN 0-415-97710-X (v. 2 : alk. paper)
1. Science--Philosophy--Encyclopedias. I. Sarkar, Sahotra. II. Pfeifer, Jessica.

Q174.7.P55 2005
501'.03--dc22 2005044344

T&F informa
Taylor & Francis Group is the Academic Division of T&F Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the Routledge Web site at
http://www.routledge-ny.com

Dedicated to the memory of Bob Nozick, who initiated this project.

TABLE OF CONTENTS

Introduction xi

List of Contributors xxvii

A to Z List of Entries xxxiii

Thematic List of Entries xxxvii

Entries A–M 1

THE PHILOSOPHY OF SCIENCE: AN INTRODUCTION

Philosophy of science emerged as a recognizable sub-discipline within philosophy only in the twentieth century. The possibility of such a sub-discipline is a result of the post-Enlightenment disciplinary and institutional separation of philosophy from the sciences. Before that separation, philosophical reflection formed part of scientific research—as, indeed, it must—and philosophy was usually guided by a sound knowledge of science, a practice that gradually lost currency after the separation. In the nineteenth century, philosophical reflection on science resulted in a tradition of natural philosophy, particularly in Britain (with the work of Mill, Pearson, Whewell, and others), but also in continental Europe, especially in Austria (with Bolzano, Mach, and others). What is called philosophy of science today has its roots in both the British and the Austrian traditions, although with many other influences, as several entries in this Encyclopedia record (see, for instance, Duhem Thesis; Poincaré, Henri).

This Encyclopedia is intended to cover contemporary philosophy of science. It is restricted to conceptual developments since the turn of the twentieth century. Its treatment of major figures in the field is restricted to philosophers (excluding scientists, no matter what the extent of their philosophical influence has been) and, with very few exceptions (notably Chomsky, Noam; Putnam, Hilary; and Searle, John), to those whose work is distant enough to allow "historical" appraisal. Conceptual issues in the general philosophy of science (including its epistemology and metaphysics) as well as in the special sciences are included; those in mathematics have been left for a different work. This Introduction will provide a guided tour of these conceptual issues; individual figures will only be mentioned in passing.

Historically, the themes treated in the Encyclopedia are those that have emerged starting with the period of the Vienna Circle (see Vienna Circle), including the figures and developments that influenced it (see Bridgman, Percy Williams; Duhem Thesis; Mach, Ernest; Poincaré, Jules Henri). The work of the members of the Vienna Circle provide a link between the older natural philosophy, especially in its Austrian version, and the later philosophy of science, which borrowed heavily from the concepts and techniques of the mathematical logic that was being created in the first three decades of the last century (see Hilbert, David; Ramsey, Frank Plumpton; Russell, Bertrand; see also Ayer [1959] and Sarkar [1996a]). The new set of doctrines—or, more accurately, methods—came to be called "logical positivism" and, later, "logical empiricism" (see Logical Empiricism; see also Sarkar [1996b]). By the 1930s these views had spread beyond the confines of Vienna and had attracted allegiance from many other similarly-minded philosophers (see Ayer, A. J.; Quine, Willard Van; Reichenbach, Hans). Two attitudes were widely shared within this group: a belief that good philosophy must be conversant with the newest developments within the sciences (see Rational Reconstruction), and a rejection of traditional metaphysics imbued with discussions with no empirical significance (see Cognitive Significance; Verifiability).

Some members of the Vienna Circle also took the so-called linguistic turn (see Carnap, Rudolf) and viewed scientific theories as systems formalized in artificial languages (Sarkar 1996c). Arguably, at least, this work lost the prized contact with the practice of science, and this development contributed to the eventual rejection of logical empiricism by most philosophers of science in the late twentieth century. However, a number of the original logical empiricists, along with many others, rejected the linguistic turn, or at least did not fully endorse it (see Neurath, Otto; Popper, Karl Raimund; Reichenbach, Hans). The tensions between the two views were never fully articulated during this period, let alone resolved, because the Vienna Circle as an institution and logical empiricism as a movement both came under political attack in Europe with the advent of Nazism. Most of the figures involved in the movement migrated to the United Kingdom and the United States. In the United States, many of the logical empiricists also later fell afoul of McCarthyism (see Logical Empiricism).

In the United States, Nagel probably best exemplifies what philosophy of science became in the period of the dominance of logical empiricism. The discussions of Nagel's (1961) *Structure of Science* typically include careful formal accounts of conceptual issues, but these are supplemented by detailed "nonformal" discussions in the spirit of the tradition of natural philosophy—this book may be viewed as a summary of where logical empiricism stood at its peak (see Nagel, Ernest). However, starting in the late 1940s, many of the theses adopted by the logical empiricists came under increasing attack even by those committed to keeping philosophy in contact with the sciences (Sarkar 1996e). (The logical empiricists had explicitly advocated and practiced intense self-criticism, and many of these attacks came from within their ranks—see Hempel, Carl Gustav.) Some of this criticism concerned whether cherished doctrines could be successfully formulated with the degree of rigor desired by the logical empiricists (see Analyticity; Cognitive Significance).

However, the most serious criticism came from those who held that the logical empiricists had failed to give an account of scientific confirmation and scientific change (see "Confirmation," "Scientific Discovery," and "Scientific Change," below). Feyerabend, for one, argued that the logical empiricists had placed science under an inadmissible rational straitjacket (see Feyerabend, Paul). As philosophy of science took a distinctly historical turn, analyzing the development of science in increasing historical detail, many felt that the logical empiricists had misinterpreted the historical processes of scientific change (see Hanson, Norwood Russell; Kuhn, Thomas). Kuhn's (1962) *Structure of Scientific Revolutions*, originally written for an encyclopedia sponsored by the logical empiricists, was particularly influential. By the mid-1960s logical empiricism was no longer the dominant view in the philosophy of science; rather, it came to be regarded as a "received view" against which philosophers of science defined themselves (Suppe 1974). However, this interpretation of logical empiricism ignores the disputes and diversity of viewpoints within the tradition (see, especially, Logical Empiricism), arguably resulting in a caricature rather than a responsible intellectual characterization. Nevertheless, for expository ease, the term "received view" will be used in this Introduction to indicate what may, at least loosely, be taken to be the majority view among the logical empiricists.

Scientific realism and various forms of naturalism, sometimes under the rubric of "evolutionary epistemology," have emerged as alternatives to the logical empiricist interpretations of science (see Evolutionary Epistemology; Scientific Realism). Meanwhile, science has also been subject to feminist and other social critiques (see Feminist Philosophy of Science). Kuhn's work has also been used as an inspiration for interpretations of science that regard it as having no more epistemological authority than "knowledge" generated by other cultural practices (see Social Constructionism). However, whether such work belongs to the philosophy of science, rather than its sociology, remains controversial. While no single dominant interpretation of science has emerged since the decline of logical empiricism, the ensuing decades have seen many innovative analyses of conceptual issues that were central to logical empiricism. There has also been considerable progress in the philosophical analyses of the individual sciences. The rest of this Introduction will briefly mention these with pointers to the relevant entries in this work.

Theories

The analysis of scientific theories—both their form and content—has been a central theme within the philosophy of science. According to what has become known as "the received view," which was developed in various versions by the logical empiricists between the 1920s and 1950s, theories are a conjunction of axioms (the laws of nature) and correspondence rules specified in a formalized ideal language. The ideal language was supposed to consist of three parts: logical terms, observational terms, and theoretical terms. Logical claims were treated as analytic truths (see Analyticity), and were thought by many to be accepted as a matter of convention (see Conventionalism). Observational claims were also thought to be unproblematic, initially understood as referring to incorrigible sense-data and later to publicly available physical objects (see Phenomenalism; Physicalism; Protocol Sentences). The correspondence rules were supposed to allow the logical empiricists to give cognitive significance (see Cognitive Significance; Verifiability) to the theoretical portion of the language, by specifying rules for connecting theoretical and observational claims. In their extreme version, these correspondence rules took the form of operational definitions (see Bridgeman, Percy Williams). One goal of such attempts was to distinguish science from non-science, especially what the logical empiricists derided as "metaphysics" (see Demarcation, Problem of).

Starting in the 1960s, the received view encountered a number of problems. Even earlier, difficulties had arisen for the correspondence rules, which took various forms over the years as a result of these problems. Initially understood as explicit definitions, they were later treated as partial definitions, and in the end the theoretical terms were merely required to make a difference to the observational consequences of the theory. One central focus of the criticism was on the observation-theory distinction (see Observation). It was argued that the theoretical and observational portions of language are not distinct (Putnam 1962; Achinstein 1968; see also Putnam, Hilary), that the distinction between entities that are observable and those that are not is vague (Maxwell 1962), and that observations are theory-laden (Hanson 1958; see also Hanson, Norwood Russell; Observation). In addition, there were problems ruling out unintended models of theories, which became a source of counterexamples. In hindsight, it is also clear that the problem of demarcating science from non-science was never fully solved.

More recently, a number of philosophers have questioned the important place given to laws of nature on this view, arguing that there are scientific theories in which laws do not appear to play a significant role (see Biology, Philosophy of; Laws of Nature). Others have questioned not the occurrence of laws within theories, but whether any of these entities should be conceptualized as linguistic entities (which is quite foreign to the practice of science). Still others have wondered whether the focus on theories has been an artifact of the received view being based primarily on physics, to the detriment of other sciences. As the received view fell out of favor, starting in the 1960s, a number of philosophers developed various versions of what is known as the semantic view of theories, which understands theories as classes of models, rather than as linguistic entities specifiable in an axiomatic system. While not without its problems, the semantic view seemed to bring philosophical accounts of theories more in line with the practices of scientists and has become the generally accepted view of theories (see Scientific Models; Theories). Nevertheless, there is at present no consensus within the discipline as to how theories should be philosophically characterized.

Scientific Models

Models are central to the practice of science and come in a bewildering variety of forms, from the double helix model of DNA to mathematical models of economic change (see Scientific Models). Scientific models were regarded as being of peripheral philosophical interest by the received view. Little philosophical work was done on them until the 1970s, with Hesse's (1963) *Models and Analogies in Science* being a notable exception. That situation has changed drastically, with models probably now being the locus of even more philosophical attention than theories.

Two developments have contributed to the burgeoning philosophical interest in models:

(i) *The Semantic Interpretation of Theories*. The development of various versions of the semantic interpretation of theories has put models at the center of theoretical work in science (see Theories). For many proponents of the semantic view, the received view provided a syntactic interpretation of theories, regarding theories as formalized structures. Scientific models are then supposed to be construed in analogy with models in formal logic, providing semantic interpretations of syntactic structures. The semantic view inverts this scheme to claim that models are epistemologically privileged and that theories should be regarded as classes of models. The various semantic views have made many contributions to the understanding of science, bringing philosophical analysis closer to the practice of science than the received view. Nevertheless, almost all versions of the semantic view are at least partly based on a dubious assumption of similarity between models in logic and what are called "models" in science.

(ii) *Historical Case Studies*. How dubious that presumed similarity has been underscored by the second development that helped generate the current focus on scientific models: the detailed studies of the role of models in science that has been part of the historical turn in the philosophy of science since the 1960s. That turn necessitated a focus on models because much of scientific research consists of the construction and manipulation of models (Wimsatt 1987). These studies have revealed that there are many different types of models and they have a variety of dissimilar functions (see Scientific Models for a taxonomy). At one end are models of data and representational material models such as the double helix. At the other are highly idealized models (see Approximation), including many of the mathematical models in the different sciences. Some models, such as the Bohr model of the atom (see Quantum Mechanics) or the Pauling models of chemical bonds (see Chemistry, Philosophy of), are both mathematical and accompanied by a visual picture that help their understanding and use (see also Visual Representation).

At present, no unified treatment of the various types and functions of scientific models seems possible. At the very least, the rich tapestry of models in science cannot entirely be accommodated to the role assigned to them by the semantic interpretation of theories or any other account that views models as having only explanatory and predictive functions. The ways in which models also function as tools of exploration and discovery remain a topic of active philosophical interest (Wimsatt 1987).

Realism

A central concern of philosophers of science has long been whether scientists have good reason to believe that the entities (in particular the unobservable entities) referred to by their theories exist and that what their theories say about these entities is true or approximately true (see Realism). In order for theories to refer to or be true about unobservable entities, they must actually be claims about these entities. This was denied by many logical empiricists, building on concerns raised by Mach, Duhem, and Poincaré (see Mach, Ernest; Poincaré, Henri). As noted above, the logical empiricists were interested in providing cognitive significance to theoretical terms by attempting to reduce theoretical claims to claims in the observation language. Even when this proved impossible, many nevertheless argued that theoretical terms are simply convenient instruments for making predictions about observable entities, rather than claims about unobservable entities (see Instrumentalism).

Because of the difficulties with theory-observation distinction discussed above (see Observation; Theories), this view fell out of favor and was replaced with a milder version of anti-realism. Van Fraassen (1980), for example, argues that while claims about unobservables might have a truth–value, scientists only have good reason to believe in their empirical adequacy, not their truth. Such a view might broadly be understood as instrumentalist in the sense that the truth of theories does not underwrite the functions they serve. There are two main arguments provided in support this version of anti-realism. First, given the problem of underdetermination raised by Duhem and Quine, there will always be more than one rival hypothesis compatible with any body of evidence (see Duhem Thesis; Underdetermination of Theories). Therefore, since these hypotheses are incompatible, the evidence cannot provide adequate reason to believe that one or the other theory is true. Second, some have argued that history provides evidence against believing in the truth of scientific theories. Given the large number of theories once thought true in the past that have since been rejected as false, history provides inductive evidence that science's current theories are likely to be false as well (see Laudan 1981).

There have been a number of responses to these arguments, including attempts to show that the problem of underdetermination can be solved, that anti-realism depends on a distinction between observable and unobservable entities that cannot be sustained, and that the realist need only claim that theories are approximately true or are getting closer to the truth (see Verisimilitude). In addition, arguments have been provided in support of realism about theories, the most influential of which is Putnam's miracle argument (see Putnam, Hilary). There are various versions of this argument, but the central premise is that science is successful (what this success amounts to varies). The contention is that the only way this success can be explained is if scientific theories are approximately true (see Abduction); otherwise the success of science would be a miracle.

This argument has been criticized in three central ways. First, Fine (1986) criticizes the miracle argument for being viciously circular. Second, some have argued that science is in fact not very successful, for reasons outlined above. Third, it is argued that the success of science does not depend on its truth, or perhaps does not even require an explanation. Van Fraassen (1980), for example, has argued that it is not surprising that scientific theories are predictively successful, since they are chosen for their predictive success. Therefore, the success of theories can be explained without supposing their truth. Others have responded that this would not, however, explain the predictive success of theories in novel situations (e.g., Leplin 1997).

Due to these problems, other forms of realism have been defended. Hacking (1983), for example, defends entity realism. He argues that, while scientists do not have good reason to believe their theories are true, they do have good reason to believe that the entities referred to in the theories exist, since scientists are able to manipulate the entities. Others have attempted to defend a more radical form of anti-realism, according to which the entities scientists talk about and the theories they invent to discuss them are merely social constructs (see Social Constructionism).

Explanation

In an attempt to avoid metaphysically and epistemically suspect notions such as causation (see Causality), Hempel and Oppenheim (1948) developed a covering law model of explanation: the deductive-nomological (D-N) account (see Explanation; Hempel, Carl). Rather than relying on causes, they argued that scientific explanations cite the law or laws that cover the phenomena to be explained. According to the D-N model, explanations are deductive arguments, where the conclusion is a statement expressing what is to be explained (the *explanandum*), and the premises (the *explanans*) include at least one law-statement. Often statements about particular antecedent conditions from which the *explanandum* can be derived. Initially developed only to cover explanations of particular facts, the D-N model was expanded to include explanations of laws, such as the explanation of Kepler's laws by deriving them from Newton's laws of motion (along with particular facts about the planets). To account for explanations of particular events and laws governed by *statistical* laws, the inductive-statistical (I-S) and deductive-statistical (D-S) models were developed (Hempel 1965). According to the D-S model, statistical laws are explained by deductively deriving them from other statistical laws. However, statements describing particular facts cannot be deduced from statistical laws. Instead, according to the I-S model, the explanans containing statistical laws must confer a high inductive probability to the particular event to be explained. In this way, the covering law model of explanation was able to link explanation with predictability (see Prediction) and also make clear why the reduction of, say, Kepler's laws to Newton's laws of motion could be explanatory (see Reductionism).

In the ensuing years, these accounts ran into a number of problems. The covering law model seemed unable to account for cases where scientists and non-scientists appear to be giving perfectly good explanations without citing laws (see Biology, Philosophy of; Function; Mechanism; Social Sciences, Philosophy of). Several counterexamples were developed against the D-N model, including the purported explanation of events by citing irrelevant factors, such as the explanation of Joe's failure to get pregnant by citing the fact that he took birth-control pills, and the explanation of causes by citing their effects, such as the explanation of the height of a flagpole by citing the length of its shadow. Deductive relations, unlike explanatory relations, can include irrelevant factors and need not respect temporal asymmetries. The I-S model also encountered difficulties. According to the I-S model, improbable events cannot be explained, which runs counter to many philosophers' intuitions about such cases as the explanation of paresis by citing the fact that a person had untreated syphilis. Moreover, developing an account of inductive probability proved difficult (see Inductive Logic; Probability). Attempts to provide an adequate account of laws within an empiricist framework also encountered problems. According to Hempel and Oppenheim, laws are expressed by universal generalizations of unlimited scope, with purely qualitative predicates, and they do not refer to particular entities. The problem is that there are accidental generalizations, such as 'All pieces of gold have a mass of less than 10,000 kg,' that satisfy these conditions. Laws appear to involve the modal features that Hume and the logical empiricists were intent on avoiding; unlike accidental generalization, laws seem to involve some sort of natural necessity. The difficulty is to develop an account of laws that makes sense of this necessity in a way that does not make knowledge of laws problematic (see Laws of Nature).

In response to these problems, some have attempted to rescue the covering-law model by supplementing it with additional conditions, as in unificationist accounts of explanation. According to these accounts, whether an argument is explanatory depends not just on the argument itself, but on how it fits into a unified theory (see Unity and Disunity of Science). Scientists explain by reducing the number of brute facts (Friedman 1974) or argument patterns (Kitcher 1989) needed to derive the largest number of consequences. Others have developed alternatives to the covering law model. Van Fraassen (1980) has defended a pragmatic account of explanation, according to which what counts as a good explanation depends on context. Others have developed various causal accounts of explanation. Salmon (1971) and others have argued that explanatory and causal relations can be understood in terms of statistical relevance; scientists explain by showing that the *explanans* (a causal factor) is statistically relevant for the event to be explained. Salmon (1984) eventually rejected this view in favor of a causal mechanical model, according to which explanations appeal to the mechanisms of causal propagation and causal interactions (see Mechanism). Along with the development of various causal accounts of explanation have come numerous accounts of causation, as well as attempts to develop a better epistemology for causal claims through, for example, causal modeling (see Causality).

Prediction

Traditionally, prediction has been regarded as being as central to science as explanation (see Prediction). At the formal level, the received view does not distinguish between explanation and prediction. For instance, in the D-N model, the conclusion derived from the laws and other assumptions can be regarded as predictions in the same way that they can be regarded as explanations. While prediction is generally taken to refer to the future—one predicts future events—philosophically, the category includes retrodiction, or prediction of past events, for instance the past positions of planets from Newton's laws and their present positions and momenta. (On some accounts of hypothesis confirmation, retrodiction is even more important than forward prediction—see Bayesianism.)

The D-N model assumes that the laws in question are deterministic (see Determinism). Statistical explanations are also predictive, but the predictions are weaker: they hold probabilistically and can only be confirmed by observing an ensemble of events rather than individual events (see Confirmation Theory). Interest in statistical explanation and prediction initially arose in the social sciences in the nineteenth century (Stigler 1986; see also Social Sciences, Philosophy of the). In this case, as well as in the case of prediction in classical statistical physics, the inability to predict with certainty arises because of ignorance of the details of the system and computational limitations. A different type of limitation of prediction is seen when predictions must be made about finite samples drawn from an ensemble, for instance, biological populations (see Evolution; Population Genetics). Finally, if the laws are themselves indeterministic, as in the case of quantum mechanics, prediction can only be statistical (see Quantum Mechanics). The last case has generated the most philosophical interest because, until the advent of quantum mechanics, the failure to predict exactly was taken to reflect epistemological limitations rather than an ontological feature of the world. That the models of statistical explanation discussed earlier do not distinguish between these various cases suggests that there remains much philosophical work to be done. Meanwhile, the failure of determinism in quantum mechanics has led to much re-examination of the concept of causality in attempts to retain the causal nature of physical laws even in a probabilistic context (see Causality).

Prediction, although not determinism, has also been recently challenged by the discovery that there exist many systems that display sensitivity to initial conditions, the so-called chaotic systems. Determinism has usually been interpreted as an ontological thesis: for deterministic systems, if two systems are identical at one instant of time, they remain so at every other instant (Earman 1986; see Determinism). However, satisfying this criterion does not ensure that the available—and, in some cases, all obtainable—knowledge of the system allows prediction of the future. Some physical theories may prevent the collection of the required information for prediction (Geroch 1977; see also Space-Time). Even if the information can be collected, pragmatic limitations become relevant. The precision of any information is typically limited by measurement methods (including the instruments). If the dynamical behavior of systems is exceedingly sensitive to the initial conditions, small uncertainties in the initial data may lead to large changes in predicted behavior—chaotic systems exemplify this problem (see Prediction).

Confirmation

Hume's problem—how experience generates rational confidence in a theory—has been central to philosophy of science in the twentieth century and continues to be an important motivation for contemporary research (see Induction, Problem of). Many of the logical empiricists initially doubted that there is a logical canon of confirmation. Breaking with earlier logical traditions, for many of which inductive logic was of central importance, these logical empiricists largely regarded confirmation as a pragmatic issue not subject to useful theoretical analyses. That assessment changed in the 1940s with the work of Carnap, Hempel, and Reichenbach, besides Popper (see Carnap, Rudolf; Hempel, Carl Gustav; Popper, Karl Raimund; Reichenbach, Hans). Carnap, in particular, began an ambitious project of the construction of a logic of confirmation, which he took to be part of semantics, in the process reviving Keynes' logical interpretation of probability. Early versions of this project were distant from the practice of science, being restricted to formal languages of excessively simplified structures incapable of expressing most scientific claims. Later versions came closer to scientific practice, but only to a limited extent (see Carnap, Rudolf). Whether or not the project has any hope remains controversial among philosophers. Although the relevant entries in this Encyclopedia record some progress, there is as yet no quantitative philosophical theory of confirmation (see Confirmation Theory; Inductive Logic; Probability).

Meanwhile, within the sciences, the problem of confirmation was studied as that of statistical inference, bringing standard statistical methods to bear on the problem of deciding how well a hypothesis is supported by the data. Most of these methods were only invented during the first half of the twentieth century. There are two approaches to statistics, so-called orthodox statistics (sometimes called "frequentist" statistics) and Bayesian statistics (which interprets some probabilities as degrees of belief). The former includes two approaches to inference, one involving confidence intervals and largely due to Neyman and E. S. Pearson and the other due to Fisher. These have received some attention from philosophers but, perhaps, not as much as they deserve (Hacking 1965; see Statistics, Philosophy of). In sharp contrast, Bayesian inference has been at the center of philosophical attention since the middle of the twentieth century. Interesting work points to common ground between traditional confirmation theory and Bayesian methodology. Meanwhile, within the sciences, newer computational methods have made Bayesian statistics increasingly popular (see Statistics, Philosophy of), for instance, in the computation of phylogenies in evolutionary biology (see Evolution). Bayesian inference methods also have the advantage of merging seamlessly with contemporary decision theory (see Decision Theory), even though most of the methods within decision theory were invented in an orthodox context.

Philosophically, the differences between orthodox and Bayesian methods remain sharply defined. Orthodox methods do not permit the assignment of a probability to a hypothesis, which, from the perspective of most Bayesians, makes them epistemologically impotent. (Bayesians also usually argue that orthodox inferential recipes are *ad hoc*—see Bayesianism.) Meanwhile Bayesian methods require an assignment of *prior* probabilities to hypotheses before the collection of data; for the orthodox such assignments are arbitrary. However, in the special sciences, the trend seems to be one of eclecticism, when orthodox and Bayesian methods are both used with little concern for whether consistency is lost in the process. This situation calls for much more philosophical analysis.

Experimentation

The logical empiricists' focus on the formal relations between theory and evidence resulted in Anglo-American philosophers neglecting the role of experimentation in science. Experimentation did receive some philosophical treatment in the late nineteenth and early twentieth centuries, in particular by Mill, Mach, and Bernard (see Mach, Ernest). In twentieth century Germany, two traditions developed around the work of Dingler and Habermas. It is only in the past three decades that experimentation has received more attention from Anglo-American philosophers, historians, and sociologists. Since then, there have been a number of careful analyses of the use of experiments by practicing scientists, with historians and sociologists focusing largely on the social and material context of experiments and philosophers focusing on their epistemic utility.

From a philosophical perspective, the neglect of experimentation was particularly problematic, since experimentation seems to affect the very evidential relations empiricists were interested in formalizing. Whether experimental results are good evidence for or against a hypothesis depends on how the results are produced—whether the data are reliably produced or a mere artifact of the experimental procedure. Moreover, this reliability often comes in degrees, thereby affecting the degree to which the data confirms or disconfirms a hypothesis. In addition, how data are produced affects what sorts of inferences can be drawn from the data and how these inferences might be drawn. As Mill argues, "Observations, in short, without experiment . . . can ascertain sequences and coexistences, but cannot prove causation" (1874, 386). How experimental results are obtained can also affect whether replication is necessary and how statistical methods are used. In some cases, statistics is used to analyze the data, while in others, it is involved in the very production of the data itself (see Experimentation; Statistics, Philosophy of).

One of the central issues in the philosophy of experimentation is what experiments are. Experiments are often distinguished from observations in that the former involve active intervention in the world, whereas the latter are thought to be passive. However, it is unclear what counts as an intervention. For example, are the use of sampling methods or microscopes interventions? There are also questions about whether thought experiments or computer simulations are "real" experiments or if they merely function as arguments. Moreover, it is not always clear how to individuate experiments—whether it is possible, especially with the increasing use of computers as integral parts of the experimental set-up, to disambiguate the experiment from the analysis of the data.

Another fundamental issue is whether and what epistemic roles experiments can play (Rheinberger 1997). They are purportedly used in the testing of theories, in garnering evidence for the existence of entities referred to by our theories (see Realism), in the creation (and thereby discovery) of new phenomena, in the articulation of theories, in the development of new theories, in allowing scientists to "observe" phenomena otherwise unobservable (see Observation), and in the development and refinement of technologies.

Whether experiments can reliably serve these epistemic functions has been called into question in a number of ways. First, sociologists and historians have argued that social factors affect or even determine whether an experiment "confirms" or "disconfirms" a theory (see Social Constructionism). It is also argued that experiments are theory-laden, since experiments require interpretation and these interpretations rely on theories (Duhem 1954). Whether this is a problem depends in part on what use is made of the experiment and what sorts of theories are needed—the theory being tested, theories of the phenomena being studied but not being tested, or theories about the experimental apparatus being used. As Hacking (1983) and Galison (1987) both argue, experiments and experimental traditions can have a life of their own independent of higher-level theories.

The theory-ladenness of experimentation also raises questions about whether experiments can be used to test hypotheses in any straightforward way no matter which level of theory is used, since predictions about experimental results rely on auxiliary hypotheses that might be called into question (see Duhem Thesis). Experiments are also purported to be "practice-laden," relying on tacit knowledge that cannot be fully articulated (Collins 1985; see also Polanyi 1958). According to Collins, this leads to problems with replication. The reliability of experiments is often judged by the ability of scientists to replicate their results. However, what counts as replication of the "same" experiment is often at issue in scientific disputes. Since, according to Collins, tacit knowledge (which cannot be made explicit) is involved in the replication of experiments and even in judgments about what constitutes the "same" experiment, adjudicating these disputes on rational grounds is problematic. Collins, in addition, questions whether there can be independent grounds for judging whether an experiment is reliable, which he calls "the experimenters' regress." Whether an experimental procedure is reliable depends on whether it consistently yields correct results, but what counts as a correct result depends on what experimental procedures are deemed reliable, and so on (Collins 1985; for a reply, see Franklin 1994). Experiments also typically involve manipulation of the world, often creating things that are not naturally occurring, which has led some to question whether experiments represent the world as it naturally is. At one extreme are those who argue that experimentation actually constructs entities and facts (Latour and Woolgar 1979; Pickering 1984; Rheinberger 1997; see also Social Constructionism). Others argue that experiments can produce artifacts, but that these can be reliably distinguished from valid results (Franklin 1986). A milder version of this worry is whether laboratory settings can accurately reproduce the complexities of the natural world, which is exemplified in debates between field and experimental biologists. The effect of interventions on experimental outcomes is even more problematic in quantum physics (see Quantum Measurement Problem).

Scientific Change

Scientific change occurs in many forms. There are changes in theory, technology, methodology, data, institutional and social structures, and so on. The focus in the philosophy of science has largely been on theory change and whether such changes are progressive (see Scientific Change; Scientific Progress). The primary concern has also been with how scientific theories are justified and/or become accepted in the scientific community, rather than how they are discovered or introduced into the community in the first place. Over the years, there have been various notions of progress correlated with the different goals scientific theories are purported to have: truth, systematization, explanation, empirical adequacy, problem solving capacity, and so on. (Notice that if the focus were on, say, technological or institutional changes, the goals attended to might be very different; for example, does the technology have greater practical utility or is the institutional change just?)

Traditionally, scientific change has been thought of as governed by rational procedures that incrementally help science achieve its goals. For the logical empiricists, the aim of scientific theories was to systematize knowledge in a way that yields true predictions in the observational language (see Theories). As such, science progresses through the collection of additional confirming data, through the elimination of error, and through unification, typically by reducing one theory to another of greater scope. To make sense of these sorts of changes, the logical empiricists developed accounts of reduction, explanation, and inductive logic or confirmation theory (see Confirmation Theory; Explanation; Inductive Logic; Reductionism; Unity and Disunity of Science). Others, such as Popper, offered a different account of theory change. Popper defended an eliminativist account much like Mill's, whereby science attempts to eliminate or falsify theories. Only those theories that pass severe tests ought to be provisionally accepted (see Corroboration). This was also one of the earliest versions of evolutionary epistemology (see Evolutionary Epistemology; Popper, Karl Raimund).

As discussed in the previous sections, these accounts ran into difficulties: Quine extended Duhem's concerns about falsification, criticized the analytic/synthetic distinction, and raised questions about the determinacy of translation (see Duhem Thesis; Quine, Willard Van; Underdetermination); Popper and Hanson argued that observations are theory-laden (see Hanson, Norwood Russell; Observation; Popper, Karl Raimund); there were problems with Carnap's inductive logic; and so on. Partly influenced by these difficulties and partly motivated by a concern that philosopher's theories about science actually fit the practices of science, Kuhn's *The Structure of Scientific Revolutions* (1962) challenged the way philosophers, historians, sociologists, and scientists thought about scientific change (see Kuhn, Thomas). He argued that scientific change is not in general cumulative and progressive, but develops through a series of distinct stages: immature science (when there is no generally accepted paradigm), normal science (when there is an agreed upon paradigm), and revolutionary science (when there is a shift between paradigms). Kuhn's notion of paradigms also expanded the focus of scientific change beyond theories, since paradigms consisted, not just of theories, but of any exemplary bit of science that guides research. While the development of normal science might in some sense be incremental, Kuhn argued that the choice between paradigms during a revolution involves something like a *Gestalt* shift. There are no independent methods and standards, since these are paradigm-laden; there is no independent data, since observations are paradigm-laden; and the paradigms may not even be commensurable (see Incommensurability). Consequently, paradigm shifts seemed to occur in an irrational manner.

The responses to Kuhn's influential work took two very different paths. On the one hand, strongly influenced by Kuhn, members of the Strong Programme argued that scientific change ought to be explained sociologically—that the same social causes explain both "good" and "bad" science. Others (e.g. Latour and Woolgar 1979) went further, arguing that scientists in some sense construct facts (see Social Constructionism). Focus on the social aspects of scientific research also led to developments in feminist philosophy of science, both in the close analysis of the gender and racial biases of particular sciences and in the development of more abstract feminist theories about science (see Feminist Philosophy of Science).

The other, a very different sort of response, involved a defense of the rationality and progress of science. There were attempts to show that competing scientific theories and paradigms are not incommensurable in the sense of being untranslatable. Davidson (1974) argues the very idea of a radically different, incommensurable paradigm does not make sense; others (e.g., Scheffler 1967) argued that sameness of reference is sufficient to ensure translatability, which was later buttressed by referential accounts of meaning (see Incommensurability). The rationality of scientific change was also defended on other grounds. Lakatos developed Popper's ideas in light of Kuhn into his methodology of scientific research programs (see Lakatos, Imre; Research Programmes); and Laudan (1977) argued that progress can be made sense of in terms of problem solving capacity. Another approach to showing that scientific change is progressive can be found in realism. Rather than arguing that each change involves a rational choice, defenses of realism can be seen as attempts to establish that science is approaching its goal of getting closer to the truth (see Realism). Of course, anti-realists might also argue that science is progressing, not toward truth, but toward greater empirical adequacy.

More recently, there have been attempts to develop formal methods of theory choice beyond confirmation theory and inductive logic (see Bayesianism;

Statistics, Philosophy of). There have also been attempts to model discovery computationally, which had been thought not to be rule governed or formalizable. Some of these try to model the way humans discover; others were developed in order to make discoveries (e.g., data mining), whether or not humans actually reason in this way. As a normative enterprise, such modeling can also be used as a defense of the rationality of scientific discovery and, therefore, scientific change (see Scientific Change).

Perhaps the longest-lasting influence in the philosophy of science of Kuhn's influential work has been to encourage philosophers to look more closely at the actual practices of the various sciences. This has resulted in a proliferation of philosophies of the special sciences.

Foundations of the Special Sciences

The logical empiricists believed in the unity of science (see Unity of Science Movement). However, the theme was interpreted in multiple ways. At one extreme were views according to which unification was to be achieved through hierarchical reduction (see Reductionism) of sociology to individual psychology (see Methodological Individualism), psychology to biology (see Psychology, Philosophy of), biology to physics and chemistry (see Biology, Philosophy of), and chemistry to physics (see, Chemistry, Philosophy of); for an influential defense of this view, see Oppenhiem and Putnam (1958). At the other extreme were those who believed that unification required no more than to be able to talk of the subjects of science in an interpersonal (that is, non-solipsistic) language—this was Carnap's (1963) final version of physicalism. Somewhere in between were stronger versions of physicalism, which, for most logical empiricists and almost all philosophers of science since them, provides some vision of the unity of science (see Physicalism).

Perhaps with the exception of the most extreme reductionist vision of the unity of science, all other views leave open the possibility of exploring the foundations and interpretations of the special sciences individually. During the first few decades of the twentieth century, most philosophical attention to the special sciences was limited to physics; subsequently, psychology, biology, and the social sciences have also been systematically explored by philosophers. In many of these sciences, most notably biology and cognitive science, philosophical analyses have played a demonstrable role in the further development of scientific work (see Biology, Philosophy of; Cognitive Science; Intentionality).

Physical Sciences

The first three decades of the twentieth century saw the replacement of classical physics by relativity theory and quantum mechanics, both of which abandoned cherished classical metaphysical principals (see Quantum Mechanics; Space-Time). It is therefore not surprising that many philosophers interested in "scientific philosophy" (see Logical Empiricism) did significant work in this field. In particular, Popper and Reichenbach made important contributions to the interpretation of quantum mechanics; Reichenbach and, to a lesser extent, Carnap also contributed to the philosophy of space-time (see Carnap, Rudolf; Popper, Karl Raimund; Reichenbach, Hans). In both quantum mechanics and relativity, philosophers have paid considerable attention to issues connected with causality and determinism, which became problematic as the classical world-view collapsed (see Causality; Determinism). Arguably, Reichenbach's work on space-time, especially his arguments for the conventionality of the metric, set the framework for work in the philosophy of space-time until the last few decades (see Conventionalism). Reichenbach also produced important work on the direction of time.

Several philosophers contributed to the clarification of the quantum measurement problem (see Quantum Measurement Problem), the concept of locality in quantum mechanics (see Locality), and the nature and role of quantum logic (see Putnam, Hilary; Quantum Logic). Meanwhile, many physicists, including Bohr, Einstein, Heisenberg, and Schrödinger, also produced seminal philosophical work on the foundations of physics (see also Bridgman, Percy Williams; Duhem Thesis). The only consensus that has emerged from all this work is that, whereas the foundations of relativity theory (both special and general) are relatively clear, even after eighty years, quantum mechanics continues to be poorly understood, especially at the macroscopic level (see Complementarity).

Perhaps because of the tradition of interest in quantum mechanics, philosophers of physics, starting mainly in the 1980s, also began to explore the conceptual structure of quantum field theory and particle physics (see Particle Physics; Quantum Field Theory). However, one unfortunate effect of the early focus on quantum mechanics and relativity is that other areas of physics that also deserve philosophical scrutiny did not receive adequate attention, as Shimony (1987) and others have emphasized. (See the list of questions in the entry, Physical Sciences, Philosophy of.) Only in recent years have philosophers begun to pay attention to questions such as reductionism and irreversibility in kinetic theory (see Irreversibility; Kinetic Theory) and condensed matter physics (see Batterman [2002] and Reductionism). One interesting result has been that the question of reductionism within physics is now believed to be far more contentious than what was traditionally thought (when it was assumed that biology, rather than the physics of relatively large objects, presented a challenge to the program of physical reductionism—see Emergence).

Finally, beyond physics, some philosophical attention is now being directed at chemistry (see Chemistry, Philosophy of) and, so far to a lesser extent, astronomy (see Astronomy, Philosophy of). As in the case of macroscopic physics, the question of the reduction of chemistry to physics has turned out to be unexpectedly complicated with approximations and heuristics playing roles that make orthodox philosophers uncomfortable (see Approximation). It is likely that the future will see even more work on these neglected fields and further broadening of philosophical interest in the physical sciences.

Biology

Professional philosophers paid very little attention to biology during the first few decades of the twentieth century, even though the advent of genetics (both population genetics and what came to be called classical genetics [see Genetics]) was transforming biology in ways as profound as what was happening in physics. Professional biologists—including Driesch, J. B. S. Haldane, J. S. Haldane, and Hogben—wrote philosophical works of some importance. However, the only philosopher who tried to interpret developments in biology during this period was Woodger (1929, 1937), better known among philosophers as the translator of Tarski's papers into English. Philosophers paid so little attention to biology that not only the evolutionary "synthesis" (see Evolution), but even the formulation of the double helix model for DNA (see Reduction), went unnoticed by philosophers of those generations (Sarkar 2005).

All that changed in the 1960s, when the philosophy of biology emerged as a recognizable entity within the philosophy of science. The first question that occupied philosophers was whether molecular biology was reducing classical biology (see Molecular Biology; Reductionism). Initial enthusiasm for reductionism gave place to a skeptical consensus as philosophers began to question both the standard theory-based account of reductionism (due to Nagel 1961; see Nagel, Ernest) and whether molecular biology had laws or theories at all (Sarkar 1998). In the 1970s and 1980s, attention shifted almost entirely to evolutionary theory (see Evolution), to the definitions of "fitness" (see Fitness) and "function" (see Function), the nature of individuals and species (see Individual; Species), the significance of adaptation and selection (see Adaptation and Adaptationism; Population Genetics), and, especially, the units and levels of selection. Philosophical work has contributed significantly to scientific discussions of problems connected to units of selection, although no consensus has been reached (see Altruism; Natural Selection). Besides evolution, there was some philosophical work in genetics (see Genetics; Heredity and Heritability).

As in the case of the philosophy of physics, the last two decades have seen a broadening of interest within the philosophy of biology. Some of the new work has been driven by the realization that molecular biology, which has become most of contemporary biology, is not simply the study of properties of matter at lower levels of organization, but has a conceptual framework of its own. This framework has largely been based on a concept of information that philosophers have found highly problematic (see Biological Information). Formulating an adequate concept of biological information—if there is one—remains a task to which philosophers may have much to contribute (see Molecular Biology).

There has also been some attention paid to biodiversity (see Conservation Biology), ecology (see Ecology), immunology (see Immunology), and developmental biology, especially in the molecular era (see Molecular Biology). Neurobiology has sometimes been approached from the perspective of the philosophy of biology, although philosophical work in that area typically has more continuity with psychology (see "Psychology" below and Neurobiology). Philosophers have also argued on both sides of attempts to use biology to establish naturalism in other philosophical areas, especially epistemology and ethics—this remains one of the most contested areas within the philosophy of biology (see Evolutionary Epistemology; Evolutionary Psychology). Some philosophers of science have also interpreted the philosophy of medicine as belonging within the conceptual terrain of the philosophy of biology (Schaffner 1993). Finally, work in the philosophy of biology has also led to challenges to many of the traditional epistemological and metaphysical assumptions about science, about the nature of explanations, laws, theories, and so on (see Biology, Philosophy of; Mechanism).

Psychology

Philosophy and psychology have an intimate historical connection, becoming distinct disciplines only in the late nineteenth and early twentieth centuries. Even since then, many of the topics covered by psychology have remained of interest to philosophers of mind and language, although the route taken to address these questions might be very different. However, while philosophers of science did address concerns about the human sciences more generally (see "Social Sciences" below), it is only in the last twenty years or so that philosophy of psychology has developed as a distinct area of philosophy of science.

The intimate connection between philosophy and psychology can be seen throughout the history of psychology and the cognitive sciences more broadly. In an attempt to make psychology scientific, Watson (1913), a philosopher, founded behaviorism, which dominated the field of psychology for the first half of the twentieth century (see Behaviorism). This view fit well with empiricist attempts to reduce theoretical claims to those in the observational language by providing operational definitions (see Hempel 1949; see also Bridgeman, Percy Williams; Theories; Verificationism). However, the combined weight of objections from philosophers, linguists, and psychologists led to the demise of behaviorism. These criticisms, along with developments in mathematical computation (see Artificial Intelligence; Turing, Alan) and the influential work of Chomsky (see Chomsky, Noam; Linguistics, Philosophy of), resulted in the cognitive revolution in psychology; it became generally agreed upon that psychological theories must make reference to internal representations (see Intentionality; Searle, John). These developments also led to the creation of the interdisciplinary field of cognitive science, which included psychology, linguistics, computer science, neuroscience, and philosophy (see Cognitive Science).

Philosophers of psychology have been broadly interested in foundational issues related to the cognitive sciences. Among the topics of concern are the content of representation, the structure of thought, psychological laws and theories, and consciousness, each of which is briefly discussed below:

(i) *The Content of Representations.* One central question is what fixes the content of representations—is content determined by internal features of the agent (e.g., conceptual role semantics), features of the external physical environment (e.g., causal and teleological theories), or features of the external social environment? There are also debates about whether the representations are propositional in form, whether they require language (see Linguistics, Philosophy of), whether some are innate (see Empiricism; Innate/acquired Distinction), and whether representations are local or distributed (see Connectionism).

(ii) *The Structure of Thought.* The nature of cognition has also been a topic of dispute. Some argue that human cognition takes the form of classical computation (see Artificial Intelligence; Cognitive Science); connectionists argue that it is more similar to parallel distributed processing (see Connectionism); and more recently other accounts have been proposed, such as dynamical and embodied approaches to cognition. Also at issue is whether the cognitive structures in the mind/brain are modular (see Evolutionary Psychology), whether cognition is rule-governed, and whether some of the rules are innate (see Chomsky, Noam; Innate/Acquired Distinction).

(iii) *Theories and Laws.* Questions have been raised about the nature of theories in the cognitive sciences (see Neurobiology), about whether there are psychological or psychophysical laws (see Laws of Nature), and about how the theories and laws in different areas of the cognitive sciences relate, such as whether psychology is reducible to neurobiology (see Neurobiology; Physicalism; Reductionism; Supervenience). In addition, there is disagreement about how to interpret theories in the cognitive sciences—whether to interpret them realistically, as an attempt to represent how the mind/brain actually works, or merely instrumentally, as a means of saving the phenomena or making predictions (see Instrumentalism; Realism). Moreover, the problems of reflexivity and the intentional circle discussed below, along with difficulties peculiar to the various areas of the cognitive sciences, raise questions about the testability of psychological theories (see Neurobiology; Psychology, Philosophy of).

(iv) *Consciousness*. There has been a resurgence of interest in consciousness (see Consciousness; Searle, John). There have been attempts to clarify what "consciousness" involves in its various senses, as well as debates about how to explain consciousness. To this end, a number of theories of consciousness have been proposed, including higher-order theories, neurological theories, representational theories, and various non-physical theories.

Social Sciences

Philosophical interest in the foundations of the social sciences has a long history, dating back at least to Mill's influential work on the social sciences. Some foundational issues have also been systematically discussed by social scientists themselves, such as Durkheim (1895/1966) and Weber (1903/1949). Around the middle of the twentieth century, the social sciences again received serious philosophical attention. The focus was largely on their being *human* sciences and the philosophical issues this raised. More recently, philosophers have directed their attention to the different social sciences in their own right, especially economics (see Economics, Philosophy of).

A central focus of discussion is whether the social sciences are fundamentally different from the natural sciences. Logical empiricists attempted to incorporate the social sciences into their models for the natural sciences (see Unity of Science Movement). Others have argued that the social sciences are unique. This has framed many of the debates within the philosophy of the social sciences, a number of which are briefly discussed in what follows (see Social Sciences, The Philosophy of):

(i) *Are There Social Science Laws?* Laws played important roles in empiricist accounts of explanation, theories, confirmation, and prediction, but it is unclear whether there are laws of the social sciences (see Laws of Nature). Social phenomena are complex, involve reference to social kinds, and require idealizations. As a result, many argue that generalizations of the social sciences, if they are laws at all, require ineliminable ceteris paribus clauses. Others argue that the social sciences ought not even attempt to create generalizations or grand theories, as social phenomena are essentially historical and local.

(ii) *Do Social Scientific Theories Yield Testable Predictions?* Because of the complexity of social systems, social scientific theories require idealizations. Given the nature of these idealizations, deriving empirical predictions from social scientific theories is difficult at best (see Prediction). As a result, many argue that social scientific theories are not testable. This is exacerbated by the reflexive nature of social science theories: the very act of theorizing can change the behavior one is theorizing about. Moreover, if human action is explained by agents' desires and beliefs, social scientists seem to be caught in an intentional circle, making it difficult to derive any testable claims (see Rosenberg 1988).

(iii) *Is the Methodology of the Social Sciences Distinct?* Given that social sciences involve humans and human behavior on a large scale, experimentation has not played a significant role in the social sciences (see Experimentation). There are also many who question whether the social sciences can be naturalized. Some argue that understanding social action is essentially a hermeneutic enterprise, distinctly different from the natural sciences.

(iv) *What Are the Ontological Commitments of Scientific Theories?* Beginning with Mill and, subsequently, Durkheim and Weber, there have been debates as to whether social scientific theories are reducible to theories about individual behavior (see Methodological Individualism). Moreover, after Nagel's influential account of intertheoretic reduction, it has been argued that social phenomena are multiply realizable, and therefore, social science theories are not reducible to lower-level theories (see Emergence; Reductionism; Supervenience). Additionally, given that social scientific theories involve idealizations, there are questions about whether these theories ought to be interpreted realistically or instrumentally (see Instrumentalism; Realism).

(v) *What Is the Nature of Social Scientific Explanations?* Some, such as Hempel (1962), have argued that social scientific explanations are no different than in the physical sciences. Others, however, have questioned this. If there are no social scientific laws, then social scientific explanation cannot be captured by the covering law model (see Explanation). Social sciences also often rely on functional explanations, which, while similar to biology, seem to be different from explanations in physics (see Function). Others, following Winch (1958), have argued that social sciences explain action, not behavior, which requires understanding the meaning of the action (not its causes), and therefore must include the actors' intentions and social norms. Moreover, some have argued that actions are governed by reasons, and are therefore not susceptible to causal explanation, a view that was later convincingly refuted by Davidson (1963). An alternative account of how beliefs and desires can explain actions has been formalized in rational choice theory (see Decision Theory), although there are questions about whether such explanations capture how people actually behave, rather than how they ought to behave.

(vi) *What Is the Relationship Between Social Science and Social Values?* There has also been concern with the connection between social values and the social sciences. Taylor (1971), for example, argues that social theory is inherently value-laden, and Habermas (1971) argues that social theory *ought* to engage in social criticism.

Concluding Remarks

Philosophy of science remains a vibrant sub-discipline within philosophy today. As this introduction has documented, many of the traditional questions in epistemology and metaphysics have been brought into sharper profile by a focus on scientific knowledge. Moreover, philosophical engagement with the special sciences has occasionally contributed to the development of those sciences and, as philosophers become more immersed in the practice of science, the number and level of such contributions can be expected to increase. The trend that philosophers of science engage all of the special sciences—not just physics—will also help produce a more complete picture of the growth of science, if not all knowledge, in the future.

With few exceptions (e.g., Demarcation, Problem of and Feminist Philosophy of Science) the entries in the Encyclopedia are not concerned with the social role of science. But, as science and technology continue to play dominant roles in shaping human and other life in the near future, philosophers may also contribute to understanding the role of science in society. Moreover, in some areas, such as the environmental sciences and evolutionary biology, science is increasingly under ill-motivated attacks in some societies, such as the United States. This situation puts philosophers of science, because of their professional expertise, under an obligation to explain science to society, and, where ethically and politically appropriate, to defend the scientific enterprise. How such defenses should be organized without invoking a suspect criterion of demarcation between science and non-science remains a task of critical social relevance. The Encyclopedia should encourage and help such efforts.

Jessica Pfeifer
Sahotra Sarkar

References

Achinstein, P. (1968), *Concepts of Science: A Philosophical Analysis*. Baltimore: Johns Hopkins University Press.

Ayer, A. J. 1959. Ed. Logical Positivism. New York: Free Press.

Batterman, R. W. 2002. *The Devil in the Details*. Oxford: Oxford University Press.

Carnap, R. 1963. "Replies and Systematic Expositions." In Schilpp, P. A. Ed. *The Philosophy of Rudolf Carnap*. La Salle: Open Court, pp. 859–1013.

Collins, Harry (1985), *Changing Order: Replication and Induction in Scientific Practice*. London: Sage.

Davidson, D. (1963), "Actions, Reasons, and Causes", *Journal of Philosophy*, 60, reprinted in Davidson (1980), *Essays on Actions and Events*, Oxford: Clarendon Press.

——— (1974), *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Duhem, Pierre (1954), *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.

Durkheim, E. (1895/1966), *The Rules of Sociological Method, 8th edition.* Solovay and Mueller (transl.) and Catlin (ed.), New York: Free Press.

Earman, J. 1986. *A Primer on Determinism*. Dordrecht: Reidel.

Fine, A. (1986), *The Shaky Game: Einstein, Realism, and the Quantum Theory*. Chicago: University of Chicago Press.

Franklin, Allan (1994), "How to Avoid the Experimenters' Regress." *Studies in the History and Philosophy of Science* 25: 97–121.

——— (1986), *The Neglect of Experiment*. Cambridge: Cambridge University Press.

——— (1997), "Calibration," *Perspectives on Science* 5: 31–80.

Friedman, M. (1974), "Explanation and Scientific Understanding," *Journal of Philosophy* 71: 5–19.

Galison, P. (1987), *How Experiments End*. Chicago: University of Chicago Press.

Geroch, R. (1977), "Prediction in General Relativity." *Minnesota Studies in the Philosophy of Science* 8: 81–93.

Habermas, J. (1971), *Knowledge and Human Interests*. McCarthy (transl.), Boston: Beacon Press.

Hacking, I. (1965), *Logic of Statistical Inference*, Cambridge: Cambridge University Press.

Hacking, Ian (1983), *Representing and Intervening*. Cambridge: Cambridge University Press.

Hanson, N. R. (1958), *Patterns of Discovery*. Cambridge: Cambridge University Press.

Hempel, C. (1949), "The Logical Analysis of Psychology," in H. Feigl and W. Sellars (eds.), *Readings in Philosophical Analysis*. New York: Appleton-Century-Crofts, 373–384.

——— (1962), "Explanation is Science and in History," in Colodny (ed.) *Frontiers of Science and Philosophy*. Pittsburgh: University of Pittsburgh Press.

——— (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.

Hempel, C. and P. Oppenheim (1948), "Studies in the Logic of Explanation," *Philosophy of Science* 15: 135–175.

Hesse, M. (1963), *Models and Analogies in Science*. London: Sheed and Ward.

Kitcher, P. (1989), "Explanatory Unification and the Causal Structure of the World," in Kitcher and Salmon (eds.), *Scientific Explanation. Minnesota Studies in the Philosophy of Science*, Vol. XIII. Minneapolis: University of Minnesota Press, 410–505.

Kuhn, T. (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kuhn, Thomas (1977), *The Essential Tension*. Chicago: Chicago University Press.

Latour, Bruno and Steve Woolgar (1979), *Laboratory Life: The Social Construction of Scientific Facts*. London: Sage.

Laudan, L. (1977), *Progress and its Problems: Towards a Theory of Scientific Growth*. Berkeley: University of California Press.

——— (1981), "A Confutation of Convergent Realism," *Philosophy of Science* 48: 19–50.

Leplin, J. (1997). *A Novel Defense of Scientific Realism*. New York: Oxford University Press.

Maxwell, G. (1962), "The Ontological Status of Theoretical Entities," in Feigl and Maxwell (eds.) *Minnesota Studies in the Philosophy of Science, Vol III*. Minneapolis: University of Minnesota Press, 3–27.

Mayo, Deborah (1996), *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Mill, John Stuart (1874), *A System of Logic: Ratiocinative and Inductive, 8th edition*. Reprinted in *Collected Works of John Stuart Mill, Vols. VII-VIII*, edited by J. M. Robson. London and Toronto: Routledge and Kegan Paul and University of Toronto Press, 1963–1991.

Nagel, E. (1961), *The Structure of Science*. New York: Harcourt, Brace and World.

Norton, John (1996) "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy*, *26*, pp. 333–66.

Oppenheim, P. and Putnam, H. (1958), "The Unity of Science as a Working Hypothesis." In Feigl, H., Scriven, M., and Maxwell, G. Eds. *Concepts, Theories, and the Mind-Body Problem*. Minneapolis: University of Minnesota Press, pp. 3–36.

Pickering, Andrew (1984), *Constructing Quarks*. Chicago: University of Chicago Press.

Polanyi, Michael (1958), *Personal Knowledge*. Chicago: University of Chicago Press.

Putnam, H. (1962), "What Theories are Not," in Nagel, Suppes, Tarski (eds.) *Logic, Methodology, and Philosophy of Science: Proceedings of the Nineteenth International Congress*. Stanford: Stanford University Press, 240–251.

Rheinberger, H. J. (1997), *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford: Stanford University Press.

Rosenberg, A. (1988), *Philosophy of Social Science*. Boulder: Westview Press.

Salmon, W. (1971), "Statistical Explanation and Statistical Relevance", in Salmon, Greeno, and Jeffrey (eds.), *Statistical Explanation and Statistical Relevance*. Pittsburgh, University of Pittsburgh Press, 29–87.

——— (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Sarkar, S. Ed. (1996a), *Science and Philosophy in the Twentieth Century: Basic Works of Logical Empiricism. Vol. 1. The Emergence of Logical Empiricism: From 1900 to the Vienna Circle*. New York: Garland.

——— Ed. (1996b), *Science and Philosophy in the Twentieth Century: Basic Works of Logical Empiricism. Vol. 2. Logical Empiricism at Its Peak: Schlick, Carnap, and Neurath*. New York: Garland.

——— Ed. (1996c), *Science and Philosophy in the Twentieth Century: Basic Works of Logical Empiricism. Vol. 3. Logic, Probability, and Epistemology: The Power of Semantics*. New York: Garland.

——— Ed. (1996d), *Science and Philosophy in the Twentieth Century: Basic Works of Logical Empiricism. Vol. 4. Logical Empiricism and the Special Sciences: Reichenbach, Feigl, and Nagel*. New York: Garland.

——— Ed. (1996e), *Science and Philosophy in the Twentieth Century: Basic Works of Logical Empiricism. Vol. 5. Decline and Obsolescence of Logical Empiricism: Carnap vs. Quine and the Critics*. New York: Garland.

——— (1998), *Genetics and Reductionism*. New York: Cambridge University Press.

——— (2005), *Molecular Models of Life: Philosophical Papers on Molecular Biology*. Cambridge, MA: MIT Press.

Schaffner, K. F. (1993), *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.

Scheffler, I. (1967), *Science and Subjectivity*. Indianapolis: Bobbs-Merrill.

Shimony, A. (1987), "The Methodology of Synthesis: Parts and Wholes in Low-Energy Physics." In Kargon, R. and Achinstein, P. Eds. *Kelvin's Baltimore Lectures and Modern Theoretical Physics*. Cambridge, MA: MIT Press, pp. 399–423.

Stigler, S. M. (1986), *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge: Harvard University Press.

Suppe, F. Ed. (1974), *The Structure of Scientific Theories*. Urbana: University of Illinois Press.

Taylor, C. (1971), "Interpretation and the Sciences of Man," *Review of Metaphysics* 25: 3–51.

Van Fraasen, Bas. *The Scientific Image*. Oxford: Clarendon Press, 1980.

Watson, John B. (1913), "Psychology as a Behaviorist Views It," *Psychological Review* 20: 158–177.

Weber, Max (1903/1949) *The Methodology of the Social Sciences*. Shils and Finch (eds.), New York: Free Press.

Wimsatt, W. (1987), "False Models as Means to Truer Theories." In Nitecki, N. and Hoffman, A. Eds. *Neutral Models in Biology*. Oxford: Oxford University Press, pp. 23–55.

Winch, P. (1958), *The Idea of Social Science and its Relation to Philosophy*. London: Routledge and Kegan Paul.

Woodger, J. H. (1929), *Biological Principles*. Cambridge: Cambridge University Press.

——— (1937), *The Axiomatic Method in Biology*. Cambridge: Cambridge University Press.

LIST OF CONTRIBUTORS

Alexander, Jason London School of Economics, United Kingdom

Armendt, Brad Arizona State University

Backe, Andrew Independent Scholar

Barrett, Jeff University of California at Irvine

Bechtel, William Washington University in St. Louis

Bouchard, Frederic University of Montreal, Canada

Bradie, Michael Bowling Green State University

Brown, Joshua University of Michigan

Byrne, Alex Massachusetts Institute of Technology

Cat, Jordi Indiana University

Craver, Carl Washington University in St. Louis

de Regt, Henk W. Vrije Universiteit, The Netherlands

Dickson, Michael Indiana University

DiSalle, Robert The University of Western Ontario, Canada

Downes, Stephen University of Utah

Eckardt, Barbara von Rhode Island School of Design

Eells, Ellery University of Wisconsin-Madison

Elga, Adam Princeton University

Ewens, Warren J. University of Pennsylvania

Falk, Raphael The Hebrew University, Israel

Fetzer, James University of Minnesota

Fitelson, Branden University of California at Berkeley

Folina, Janet Macalester College

Frigg, Roman London School of Economics, United Kingdom

Garson, Justin University of Texas at Austin

Gillies, Anthony University of Texas at Austin

Glennan, Stuart Butler University

Grandy, Richard Rice University

Gregory, Paul Washington and Lee University

Griffiths, Paul University of Queensland, Australia

Hájek, Alan California Institute of Technology

Hall, Ned Massachusetts Institute of Technology

Halvorson, Hans Princeton University

Hankinson-Nelson, Lynn University of Missouri at St. Louis

Hardcastle, Gary Bloomsburg University of Pennsylvania

Hardcastle, Valerie Virginia Polytechnic Institute and State University

Hartmann, Stephan London School of Economics and Political Science, United Kingdom

Hochberg, Herbert University of Texas at Austin

Hodges, Andrew University of Oxford

Hooker, Cliff A. University of Newcastle, Australia

Hull, David Northwestern University

Irvine, Andrew University of British Columbia, Canada

Joyce, Jim University of Michigan

Juhl, Cory University of Texas at Austin

Justus, James University of Texas at Austin

Kamlah, Andreas University of Osnabrück, Germany

Kincaid, Harold Center for Ethics and Values in the Sciences

Koertge, Noretta Indiana University

Larvor, Brendan University of Hertfordshire, United Kingdom

Laubichler, Manfred Arizona State University

Leplin, Jarrett University of North Carolina

Lipton, Peter University of Cambridge, United Kingdom

Little, Daniel University of Michigan-Dearborn

Lloyd, Elisabeth Indiana University

Loomis, Eric University of South Alabama

Ludlow, Peter University of Michigan

Lynch, Michael Cornell University

Lyre, Holger University of Bonn, Germany

MacLaurin, James University of Otago, New Zealand

Magnus, P. D. State University of New York at Albany

Majer, Ulrich University of Goettingen, Germany

Martinich, A. P. University of Texas at Austin

Motterlini, Matteo University of Trento, Italy

Nagel, Jennifer University of Toronto, Canada

Nickles, Thomas University of Nevada, Reno

Niiniluoto, Ilkka University of Helsinki

Noe, Alva University of California at Berkeley

Nyberg, Ian University of Texas at Austin

Odenbaugh, Jay University of California at San Diego

Okasha, Samir University of Bristol, United Kingdom

Perini, Laura Virginia Polytechnic Institute and State University

Pfeifer, Jessica University of Maryland, Baltimore County

Piccinini, Gualtiero Washington University in St. Louis

Plutynski, Anya University of Utah

Pojman, Paul University of Utah

Radder, Hannes Vrije Universiteit Amsterdam

Ramsey, Jeffry Smith College

Ratcliffe, Matthew University of Durham, United Kingdom

Richardson, Alan University of British Columbia

Roberts, John University of North Carolina

Rosenkrantz, Roger Independent Scholar

Roush, Sherrilyn Rice University

Ruetsche, Laura University of Pittsburgh

Sandell, Michelle Independent Scholar

Sankey, Howard University of Melbourne, Australia

Sarkar, Sahotra University of Texas at Austin

Shapere, Dudley Wake Forest University

Sigmund, Karl University of Vienna, Austria

Simchen, Ori University of British Columbia, Canada

Sober, Elliott University of Wisconsin-Madison

Stachel, John Boston University

Stadler, Friedrich K. University of Vienna, Austria

Stanford, P. Kyle University of California, Irvine

Stoljar, Daniel Australian National University

Stöltzner, Michael Institute Vienna Circle, Austria

Sundell, Tim University of Michigan

Suppes, Patrick Stanford University

Tauber, Alfred I. Boston University

Thornton, Stephen P. University of Limerick, Ireland

Vineberg, Susan Wayne State University

Wayne, Andrew Concordia University, Canada

Wilson, Jessica University of Toronto, Canada

Wilson, Robert A. University of Alberta, Canada

Wimsatt, William The University of Chicago

Witmer, D. Gene University of Florida

A TO Z LIST OF ENTRIES

A

Abduction
Adaptation and Adaptationism
Altruism
Analyticity
Anthropic Principle
Approximation
Artificial Intelligence
Astronomy, Philosophy of
Ayer, A. J.

B

Bayesianism
Behaviorism
Biological Information
Biology, Philosophy of
Bridgman, Percy Williams

C

Carnap, Rudolf
Causality
Chemistry, Philosophy of
Chomsky, Noam
Classical Mechanics
Cognitive Science
Cognitive Significance
Complementarity
Confirmation Theory
Connectionism
Consciousness
Conservation Biology
Conventionalism
Corroboration

D

Decision Theory
Demarcation, Problem of
Determinism
Duhem Thesis
Dutch Book Argument

E

Ecology
Economics, Philosophy of
Emergence
Empiricism
Epistemology
Evolution
Evolutionary Epistemology
Evolutionary Psychology
Experiment
Explanation
Explication

F

Feminist Philosophy of Science
Feyerabend, Paul Karl
Fitness
Function

G

Game Theory
Genetics

H

Hahn, Hans
Hanson, Norwood Russell
Hempel, Carl Gustav
Heritability
Hilbert, David

I

Immunology
Incommensurability
Individuality
Induction, Problem of
Inductive Logic
Innate/Acquired Distinction
Instrumentalism
Intentionality
Irreversibility

K

Kinetic Theory
Kuhn, Thomas

L

Lakatos, Imre
Laws of Nature
Linguistics, Philosophy of
Locality
Logical Empiricism

M

Mach, Ernest
Mechanism
Methodological Individualism
Molecular Biology

N

Nagel, Ernest
Natural Selection
Neumann, John von
Neurath, Otto
Neurobiology

O

Observation

P

Parsimony
Particle Physics
Perception
Phenomenalism
Physical Sciences, Philosophy of
Physicalism
Poincaré, Jules Henri
Popper, Karl Raimund
Population Genetics
Prediction
Probability
Protocol Sentences
Psychology, Philosophy of
Putnam, Hilary

Q

Quantum Field Theory
Quantum Logic
Quantum Measurement Problem
Quantum Mechanics
Quine, Willard Van

R

Ramsey, Frank Plumpton
Rational Reconstruction
Realism
Reductionism
Reichenbach, Hans
Research Programs
Russell, Bertrand

S

Schlick, Moritz
Scientific Change
Scientific Domains
Scientific Metaphors
Scientific Models
Scientific Progress
Scientific Revolutions
Scientific Style

Searle, John
Social Constructionism
Social Sciences, Philosophy of the
Space-Time
Species
Statistics, Philosophy of
Supervenience

T

Theories
Time
Turing, Alan

U

Underdetermination of Theories
Unity and Disunity of Science
Unity of Science Movement

V

Verifiability
Verisimilitude
Vienna Circle
Visual Representation

THEMATIC LIST OF ENTRIES

Biology

Adaptation and Adaptationism
Altruism
Biological Information
Biology, Philosophy of
Conservation Biology
Ecology
Evolution
Fitness
Genetics
Heritability
Immunology
Individuality
Innate/Acquired Distinction
Molecular Biology
Natural Selection
Population Genetics
Species

Epistemology and Metaphysics

Abduction
Analyticity
Approximation
Bayesianism
Causality
Cognitive Significance
Confirmation Theory
Conventionalism
Corroboration
Decision Theory
Demarcation, Problem of
Determinism
Duhem Thesis
Dutch Book Argument
Emergence
Empiricism
Epistemology
Evolutionary Epistemology
Experiment
Explanation
Explication
Feminist Philosophy of Science
Function
Incommensurability
Induction, Problem of
Inductive Logic
Instrumentalism
Laws of Nature
Logical Empiricism
Mechanisms
Observation
Parsimony
Perception
Phenomenalism
Physicalism
Prediction
Probability
Protocol Sentences
Rational Reconstruction
Realism
Reductionism
Research Programs
Scientific Change
Scientific Domains
Scientific Metaphors
Scientific Models
Scientific Progress
Scientific Revolutions
Scientific Style
Social Constructionism
Supervenience
Theories
Underdetermination of Theories
Unity and Disunity of Science
Unity of Science Movement
Verifiability
Verisimilitude
Vienna Circle
Visual Representation

Physical Sciences

Anthropic Principle
Astronomy, Philosophy of
Chemistry, Philosophy of
Classical Mechanics
Complementarity
Irreversibility
Kinetic Theory
Locality
Particle Physics
Physical Sciences, Philosophy of
Quantum Field Theory
Quantum Logic
Quantum Measurement Problem
Quantum Mechanics
Space-Time
Time

Principal Figures

Ayer, A. J.
Bridgman, Percy Williams
Carnap, Rudolf
Chomsky, Noam
Feyerabend, Paul Karl
Hahn, Hans
Hanson, Norwood Russell
Hempel, Carl Gustav
Hilbert, David
Kuhn, Thomas
Lakatos, Imre
Mach, Ernest
Nagel, Ernest
Neumann, John von
Neurath, Otto
Poincaré, Jules Henri
Popper, Karl Raimund
Putnam, Hilary
Quine, Willard Van
Ramsey, Frank Plumpton
Reichenbach, Hans
Russell, Bertrand
Schlick, Moritz
Searle, John
Turing, Alan

Psychology and Mind

Artificial Intelligence
Cognitive Science
Connectionism
Consciousness
Evolutionary Psychology
Intentionality
Neurobiology
Psychology, Philosophy of

Social Sciences

Behaviorism
Economics, Philosophy of
Game Theory
Linguistics, Philosophy of
Methodological Individualism
Social Sciences, Philosophy of the

Statistics

Statistics, Philosophy of

A

ABDUCTION

Scientific hypotheses cannot be deduced from the empirical evidence, but the evidence may support a hypothesis, providing a reason to accept it. One of the central projects in the philosophy of science is to account for this nondemonstrative inductive relation. The justification of induction has been a sore point since the eighteenth century, when David Hume ([1777] 1975) gave a devastating skeptical argument against the possibility of any reason to believe that nondemonstrative reasoning will reliably yield true conclusions (see Induction, Problem of). Even the more modest goal of giving a principled description of our inductive practices has turned out to be extremely difficult. Scientists may be very good at weighing evidence and making inferences; but nobody is very good at saying how they do it.

The nineteenth-century American pragmatist Charles Sanders Peirce (1931, cf. 5.180–5.212) coined the term *abduction* for an account, also now known as "Inference to the Best Explanation," that addresses both the justification and the description of induction (Harman 1965; Lipton 1991, 2001; Day and Kincaid 1994; Barnes 1995). The governing idea is that explanatory considerations are a guide to inference, that the hypothesis that would, if correct, best explain the evidence is the hypothesis that is most likely to be correct. Many inferences are naturally described in this way (Thagard 1978). Darwin inferred the hypothesis of natural selection because although it was not entailed by his biological evidence, natural selection would provide the best explanation of that evidence. When astronomers infer that a galaxy is receding from the Earth with a specified velocity, they do so because the recession would be the best explanation of the observed redshift of the galaxy's spectrum.

On the justificatory question, the most common use of abduction has been in the miracle argument for scientific realism. Hilary Putnam (1978, 18–22) and others have argued that one is entitled to believe that empirically highly successful hypotheses are at least approximately true—and hence that the inductive methods scientists use are reliable routes to the truth—on the grounds that the truth of those hypotheses would be the best explanation of their empirical success (see Putnam, Hilary; Realism) because it would be a miracle if a hypothesis that is fundamentally mistaken were found to have so many precisely correct empirical consequences. Such an outcome is logically possible, but the correctness of the hypothesis is a far better explanation of its success. Thus the miracle argument is itself an abduction—an inference to the best explanation—from the predictive success of a hypothesis to its correctness.

Like all attempts to justify induction, the miracle argument has suffered many objections. The miracle argument is itself an abduction, intended as a justification of abduction. One objection is that this is just the sort of vicious circularity that Hume argued against. Moreover, is the truth of a hypothesis really the best explanation of its empirical successes? The feeling that it would require a miracle for a false hypothesis to do so well is considerably attenuated by focusing on the history of science, which is full of hypotheses that were successful for a time but were eventually replaced as their successes waned (see Instrumentalism). The intuition underlying the miracle argument may also be misleading in another way, since it may rest solely on the belief that most possible hypotheses would be empirically unsuccessful, a belief that is correct but arguably irrelevant, since it may also be the case that most successful hypotheses would be false, which is what counts.

Abduction seems considerably more promising as an answer to the descriptive question. In addition to the psychologically plausible account it gives of many particular scientific inferences, it may avoid weaknesses of other descriptive accounts, such as enumerative induction, hypothetico-deductivism, and Bayesianism. Enumerative inferences run from the premise that observed *F*s are *G* to the conclusion that all *F*s are *G*, a scheme that does not cover the common scientific case where hypotheses appeal to entities and processes not mentioned in the evidence that supports them. Since those unobservables are often introduced precisely because they would explain the evidence, abduction has no difficulty allowing for such "vertical" inferences (see Confirmation theory).

If the enumerative approach provides too narrow an account of induction, hypothetico-deductive models are too broad, and here again abduction does better. According to hypothetico-deductivism, induction runs precisely in the opposite direction from deduction, so that the evidence supports the hypotheses that entail it. Since, however, a deductively valid argument remains so whatever additional premises are inserted, hypothetico-deductivism runs the risk of yielding the absurd result that any observation supports every hypothesis, since any hypothesis is a member of a premise set that entails that observation. Abduction avoids this pitfall, since explanation is a more selective relationship than entailment: Not all valid arguments are good explanations.

The relationship between abduction and Bayesian approaches is less clear. Like abduction, Bayesianism avoids some of the obvious weaknesses of both enumerative and hypothetico-deductive accounts. But the Bayesian dynamic of generating judgments of prior probability and likelihood and then using Bayes's theorem to transform these into judgments of posterior probability appears distant from the psychological mechanisms of inference (see Bayesianism). It may be that what abduction can offer here is not a replacement for Bayesianism, but rather a way of understanding how the Bayesian mechanism is "realized" in real inferential judgment. For example, consideration of the explanatory roles of hypotheses may be what helps scientists to determine the values of the elements in the Bayesian formula. Bayesianism and abduction may thus be complementary.

One objection to an abductive description of induction is that it is itself only a poor explanation of our inductive practices, because it is uninformative. The trouble is that philosophers of science have had as much difficulty in saying what makes for a good explanation as in saying what makes for a good inference. What makes one explanation better than another? If "better" just means more probable on the evidence, then the abductive account has left open the very judgment it was supposed to describe. So "better" had better mean something like "more explanatory." However, it is not easy to say what makes one hypothesis more explanatory than another. And even if the explanatory virtues can be identified and articulated, it still needs to be shown that they are scientists' guides to inference. But these are challenges the advocates of abduction are eager to address.

PETER LIPTON

References

Barnes, Eric (1995), "Inference to the Loveliest Explanation," *Synthese* 103: 251–277.

Day, Timothy, and Harold Kincaid (1994), "Putting Inference to the Best Explanation in Its Place," *Synthese* 98: 271–295.

Harman, Gilbert (1965), "The Inference to the Best Explanation," *Philosophical Review* 74: 88–95.

Hume, David ([1777] 1975), *An Enquiry Concerning Human Understanding.* Edited by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Oxford University Press.

Lipton, Peter (1991), *Inference to the Best Explanation.* London: Routledge.

——— (2001), "Is Explanation a Guide to Inference?" in Giora Hon and Sam S. Rakover (eds.), *Explanation: Theoretical Approaches and Application.* Dordrecht, Netherlands: Kluwer Academic Publishers, 93–120.

Peirce, Charles Sanders (1931), *Collected Papers.* Edited by C. Hartshorn and P. Weiss. Cambridge, MA: Harvard University Press.

Putnam, Hilary (1978), *Meaning and the Moral Sciences.* London: Hutchinson.

Thagard, Paul (1978), "The Best Explanation: Criteria for Theory Choice," *Journal of Philosophy* 75: 76–92.

See also **Bayesianism; Confirmation Theory; Induction, Problem of; Instrumentalism; Realism**

ADAPTATION AND ADAPTATIONISM

In evolutionary biology, a phenotypic trait is said to be an *adaptation* if the trait's existence, or its prevalence in a given population, is the result of natural selection. So for example, the opposable thumb is almost certainly an adaptation: Modern primates possess opposable thumbs because of the selective advantage that such thumbs conferred on their ancestors, which led to the retention and gradual modification of the trait in the lineage leading to modern primates. Usually, biologists will describe a trait not as an adaptation per se but rather as an adaptation *for* a given task, where the task refers to the environmental "problem" that the trait helps the organism to solve. Thus the opposable thumb is an adaptation *for* grasping branches; the ability of cacti to store water is an adaptation *for* living in arid deserts; the brightly adorned tail of the peacock is an adaptation *for* attracting mates; and so on. Each of these statements implies that the trait in question was favored by natural selection *because* it conferred on its bearer the ability to perform the task. In general, if a trait *T* is an adaptation for task *X*, this means that *T* evolved because it enabled its bearers to perform *X*, which enhanced their Darwinian fitness. This can also be expressed by saying that the *function* of the trait *T* is to perform *X*. Thus there is a close link between the concepts of adaptation and evolutionary function (Sterelny and Griffiths 1999; Ariew, Cummins, and Perlman 2002; Buller 1999).

Many authors have emphasized the distinction between a trait that is an adaptation and a trait that is *adaptive.* To describe a trait as adaptive is to say that it is *currently* beneficial to the organisms that possess it, in their current environment. This is a statement solely about the present—it says nothing about evolutionary history. If it turned out that Darwinism were wholly untrue and that God created the universe in seven days, many phenotypic traits would still qualify as adaptive in this sense, for they undeniably benefit their current possessors. By contrast, to describe a trait as an adaptation *is* to say something about evolutionary history, namely that natural selection is responsible for the trait's evolution. If Darwinism turned out to be false, it would follow that the opposable thumb is *not* an adaptation for grasping branches, though it would still be adaptive for primates in their current environment. So the adaptive/adaptation distinction corresponds to the distinction between a trait's *current utility* and its *selective history.*

In general, most traits that are adaptations are also adaptive and vice versa. But the two concepts do not always coincide. The human gastrointestinal appendix is not adaptive for contemporary human beings—which is why it can be removed without loss of physiological function. But the appendix is nonetheless an adaptation, for it evolved to help its bearers break down cellulose in their diet. The fact that the appendix no longer serves this function in contemporary humans does not alter the (presumed) fact that this is why it originally evolved. In general, when a species is subject to rapid environmental change, traits that it evolved in response to previous environmental demands, which thus count as adaptations, may cease to be adaptive in the new environment. Given sufficient time, evolution may eventually lead such traits to disappear, but until this happens these traits are examples of adaptations that are not currently adaptive.

It is also possible for a trait to be adaptive without being an adaptation, though examples falling into this category tend to be controversial. Some linguists and biologists believe that the capacity of humans to use language was not directly selected for, but emerged as a side effect of natural selection for larger brains. According to this theory, there was a direct selective advantage to having a large brain, and the emergence of language was simply an incidental by-product of the resulting increase in brain size among proto-humans. *If* this theory is correct, then human linguistic ability does not qualify as an adaptation and has no evolutionary function; thus it would be a mistake to look for a specific environmental demand to which it is an evolved response. But the ability to use language is presumably adaptive for humans in their *current* environment, so this would be an example of an adaptive trait that is not an adaptation. It should be noted, however, that many biologists and linguists are highly suspicious of the idea that human linguistic capacity was not directly shaped by natural selection. (See Pinker and Bloom 1990 and Fodor 2000 for opposing views on this issue.)

It sometimes happens that a trait evolves to perform one function and is later co-opted by evolution for a quite different task. For example, it is thought that birds originally evolved feathers as a way of staying warm, and only later used them to assist with flight. This is an interesting evolutionary phenomenon, but it creates a potential ambiguity. Should birds' feathers be regarded as an adaptation for thermoregulation or for efficient flight? Or perhaps for both? There is no simple answer to this question, particularly since feathers underwent considerable evolutionary modification after they first began to be used as a flying aid. Gould and Vrba (1982) coined the term "exaptation" to help resolve this ambiguity. An exaptation is any trait that originally evolves for one use (or arises for nonadaptive reasons) and is later co-opted by evolution for a different use.

How is it possible to tell which traits are adaptations and which are not? And if a particular trait is thought to be an adaptation, how is it possible to discover what the trait is an adaptation *for,* that is, its evolutionary function? These are pressing questions because evolutionary history is obviously not directly observable, so can be known only via inference. Broadly speaking, there are two main types of evidence for a trait's being an adaptation, both of which were identified by Darwin (1859) in *On the Origin of Species.* First, if a trait contributes in an obvious way to the "fit" between organism and environment, this is a prima facie reason for thinking it has been fashioned by natural selection. The organism/environment fit refers to the fact that organisms often possess a suite of traits that seem specifically tailored for life in the environments they inhabit. Consider for example the astonishing resemblance between stick insects and the foliage they inhabit. It seems most unlikely that this resemblance is a coincidence or the result of purely chance processes (Dawkins 1986, 1996). Much more plausibly, the resemblance is the result of many rounds of natural selection, continually favoring those insects who most closely resembled their host plants, thus gradually bringing about the insect/plant match. It is obvious *why* insects would have benefited from resembling their host plants—they would have been less visible to predators—so it seems safe to conclude that the resemblance is an adaptation for reducing visibility to predators. Biologists repeatedly employ this type of reasoning to infer a trait's evolutionary function.

Second, if a phenotypic trait is highly *complex*, then many biologists believe it is safe to infer that it is an adaptation, even if the trait's evolutionary function is not initially known. Bodily organs such as eyes, kidneys, hearts, and livers are examples of complex traits: Each involves a large number of component parts working together in a coordinated way, resulting in a mechanism as intricate as the most sophisticated man-made device. The inference from complexity to adaptation rests on the assumption that natural selection is the only serious scientific explanation for how organic complexity can evolve. (Appealing to an intelligent deity, though intellectually respectable in pre-Darwinian days, no longer counts as a serious explanation.) Again, inferences of this sort do not strictly amount to proof, but in practice biologists routinely assume that complex organismic traits are adaptations and thus have evolutionary functions waiting to be discovered.

The definition of an adaptation given above—any trait that has evolved by natural selection—is standard in contemporary discussions. In this sense, all biologists would agree that every extant organism possesses countless adaptations. However, the term has sometimes been understood slightly differently. R. A. Fisher, one of the founders of modern Darwinism, wrote that an organism

> "is regarded as adapted to a particular situation . . . only in so far as we can imagine an assemblage of slightly different situations, or environments, to which the animal would on the whole be less well adapted; and equally only in so far as we can imagine an assemblage of slightly different organic forms, which would be less well adapted to that environment." (1930, 41)

It is easy to see that Fisher's notion of adaptation is more demanding than the notion employed above. Fisher requires a very high degree of fit between organism and environment before the concept of adaptation applies, such that any small modification of either the organism or the environment would lead to a reduction in fitness. In modern parlance, this would normally be expressed by saying that the organism is *optimally* adapted to its environment.

It is quite possible for an organism to possess many adaptations in the above sense, i.e., traits that are the result of natural selection, without being optimally adapted in the way Fisher describes. There are a number of reasons why this is so. First, natural selection is a gradual process: Many generations are required in order to produce a close adaptive fit between organism and environment. Suboptimality may result simply because selection has yet to run its full course. Second, unless there is considerable environmental constancy over time, it is unlikely that organisms will evolve traits that adapt them optimally to any particular environment, given the number of generations required. So suboptimality may result from insufficient environmental constancy. Third, there may be evolutionary trade-offs. For example, the long necks of giraffes enable them to graze on high foliage, but the price of a long neck might be too high a center of gravity and thus a suboptimal degree of stability. Evolution cannot always modify an organism's phenotypic traits independently of each other: Adjusting one trait to its optimal state may inevitably bring suboptimality elsewhere. Finally, as Lewontin (1985) and others have stressed, natural selection can drive a species from one point in phenotypic space to another only if *each intermediate stage* is fitness enhancing. So, suboptimality may result because the optimal phenotypic state cannot be accessed from the actual state by a series of incremental changes, each of which increases fitness. For all these reasons, it is an open question whether natural selection will produce optimally adapted organisms.

It is worth noting that the Fisherian concept of optimal adaptation employed above is not *totally* precise, and probably could not be made so, for it hinges on the idea of a small or slight modification to either the organism or the environment, leading to a reduction in fitness. But "small" and "slight" are vague terms. How large can a modification be before it counts as too big to be relevant to assessing whether a given organism is optimally adapted? Questions such as this do not have principled answers. However, any workable concept of optimality is likely to face a similar problem. It is unacceptable to say that an organism is optimally adapted if there is no *possible* modification that would raise its fitness, for by that token no organism would qualify as optimally adapted. With sufficient imagination, it is always possible to think of phenotypic changes that would boost an organism's fitness—for example, doubling its fecundity while leaving everything else unchanged. (As John Maynard Smith wrote, "it is clearly impossible to say what is the "best" phenotype unless one knows the range of possibilities" [Maynard Smith 1978, 32]). So to avoid trivializing the concept of optimality altogether, some restriction must be placed on the class of possible modifications whose effects on organismic fitness are relevant to judging how well adapted the organism is in its current state. Spelling out the necessary restriction will lead to a concept similar to Fisher's, with its attendant vagueness.

The constraints on optimality noted in the earlier discussion of suboptimality show that natural selection *may* fail to produce organisms that are optimally adapted. But how important these constraints are in practice is a matter of considerable controversy. Some biologists think it is reasonable to assume that most extant organisms are optimally or nearly optimally adapted to their current environment. On this view, any phenotypic trait of an organism can be studied on the assumption that selection has fine-tuned the trait very precisely, so that there is an evolutionary reason for the character being exactly the way it is. Other biologists have less confidence in the power of natural selection. While not denying that selection *has* shaped extant phenotypes, they see the constraints on optimality as sufficiently important to invalidate the assumption that what has actually evolved is optimal in Fisher's sense. They would not seek adaptive significance in every last detail of an organism's phenotype. (See Maynard Smith 1978 and Maynard Smith et al. 1985 for a good discussion of this issue.)

The optimality question is just one aspect of an important and sometimes heated debate concerning the legitimacy of what is called "adaptationism" in evolutionary biology (Sober and Orzack 2001; Dupre 1987). Adaptationism encompasses both an empirical thesis about the world and a methodology for doing evolutionary research (Godfrey-Smith 2001). Empirically, the main claim is that natural selection has been by far the most important determinant of organismic phenotypes in evolutionary history—all or most traits have been directly fashioned by natural selection. Typically, adaptationists will also show some sympathy for the view that extant organisms are optimally adapted to their environments, in at least certain respects. Methodologically, adaptationists believe that the best way to study the living world is to search for the evolutionary function of organisms' phenotypic traits. Thus, for example, if an adaptationist observes an unusual pattern of behavior in a species of insect, the adaptationist will immediately assume that the behavior has an evolutionary function and will devote effort to trying to discover that function. Opponents of adaptationism reject both the empirical thesis and the methodological strategy. They emphasize the constraints on optimality noted above, as well as others; additionally, they point out that natural selection is not the only cause of evolutionary change and that organisms possess certain features that are nonadaptive and even maladaptive. Thus, it is a mistake to view the living world through an exclusively adaptationist lens, they argue.

The basic contours of the adaptationism debate have been in place for a long time, and indeed trace right back to Darwin. But the modern debate was instigated by Stephen Jay Gould and Richard Lewontin's famous article "The Spandrels of San Marco and the Panglossian Paradigm" (Gould and Lewontin 1979). These authors launched a forthright attack on what they saw as the extreme adaptationism prevalent in many evolutionary circles. They accused adaptationists of (a) uncritically assuming that every organismic trait *must* have an evolutionary function, (b) failing to accord a proper role to forces other than natural selection in evolution, and (c) paying insufficient heed to the constraining factors that limit selection's power to modify phenotypes at will. Unusually for a scientific article, "Spandrels" contains two striking literary allusions. Firstly, adaptationists are compared to Dr. Pangloss, a protagonist in Voltaire's satirical novel *Candide,* who despite suffering terrible misfortunes continues to believe that he inhabits the "best of all possible worlds." Gould and Lewontin's suggestion is that adaptationists commit a similar absurdity by viewing every aspect of an organism's phenotype as optimized by selection. Secondly, adaptationists are accused of inventing "Just So Stories" in their relentless search for evolutionary functions, that is, devising speculative hypotheses about traits' adaptive significance that owe more to their ingenuity than to empirical evidence. The reference here is to Rudyard Kipling's famous collection of children's stories, which include "How the Leopard Got Its Spots" and "How the Camel Got His Hump."

The title of Gould and Lewontin's paper illustrates what is perhaps their central complaint against adaptationist logic: the assumption, in advance of specific empirical evidence, that every trait has adaptive significance of its own. *Spandrel* is an architectural term that refers to the roughly triangular space between two adjacent arches and the horizontal above them; they are necessary by-products of placing a dome (or a flat roof) on arches. The spandrels beneath the great dome of St. Mark's Cathedral in Venice are decorated with elaborate mosaics of the four evangelists. Gould and Lewontin's point is that despite their ornate design, the spandrels are obviously not the raison d'être of the whole construction: Rather, they are inevitable by-products of the architectural design. Similarly, they suggest, certain anatomical and morphological traits of modern organisms may be inevitable by-products of their overall design, rather than directly shaped by selection. If so, such traits would not be adaptations, and it would be inappropriate to search for their evolutionary function. The human chin is a commonly cited example of a spandrel.

Gould and Lewontin's attack on adaptationism provoked an array of different reactions. Some of their opponents accused them of caricaturing adaptationism and thus attacking a strawman, on the grounds that no evolutionist had ever claimed every phenotypic trait of every organism to be adaptation, less still an optimal adaptation. There is certainly an element of truth to this charge. Nonetheless, Gould and Lewontin were writing at the height of the controversy over human sociobiology; and it is also true that *some* of the early proponents of that discipline advanced highly speculative hypotheses about the supposed evolutionary function of various behavioral patterns in humans, often on the basis of flimsy and anecdotal evidence. (This was not true of the best work in human sociobiology.) Gould and Lewontin's critique, even if overstated, was a useful corrective to this sort of naive adaptationism and led to a greater degree of methodological self-awareness among evolutionary biologists.

With hindsight, it seems that Gould and Lewontin's article has tempered, but not altogether eliminated, the enthusiasm felt by evolutionary biologists for adaptationism (cf. Walsh, "Spandrels," forthcoming). Many biologists continue to believe that cumulative natural selection over a large number of generations is the most plausible way of explaining complex adaptive traits, and such traits are abundant in nature. And despite the potential methodological pitfalls that the "Spandrels" paper warns against, the adaptationist research program continues to be highly fruitful, yielding rich insights into how nature works, and it has no serious rivals. Moreover, it *is* possible to test hypotheses about the adaptive significance of particular traits, in a variety of different ways. The *comparative method,* which involves comparing closely related species and trying to correlate phenotypic differences among them with ecological differences among their habitats, is one of the most common (cf. Harvey and Pagel 1991); it was employed by Darwin himself in his discussion of the Galapagos finches' beaks. Experimentally altering a trait, e.g., painting the plumage of a bird, and then carefully observing the effect on the organism's survival and reproductive success is another way of learning about a trait's adaptive significance. The most sophisticated work in evolutionary biology routinely uses these and other tests to adjudicate hypotheses about evolutionary function, and they bear little relation to the crude storytelling that Gould and Lewontin criticize. (See Endler 1986 for a good discussion of these tests.)

On the other hand, there is a grain of truth to Gould and Lewontin's charge that when a particular hypothesis about a trait's adaptive function is falsified, biologists will normally invent another adaptationist hypothesis rather than conclude that the trait is not an adaptation at all. However, not everyone agrees that reasoning in this way is methodologically suspect. Daniel Dennett agrees that adaptationists like himself offer "purely theory-driven explanations, argued a priori from the assumption that natural selection tells the true story—some true story or other—about every curious feature of the biosphere," but he regards this as perfectly reasonable, given the overall success of Darwinian theory (1995, 245). It is doubtful whether what Dennett says is literally true, however. There are many "curious features" of the biosphere for which it is not known whether there is an adaptationist story to be told or not. Take for example the prevalence of repeat sequences of non-coding "junk" DNA in the eukaryotic genome. This certainly qualifies as a curious feature—it took molecular biologists greatly by surprise when it was first discovered in the 1970s. But junk DNA has no known function—hence its name—and many people suspect that it has no function at all (though the current evidence on this point is equivocal; see Bejerano et al. 2004 for a recent assessment). So although Dennett is right that there is a general presumption in favor of adaptationist explanations among biologists, it is not true that every trait is *automatically* assumed to be an adaptation.

Samir Okasha

References

Ariew, Andre, Robert Cummins, and Mark Perlman (eds.) (2002), *Functions: New Essays in the Philosophy of Psychology and Biology.* Oxford: Oxford University Press.

Bejerano, Gill, Michael Pheasant, Igor Makunin, Stuart Stephen, W. James Kent, John S. Mattick, and David Haussler (2004), "Ultraconserved Elements in the Human Genome," *Science* 304: 1321–1325.

Buller, David J. (ed.) (1999), *Function, Selection and Design.* Albany: State University of New York Press.

Darwin, Charles (1859), *On the Origin of Species by Means of Natural Selection.* London: John Murray.

Dawkins, Richard (1986), *The Blind Watchmaker.* New York: W. W. Norton.

——— (1996), *Climbing Mount Improbable.* New York: W. W. Norton.

Dennett, Daniel (1995), *Darwin's Dangerous Idea.* New York: Simon and Schuster.

Dupre, John (ed.) (1987), *The Latest on the Best: Essays on Evolution and Optimality.* Cambridge, MA: MIT Press.

Endler, John A. (1986), *Natural Selection in the Wild.* Princeton, NJ: Princeton University Press.

Fisher, Ronald A. (1930), *The Genetical Theory of Natural Selection.* Oxford: Clarendon Press.

Fodor, Jerry (2000), *The Mind Doesn't Work That Way.* Cambridge MA: MIT Press.

Godfrey-Smith, Peter (2001), "Three Kinds of Adaptationism." In Elliott Sober and Steven Hecht Orzack (eds.), *Adaptationism and Optimality.* Cambridge: Cambridge University Press, 335–357.

Gould, Stephen Jay, and Elizabeth Vrba (1982), "Exaptation: A Missing Term in the Science of Form," *Paleobiology* 8: 4–15.

Gould, Stephen Jay, and Richard Lewontin (1979), "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme," *Proceedings of the Royal Society of London,* Series B 205: 581–598.

Harvey, Paul, and Mark Pagel (1991), *The Comparative Method in Evolutionary Biology.* Oxford: Oxford University Press.

Lewontin, Richard (1985), "Adaptation." In Richard Levins and Richard Lewontin (eds.), *The Dialectical Biologist.* Cambridge MA: Harvard University Press, 65–84.

Maynard Smith, John (1978), "Optimization Theory in Evolution," *Annual Review of Ecology and Systematics* 9: 31–56.

Maynard Smith, John, Richard Burian, Stuart Kaufman, Pere Alberch, James Campbell, Brian Goodwin, Russell Lande, David Raup, and Lewis Wolpert (1985), "Developmental Constraints and Evolution," *Quarterly Review of Biology* 60: 265–287.

Pinker, Steven, and Paul Bloom (1990), "Natural Language and Natural Selection," *Behavioral and Brain Sciences* 13: 707–784.

Sober, Elliott, and Steven Hecht Orzack (eds.) (2001), *Adaptationism and Optimality.* Cambridge: Cambridge University Press.

Sterelny, Kim, and Paul Griffiths (1999), *Sex and Death: An Introduction to Philosophy of Biology.* Chicago: Chicago University Press.

Walsh, Denis (ed.) (forthcoming). *Spandrels of San Marco: 25 Years Later.* Cambridge: Cambridge University Press.

See also **Evolution; Fitness; Natural Selection**

ALTRUISM

The concept of altruism has led a double life. In ordinary discourse as well as in psychology and the social sciences, behavior is called *altruistic* when it is caused by a certain sort of motive. In evolutionary biology, the concept applies to traits (of morphology, physiology, and behavior) that enhance the fitness of others at some cost to the self.

A behavior can be altruistic in the evolutionary sense without being an example of psychological altruism. Even if a honeybee lacks a mind, it nonetheless counts as an evolutionary altruist when it uses its barbed stinger to attack an intruder to the hive: The barb disembowels the bee but allows the stinger to keep pumping venom into the intruder even after the bee has died, thus benefiting the hive.

Symmetrically, a behavior can be altruistic in the psychological sense without being an example of evolutionary altruism. If one gives another a volume of piano sonatas out of the goodness of one's heart, one's behavior may be psychologically altruistic. However, the gift will not be an example of evolutionary altruism if it does not improve the other's prospects for survival and reproductive success or does not diminish one's own fitness.

Both types of altruism have given rise to controversy. According to psychological egoism, all our motives are ultimately selfish, and psychological altruism is merely a comforting illusion. Egoism was the dominant position in all major schools of twentieth-century psychology (Batson 1991). Within evolutionary biology, there has been considerable hostility to the idea that altruistic traits evolve because they benefit the group; according to one influential alternative viewpoint, the gene—not the group or even the individual organism—is the unit of selection (Dawkins 1976; Williams 1966).

Evolutionary Altruism

Evolutionary altruism poses a puzzle—it appears to be a trait that natural selection will stamp out rather than promote. If altruists and selfish individuals live in the same group, altruists will donate "fitness benefits" to others, whereas selfish individuals will not. Altruists receive benefits from the donations of other altruists, but so do selfish individuals. It follows that altruists will be less fit than selfish individuals in the same group. Natural selection is a process that causes fitter traits to increase in frequency and less fit traits to decline. How, then, can natural selection explain the existence of evolutionary altruism?

Darwin's answer was the hypothesis of group selection. Although altruists are less fit than selfish individuals in the same group, groups of altruists will be fitter than groups of selfish individuals. Altruistic traits evolve because they benefit the group and in spite of the fact that they are deleterious to altruistic individuals. Darwin (1859, 202) applied this idea to explain the barbed stinger of the honeybee; he also invoked the hypothesis to explain why men in a tribe feel morally obliged to defend the tribe and even sacrifice their lives in time of war (1871, 163–165).

With regard to natural selection, Darwin was a "pluralist": He held that some traits evolve because they are good for the individual, while others evolve because they are good for the group (Ruse 1980). This pluralism became a standard part of the evolutionary biology practiced in the period 1930–1960, when the "modern synthesis" was created. The idea of group adaptation was often applied uncritically during this period; however, the same can be said of the idea of individual adaptation. The situation changed in the 1960s, when group selection was vigorously attacked (Hamilton 1964; Maynard Smith 1964; Williams 1966). Its exile from evolutionary theory was hailed as one of the major advances of twentieth-century biology. Since then, the hypothesis of group selection has been making a comeback; many biologists now think that group selection is well grounded theoretically and well supported empirically as the explanation of some—though by no means all—traits (Sober and Wilson 1998). However, many other biologists continue to reject it.

The arguments mustered against group selection during the 1960s were oddly heterogeneous. Some were straightforwardly empirical; for example, Williams (1966) argued that individual selection predicts that sex ratios should be close to even, whereas group selection predicts that organisms should facultatively adjust the mix of daughters and sons they produce so as to maximize group productivity. Williams thought (mistakenly) that sex ratios were almost always close to even and concluded that group selection was not involved in the evolution of this trait. Other arguments were sweeping in their generality and almost a priori in character; for example, it was argued that genes, not groups, were the units of selection, because genes provide the mechanism of heredity (Dawkins 1976; Williams 1966). A third type of argument involved proposing alternatives to group selection. One reason why group selection was largely abandoned during this period was that inclusive fitness theory (Hamilton 1964), the theory of reciprocal altruism (Trivers 1971), and game theory (Maynard Smith and Price 1973) were widely viewed as alternatives. One reason the controversy continues is that it is disputed whether these theories are alternatives to or implementations of the idea of group selection (Sober and Wilson 1998).

Psychological Altruism

Although the concept of psychological altruism is applied to people and to actions, the best place to begin is to think of psychological altruism as a property of motives, desires, or preferences. Eve's desire that Adam have the apple is other directed; the proposition that she wants to come true—that Adam has the apple—mentions another person, but not Eve herself. In contrast, Adam's desire that he have the apple is self-directed. In addition to purely self-directed and purely other-directed desires, there are mixed desires: People desire that they and specific others be related in a certain way. Had Eve wanted to share the apple with Adam, her desire would have been mixed.

An altruistic desire is an other-directed desire in which what one wants is that another person do well. Altruistic desires, understood in this way, obviously exist. The controversy about psychological altruism concerns whether these desires are ever ultimate or are always merely instrumental. When one wishes others well, does one ever have this desire as an end in itself, or does one care about others only because one thinks that how they do will affect one's own welfare? According to psychological egoism, all ultimate motives are self-directed. When Eve wants Adam to have the apple, she has this other-directed desire only because she thinks that his having the apple will benefit her.

Psychological hedonism is one variety of egoistic theory. Its proponents claim that the only ultimate motives people have are the attainment of pleasure and the avoidance of pain. The only things one cares about as ends in themselves are states of one's own consciousness. This special form of egoism is the hardest one to refute. It is easy enough to see from human behavior that people do not always try to maximize their wealth. However, when someone chooses a job with a lower salary over a job that pays more, the psychological hedonist can say that this choice was motivated by the desire to feel good and to avoid feeling bad. Indeed, hedonists think they can explain even the most harrowing acts of self-sacrifice—for example, the proverbial soldier in a foxhole who throws himself on a live grenade to protect his comrades. The soldier supposedly does this because he prefers not existing at all over living with the tormenting knowledge that he allowed his friends to die. This hedonistic explanation may sound strained, but that does not mean it must be false.

Since hedonism is difficult to refute, egoism is also difficult to refute. However, that does not mean it is true. Human behavior also seems to be consistent with a view called *motivational pluralism;* this is the claim that people have both self-directed and other-directed ultimate aims. This theory does not assert that there are human actions driven solely by other-directed ultimate desires. Perhaps one consideration lurking behind everything one does is a concern for self. However, since actions may be caused by several interacting desires, pluralism is best understood as a claim about the character of our desires, not about the purity of our actions.

It is an interesting fact about our culture that many people are certain that egoism is true and many others are certain it is false. An extraterrestrial anthropologist might find this rather curious, in view of the fact that the behaviors one observes in everyday life seem to be consistent with both egoism and pluralism. One's convictions evidently outrun the evidence one has at hand. Is the popularity of egoism due to living in a culture that emphasizes individuality and economic competition? Is the popularity of pluralism due to the fact that people find it comforting to think of benevolence as an irreducible part of human nature? These questions are as fascinating as they are difficult to answer.

Social psychologists have tried to gather experimental evidence to decide between egoism and motivational pluralism. Egoism comes in a variety of forms; if each form could be refuted, this would refute egoism itself. According to one version of egoism, one helps needy others only because witnessing their suffering makes one uncomfortable; one helps help them for the same reason that one adjusts the thermostat when the room becomes too hot. According to a second version of egoism, people help only because they want to avoid censure from others or self-censure. According to a third version, people help only because helping provides them with a mood-enhancing reward. Batson (1991) argues that the experimental evidence disconfirms all the versions of egoism formulated so far (but see Sober and Wilson 1998).

The Prisoners' Dilemma

Game theory was developed by economists and mathematicians as a methodology for modeling the process of rational deliberation when what is best for one agent depends on what other agents do (von Neumann and Morgenstern 1947). These ideas were subsequently modified and extended, the result being evolutionary game theory, in which organisms (that may or may not have minds) interact with each other in ways that affect their fitness (Maynard Smith 1982). The prisoners' dilemma began as a problem in game theory, but it also has been much discussed in evolutionary game theory. It is central to understanding the conditions that must be satisfied for evolutionary altruism to evolve. Economists and other social scientists have studied the prisoners' dilemma because they think that people's behavior in such situations throws light on whether their motivation is altruistic or selfish.

Two actors come together; each must decide independently which of two actions to perform: cooperate or defect. The payoffs to each player are shown in the illustration.

Prisoners' dilemma: payoffs		Column player	
		Cooperate	Defect
Row player	Cooperate	Both get 3.	Row gets 1. Column gets 4.
	Defect	Row gets 4. Column gets 1.	Both get 2.

A simple dominance argument shows that each player should defect. Suppose you are the row player. If the column player cooperates, you are better off defecting (since 4 > 3); and if the column player defects, you are better off defecting (since 2 > 1). The column player is in exactly the same position. However, the resulting solution—defection by both—means that both players are worse off than they would have been if both had chosen to cooperate (since 3 > 2). The lesson is that rational decision making, with full information about the consequences of one's choices, can make one worse off.

How might the players avoid this dispiriting outcome? One possibility is for them to become less selfish. If they care about payoffs to self and payoffs to others in equal measure, each player will choose to cooperate (since 6 > 5 and 5 > 4). Another possibility is for the players to be forced to make the same decision, either by a mutually binding agreement or by a third party (a "leviathan"). Each of these changes puts the players into a new game, since the defining features of the prisoners' dilemma have now been violated.

In an evolutionary setting, the utilities described in the table are reinterpreted as "fitnesses"—that is, numbers of offspring. Suppose there are two types of individuals in a population: cooperators and defectors. Individuals form pairs, the members of each pair interact, and then each reproduces asexually, with the number of an organism's offspring dictated by the table of payoffs. Offspring exactly resemble their parents (there are no mutations). The older generation then dies, and the new generation of cooperators and defectors form pairs and play the game again. If this process is iterated over many generations, what will be the final configuration of the population? Since mutual defection is the solution to the prisoners' dilemma when the problem involves rational deliberation, can we infer that the solution for the evolutionary problem is universal defection?

The answer is no. Everything depends on how the pairs are formed. If they form at random, then universal defection will evolve. But suppose that like always pairs with like. In this instance cooperators receive a payoff of 3 and defectors receive a payoff of 2, which means that cooperation will increase in frequency. The dominance argument that settles the deliberational problem has no force in the evolutionary problem (Skyrms 1996). All one can tell from the table of payoffs is that the average fitness of cooperation must be somewhere between 3 and 1, and the average fitness of defection must be somewhere between 4 and 2. Cooperation and defection in evolutionary game theory are nothing other than evolutionary altruism and selfishness. The evolution of altruism thus crucially depends on how much correlation there is between interacting individuals (Skyrms 1996; Sober 1992).

When economists run experiments in which subjects play the prisoners' dilemma game, they find that people often behave "irrationally"—they do not always defect. What could explain this? One obvious possibility is that the payoff used in the experiment (usually money) doesn't represent everything that the subjects care about. It is true that defection can be reduced in frequency by manipulating the magnitude of monetary payoffs, but this does not show that the impulse to cooperate is always nonexistent; it shows only that this impulse is sometimes weaker than the impulse of self-interest. Furthermore, defenders of psychological egoism can easily bring cooperative behavior within the orbit of their theory. If people cooperate because and only because cooperation brings them pleasure, the behavior is consistent with psychological hedonism. Indeed, there is evidence from neuroscience that subjects do experience pleasure when they cooperate in the prisoners' dilemma (Billing et al. 2002). Note, however, that motivational pluralism is consistent with this finding. This raises the question whether behavior in such experiments throws light on the dispute between psychological egoism and motivational pluralism.

Elliott Sober

References

Batson, C. Daniel (1991), *The Altruism Question: Toward a Social-Psychological Answer.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Billing, J. K., D. A. Gutman, T. R. Zeh, G. Pagnoni, G. S. Berns, and C. D. Kitts (2002), "A Neural Basis for Social Cooperation," *Neuron* 35: 395–405.

Darwin, C. (1859), *On the Origin of Species.* London: Murray.

——— (1871), *The Descent of Man and Selection in Relation to Sex.* London: Murray.

Dawkins, R. (1976), *The Selfish Gene.* New York: Oxford University Press.

Hamilton, William D. (1964), "The Genetical Evolution of Social Behavior, I and II," *Journal of Theoretical Biology* 7: 1–16, 17–52.

Maynard Smith, John (1964), "Group Selection and Kin Selection," *Nature* 201: 1145–1146.

——— (1982), *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

Maynard Smith, John, and George Price (1973), "The Logic of Animal Conflict," *Nature* 246: 15–18.

Ruse, Michael (1980), "Charles Darwin and Group Selection," *Annals of Science* 37: 615–630.

Skyrms, Brian (1996), *The Evolution of the Social Contract.* New York: Cambridge University Press.

Sober, Elliott (1992), "The Evolution of Altruism: Correlation, Cost, and Benefit," *Biology and Philosophy* 7: 177–188.

Sober, Elliott, and David Sloan Wilson (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior.* Cambridge, MA: Harvard University Press.

Trivers, R. L. (1971), "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46: 35–57.

von Neumann, John, and Oscar Morgenstern (1947), *Theory of Games and Economic Competition.* Princeton, NJ: Princeton University Press.

Williams, George C. (1966), *Adaptation and Natural Selection.* Princeton, NJ: Princeton University Press.

See also **Evolution; Game Theory; Natural Selection**

ANALYTICITY

Analyticity can be characterized vaguely as follows. Analytic sentences, statements, or propositions are either true or known to be true "in virtue of" the meanings of the terms or concepts contained within them. Although a variety of notions resembling analyticity appear in writings of the late seventeenth century, notably in the work of Gottfried Leibniz (who uses "analytic" as a synonym for "practical" [Leibniz 1981, 624]), the more contemporary use of the notion of analyticity first appears in the work of Immanuel Kant. The account below will survey several historically significant notions of analyticity and then turn to more recent developments stemming from important objections to the notion of analyticity developed by Willard Van Quine and others. A strategy in defense of analyticity in response to Quine's challenge will then be sketched.

Historical Background

In his Criti*que of Pure Reason,* Kant characterized an *analytic* judgment in two ways. Firstly, he posited that

> the relation of a subject to the predicate is . . . possible in two different ways. Either the predicate *B* belongs to the subject *A,* as something which is (covertly) contained in this concept *A;* or *B* lies outside *A,* although it does indeed stand in connection with it. In the one case I entitle the judgment analytic, in the other synthetic. (Kant 1965, A6–7/B10)

Later Kant stated:

> All analytic judgments rest wholly on the principle of contradiction. . . . For because the predicate of an affirmative analytic judgment has already been thought in the concept of the subject, it cannot be denied of the subject without contradiction. (1965, A150/B189)

Subsequent philosophers such as Ayer (1946, 78) have questioned the equivalence of these two definitions, arguing that the former appears to provide a psychological criterion for analyticity, while the latter is purely logical.

It is tempting to impute to Kant the view that analytic judgments are vacuous (Ayer 1946, 77) or that they are "true by virtue of meanings and independently of fact" only (Quine 1953a, 21), but it is not clear that such attributions are correct. Subject–predicate judgments that Kant regarded as synthetic a priori, such as "A straight line between two points is the shortest," are plausibly regarded as being both necessary and true by virtue of meaning, since the predicate is conjoined to the subject through a kind of construction in intuition, and this construction conforms to what Kant called a "necessity inherent in the concepts themselves" (1965, B16–17; cf. Proust 1986; Hintikka 1974). Moreover, Kant at one point rejected the supposition that "tautologous" or vacuous judgments such as identity statements could be analytic, precisely in virtue of their vacuity:

> Propositions which explain *idem per idem* advance cognition neither analytically nor synthetically. By them I have neither an increase in distinctness nor a growth in cognition. (1965, XXIV, 667)

What then does distinguish analytic truths for Kant? A specification of the exact nature of analytic truths in Kant's work is complicated by his rather detailed theory of concepts (and his at times imprecise formulations), but if one attributes to him a common eighteenth-century doctrine of complex concepts according to which such concepts are built up by composition from simpler ones (such as a concept's *genus* and *differentia*), then a judgment of the form "*S* is *P*" could be said to be analytic if and only if the concept *S* has an analysis in which the concept *P* appears as a composite element (cf. De Jong 1995). Kant thought that "All bodies are extended" was of this form. In contrast, synthetic a priori subject–predicate judgments forge a connection between their subject and predicate through a construction in intuition by which the predicate is somehow "attached necessarily" to the subject but not antecedently contained within it (1965, B16–17). However, Kant believed that both forms of judgment advanced cognition and that both could eventuate in necessary truth: analytic judgments through clarification of a previously given but unanalyzed content, synthetic a priori judgments through construction of the predicate in intuition.

While the Kantian notion of analyticity thus bears only a loose resemblance to more contemporary ones, certain features of the contemporary notion were anticipated soon after Kant in the *Wissenshaftslehre* of Bernhard Bolzano. Bolzano thought that Kant's characterization of analyticity was too vague because the notion of the "containment" of the predicate by the subject was merely figurative (Bolzano 1973, 201). Furthermore, Kant's characterization allowed a proposition like "The father of Alexander, King of Macedon, was King of Macedon" to be classed as analytic even though, Bolzano claimed, no one would want to describe it as such (1973, 201). In brief, Bolzano proposed characterizing analytic propositions as those propositions that contained some referring idea such that, given any arbitrary substitution of that idea with another referring idea, the truth value of the original proposition would be preserved (provided that both ideas shared the same "range of variation"). Thus, the proposition "A depraved man does not deserve respect" is regarded as analytic because it contains a referring idea (man) that can be replaced with any other within a certain range while preserving truth (1973, 198). By contrast, the proposition "God is omniscient" is synthetic, since no constituent referring idea of this proposition could be arbitrarily replaced without altering the truth of the proposition.

Bolzano's account allowed him to introduce a distinction within the class of analytic propositions between those propositions recognizable as analytic solely by virtue of "logical knowledge" (such as "An *A* that is *B* is *A*") and those that required "extra-logical knowledge." His account also anticipated much later developments, both in making propositions (as opposed to judgments) the basis of the designation *analytic* and in relating logically analytic propositions to his notions of satisfaction and logical validity (Bolzano 1973, §147f).

Despite its originality, Bolzano's account of analytic propositions has certain counterintuitive consequences. For instance, the conjunction of any true proposition like "Grass is green" with an analytic proposition (e.g., "Grass is green and (*A* is *A*)") would appear to be analytic, since *A* occurs in the second conjunct vacuously and so can be arbitrarily substituted while preserving the truth value. But by the same criterion, the proposition "A bachelor is an unmarried adult male" would be synthetic, since for example "A widower is an unmarried adult male" is false and we can produce false sentences by substituting "married," "infant," and "female" for the terms "unmarried," "adult," and "male," respectively. Moreover, Bolzano's theory of propositions remained limited by his insistence that every statement be regarded as having the simple subject–copula–predicate form "*S* has *P*" (1973, 173ff).

The latter limitation was largely overcome with Gottlob Frege's development of the first-order predicate calculus, which, through the use of variable-binding generality operators ("quantifiers"), allowed logicians to deal with statements not obviously of the subject–predicate form, such as multiply general statements, or such statements as the claim that 1 plus 1 equals 2 or that any two distinct points determine a unique line. In his *Foundations of Arithmetic,* Frege defined analytic truths as all and only those truths that were provable from general logical laws and definitions (1974, §4). Since Frege believed that he could establish that arithmetic rested upon logic alone, his characterization of analyticity within the context of his new logic allowed him to assert, *contra* Kant, that mathematical truths were analytic. While this significantly improved logical apparatus thus allowed Frege to capture a notion of logical consequence that encompassed many new intuitively valid inferences, his restriction of analytic statements to those provable from general logical laws meant that statements such as "If *A* is longer than *B*, and *B* is longer than *C*, then *A* is longer than *C*" did not count as analytic.

This problem was addressed by Frege's student, Rudolf Carnap. Carnap attempted to "explicate" analyticity by giving the concept a precise characterization within formally specified languages (see Explication). In his *Logical Syntax of Language*, Carnap identified analytic statements as those statements that were consequences—in a given specified language such as his "Language II"—of logic and classical mathematics (1937, 100–1), and he later introduced a distinction between analytic truths that depend upon the meaning of nonlogical terms and those that do not (1956, 223–4). Unlike Bolzano and Frege, who saw in nonlogical analytic truths the need for the introduction of concepts "wholly alien to logic" (cf. Bolzano 1973, 198), Carnap thought that nonlogical analytic truths could nonetheless be precisely and unproblematically accommodated through the introduction of "meaning postulates" that stipulated relations of implication or incompatibility for the nonlogical predicates of an analytic statement (Carnap 1990, 68f). Since such postulates were stipulated, knowledge of them was for Carnap no different in kind than knowledge of the laws of logic or mathematics, which he also regarded as stipulated (see Conventionalism). It is apparently only through the introduction of something like Carnap's meaning postulates or allied devices that a reasonably precise characterization of analyticity in terms of truth by virtue of meaning is possible (see Carnap, Rudolf).

Besides Carnap, other members of the Vienna Circle (see Logical Empiricism; Vienna Circle) had seen in the notion of analytic truth the possibility of explaining how certain statements, such as "Either some ants are parasitic or none are" or the statements of the laws and theorems of logic, could be regarded as knowable a priori within an empiricist epistemology. The idea was that analytic statements, understood as statements that are true by virtue of meaning alone, would be true independently of matters of fact and known a priori to be true by anyone who understood the words used in them. As mere consequences of word meaning, such statements would not express genuine knowledge about the world and so would pose no threat to an epistemology that regarded all genuine knowledge as empirical (see Ayer 1946, 71–80). It should be noted, however, that although Carnap was allied with the Vienna Circle, the extent to which he viewed analyticity as serving in this role has been questioned (cf. Friedman 1987).

Quine's Two Dogmas of Empiricism

In "Two Dogmas of Empiricism," Quine (1953a) identified two "dogmas" that he claimed had been widely accepted by logical empiricists such as Carnap and other members of the Vienna Circle. One dogma was the view that there is a distinction between analytic statements, which are true by virtue of meaning alone, and synthetic ones, which have empirical content. The other dogma was the view that individual statements have empirical content, in the sense that each such statement can be confirmed or disconfirmed "taken in isolation from its fellows" (41). Quine suggested, in opposition to this dogma of "reductionism," that "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body" (41) (see Verifiability).

Against the first dogma, Quine argued that the notion of analyticity is suspect, in that it cannot be given a satisfactory explanation in independent terms. For example, he argued that one can define analytic statements as those that become truths of logic via a sequence of replacements of expressions by synonymous ones. However, Quine then noted that synonymy can be defined in terms of analyticity, in that two terms are synonymous just in case a statement of their equivalence is analytic. Thus what might have seemed to be an explication of the notion of analyticity turns out, in Quine's view, to cast doubt on the legitimacy of the notion of synonymy instead. Quine considered other attempts at defining analyticity, including appeals to the notion of a definition. For most terms in use, there is no record of anyone introducing the term into our language by explicit stipulation. Interestingly, Quine allowed that in cases of "the explicitly conventional introduction of novel notations for purposes of sheer abbreviation . . . the definiendum becomes synonymous with the definiens simply because it has been created expressly for the purpose of being synonymous with the definiens" (1953a, 26). However, asserting that stipulative definitions create synonymies is of doubtful coherence with the rest of what Quine advocates, as will be discussed further below.

Carnap agreed that expressions within natural languages are so vague and ambiguous that it is often unclear which should count as analytic. One can, however, create more precisely defined artificial languages in which the question of which sentences are analytic is perfectly well defined, in Carnap's view. Carnap called some basic sentences *meaning postulates* and essentially defined analytic statements for such a language *L* to be those sentences that follow, via the logical rules of *L,* from the meaning postulates alone (cf. Carnap 1956, 222f). Quine objected that Carnap's appeal to artificial languages is unhelpful, since Carnap does not explain what is special about meaning postulates of a language other than that they are listed under the heading "Meaning Postulates of *L*" in Carnap's books (Quine 1953a, 34). What Quine appeared to want was a language-transcendent notion of analyticity that would enable one (given an arbitrary language) to pick out the analytic sentences, and he denied that Carnap provided this.

The second dogma is related to the first, according to Quine, since if reductionism (confirmational localism, as opposed to holism) were true, then one could consider individual sentences "in isolation" and see whether they were confirmed and remained so under all possible observational circumstances. If so, they would count as analytic in this stipulated sense. However, Quine believed that whether any sentence is confirmed on a given occasion depends on its connections to other sentences and to further sentences in an immense web of belief. He thought that because of this holism, given any sentence *S*, there would always be some imaginable circumstances under which *S* would be disconfirmed. Thus, there are no sentences that would be confirmed in all circumstances, and so no analytic statements.

During the next decade or two, a large number of counterattacks and further objections were raised, culminating in Harman's summary of the current state of the argument as of about 1967. After that, most philosophers either were converted to Quineanism or, "however mutinously" (cf. Harman 1967), stopped claiming that anything was analytic.

At the turn of the twenty-first century, the arguments commonly raised against analyticity are essentially the same as those of Quine and Harman. The very few papers one finds in defense of analyticity, while they might contain an objection to one or another Quinean argument, nevertheless admit an inability actually to present a positive view concerning the distinction. Yet at the same time, most contemporary philosophers, even most who agree with Quine on analyticity, doubt Quine's conclusions about the "indeterminacy of translation." Most seem to think that if analyticity has to go, it will be for reasons much more straightforward and less contentious than those supporting the indeterminacy of translation (and of the indeterminacy of meaning generally). In what follows, the indeterminacy-based arguments against analyticity will be ignored (for these, see Boghossian 1996). The legitimacy of modal (or "intensional") notions such as possibility, which Quine rejected as part of his rejection of analyticity (Quine 1953b, 139ff), will also not be disputed. Again, most contemporary philosophers, including those who reject analyticity, are quite happy to employ intensional notions.

Early Responses to Quine: Carnap, Grice and Strawson, and Others

Carnap himself responded to Quine in a number of places (see especially 1956, 233f; 1990, 427ff) and also noted responses to Quine by other philosophers. One is by Martin (1952), who objects that although Quine finds the notions of synonymy and analyticity to be problematic on the grounds that no one, including Carnap, has given a language-transcendent concept of synonymy, Quine nonetheless accepts the notion of truth as unproblematic. Yet, neither has anyone provided a language-transcendent notion of truth. Parity of reasoning would thus seem to require that Quine abandon this latter notion also. Benson Mates (1951) argues that in fact there are behavioristically detectable differences between our treatment of analytic statements and merely well-confirmed synthetic ones, so that even if one adopts Quine's strict behaviorist methodology, one can draw the distinction.

More recently, John Winnie (1975) has argued that Carnap's later attempt at defining analyticity overcomes most of Quine's objections. The basic idea is that one takes a theory T, "Ramsifies" it (essentially, existentially generalize on the theoretical predicate terms by using second-order variables and quantifiers) to yield R, and then constructs the sentence

$$R \rightarrow T.$$

This conditional is taken to be the analytic content of the theory. While Winnie's proposal will not be addressed, it is mentioned as a potentially fruitful avenue for the defender of one form of analyticity.

Among the earliest and best-known responses to the Quinean offensive against analyticity is Grice and Strawson's "In Defense of a Dogma" (1956). The argument in effect states that since there is broad agreement in a number of central cases of analytic and synthetic sentences/propositions, there must be some substantive distinction that is being drawn, even if it is difficult to spell it out clearly. Harman has a response to this argument that must be confronted, however (see below).

In their paper, Grice and Strawson further suggest that Quine's position on stipulative definition for the purposes of abbreviation is incoherent. If, as Quine argues at length, the concept of analyticity is unintelligible, then how can it be intelligible to stipulate that some sentence has this unintelligible feature? Some important later defenders of Quine seem to agree that Quine made a slip here. Harman (1967), for instance, rejects Quine's concession on behalf of analyticity. Even later supporters of Quine's view of analyticity such as William Lycan (1991) continue to try to make a case against stipulative definition as a source of analyticity. It will be taken for granted here, along with Grice and Strawson, Harman, and Lycan, that the opponent of analyticity should disavow stipulative varieties. Furthermore, their main arguments all work against the case of stipulation if they work at all, so that good responses on behalf of stipulation will undercut the Quinean threat as a whole. For these reasons the discussion below will focus on the crucial issue of stipulative definitions when suggesting possible responses to Quinean arguments.

Harman's Synthesis of the Case Against Analyticity

Gilbert Harman neatly summarized the state of the dialectic as of the mid-1960s (Harman 1967), and similar arguments have continued to define terms of the debate into the present. Harman presents a broadly Quinean case against analyticity, incorporating Quinean responses to a number of "first wave" criticisms. Harman's arguments can be summarized as follows:

1. *Circularity.* Analyticity can be defined or explained only in terms of a family of concepts interdefinable with it (1967, 128f).
2. *Empty extension.* There are in fact no analytic sentences, just as there are no witches (125).
   a. Any (assent disposition toward a) sentence is revisable, and since analytic sentences would have to be unrevisable, there are no such sentences.
3. *Nonexplanatoriness.* The analytic/synthetic distinction does not explain anything; it is part of a "bad empirical theory" (136ff).
4. Analyticity does not account for a priori knowledge in the way logical empiricists claimed (130f):
   a. "Truth in virtue of meaning" makes no sense.
   b. Knowledge of truth by virtue of knowledge of meanings is impossible.
   c. Stipulation does not guarantee truth, in part because there is no way to draw a distinction between "substantive postulates" and stipulative definitions (Harman differs on this point from Quine [1953a]).

Depending upon the views of a given defender of analyticity, these objections might be faced in a variety of ways. Rather than present a panoply of possible responses and counterobjections, one strand of the dialectic will be emphasized, exhibiting one possible avenue of defense against these objections. Keeping in mind that this is only one possible view concerning analyticity, one can distinguish between statements that no empirical evidence is allowed to count for or against and those that are empirically defeasible. Call the former statements analytic and the latter synthetic. Examples of analytic statements might be that bachelors are unmarried men, that trilaterals have three sides, and that for any A and B, if A murdered B, then B died. Stipulative definitions of newly introduced terms, for purposes of abbreviation, are in the present view paradigmatic cases of analytic statements. A necessary condition for being a *statement,* in the present use, is that a sentence used on an occasion be in accord with certain norms of use. The notion of a norm may well appear suspect to a Quinean (cf. Quine 1953b, 32f), and this raises a general question as to whether and to what extent a defender of the notion of analyticity is required to use only notions that Quine himself regards as legitimate. This important issue will be discussed below.

Many authors have responded to the first argument presented by Harman above, that analyticity has no noncircular characterization. A particularly clear response is presented by Hans-Johann Glock (1992), who examines Quine's criteria of adequacy for a satisfactory definition of the analytic. Quine's objections to various proposed definitions invariably involved his claiming that concepts used in the definienda are themselves definable in terms of analyticity. But here Glock points out that it appears that Quine has unfairly stacked the rules against his opponents:

> [Quine] rejects Carnap's definition—via "meaning postulates"—of metalinguistic predicate "analytic in L_0" for formal languages on the grounds that this only explains "analytic in L_0" but not "analytic." . . . And this objection must surely amount to the requirement that an explanation of "analytic" give its *intension,* and not merely its *extension*. . . . It follows that Quine's challenge itself depends essentially upon the use of an intensional concept . . . whilst he forbids any response which grants itself the same license. Yet, it is simply inconsistent to demand that the meaning of "analytic" be explained, but to reject putative definitions solely on the grounds that they depend on the use of intensional concepts.

There is another inconsistency between Quine's insistence that the notion of analyticity be reduced to extensional ones and his requests for an explanation of the meaning of analyticity. For the latter demand cannot be fulfilled without using those concepts to which, as Quine himself has shown, the explanandum is *synonymous* or *conceptually* related, that is, the notions he prohibits. Consequently Quine's circularity charge comes down to the absurd complaint that the analytic can be explained only via synonyms or notions to which it is conceptually related and not via notions to which it is conceptually unrelated. (Glock 1992, 159). In other words, it appears that Quine is requiring an explanation of the meaning of the term *analytic* (as opposed to simply a specification of its extension), while simultaneously denying the intelligibility of the distinction. Furthermore, if the circularity worry is that analyticity cannot be defined except in terms that are interdefinable with it, it is unclear that analyticity fares worse in this respect than any concept whatsoever. The Quinean may retort that there are more or less illuminating explanations of meaning, and those given by defenders involve such small circles that they are not helpful at all. Even granting this, however, the concern remains that Quine's demands are inconsistent with his own expressed views.

Harman's second argument states that just as scientists have discovered that there are no witches and no phlogiston, Quine and others have discovered that there are no analytic sentences. A possible reply to this objection would be to suggest that the cases of witches and phlogiston are crucially disanalogous to the case of analyticity. Witches, for instance, are alleged to possess special causal powers. It is plausible to think that scientists can and have discovered that there are no beings with such causal powers. Furthermore, whether someone is a witch is logically or conceptually independent of whether anyone believes that that person is a witch, or treats that person as a witch. By contrast, analytic sentences are arguably more like chess bishops. In such a view, what makes something a chess bishop is that it is treated as one. Similar accounts might be given for presidents or tribal chieftains, laws, and a host of other familiar items.

Suppose, for the purposes of analogy, that some gamers invent a game at one of the weekly meetings of the Game-Definers' Club and call it Chess. They stipulate that bishops move only diagonally and begin at a square on which there appears a small sculpture resembling a Roman Catholic bishop. Suppose now that someone who has observed the proceedings and agrees to the stipulation says that it has been discovered that chess bishops move horizontally, or alternatively that there are no chess bishops. What could the gamers say to such a bizarre bishop skeptic? It seems as though skepticism is beside the point in such cases, or is even unintelligible. Similarly, one might respond to Quine's and Harman's second objection by noting that their talk of "discovering that there are no analytic sentences" makes no sense, in a way akin to that in which talk of discovering that there are no chess bishops, or that chess bishops move horizontally rather than diagonally, makes no sense. Someone who denies a stipulated rule of a game simply fails to understand the proceedings.

The second argument presented by Harman above goes on to state that there are in fact no analytic sentences, since scientists do (or would, in some conceivable circumstance) count some evidence against any sentence. There are a number of issues raised by this objection, to which several replies are possible. First, even if one decided that, as a matter of empirical fact, no one has ever used a sentence in an "evidentially isolated" way, this would be irrelevant to the question whether someone *could* do so, or whether scientists or philosophers can do so now, for instance in an attempt to "rationally reconstruct" or explicate scientific practice. Second, no advocate of analyticity ever denied that any given *sentence* could be used in a variety of ways in different contexts, for instance as evidentially isolated in one context or to express a synthetic claim in another. Here the analogy with game pieces can be extended. The fact that a miniature piece of sculpture could be used as a chess piece, and later as a paperweight, need not undercut the claim that it is a chess bishop as employed on a particular occasion. Finally, the defender of analyticity need not deny that it may be unclear in particular cases whether one should count a sentence as it was used on a particular occasion as analytic or as synthetic. One might grant that the language "game" often proceeds without completely determinate rules governing the uses of sentences while at the same time maintaining that some applications of expressions are evidence independent in the sense required for analyticity.

Harman's third major claim is that the notion of analyticity is part of a bad empirical theory. In reply, one might propose that analyticity need not be conceived as part of an empirical theory at all, any more than chess bishophood is. Carnap and some other logical empiricists treated metalinguistic notions as ultimately part of a (behavioristic) psychology, thereby ultimately to be evaluated in terms of their causal-explanatory virtues. The advocate of such a position would presumably have to address Harman's objection by pointing out ways in which standard behavior with respect to analytic sentences could be distinguished from standard behavior with respect to nonanalytic ones (cf. Mates 1951). However, a defender of analyticity might suggest instead that analyticity is a concept that is not typically employed within a predictive theory. For instance, when one says that a certain piece on a board is a bishop, one is typically not making a predictive claim about how it will move (although on some occasions one might be using the sentence in this way), but rather specifying what rules govern the movements of a certain object within a certain game. Likewise, when someone is described as a local chieftain, it is often simply to note how the chieftain is to be treated, what counts as appropriate responses to the chieftain's commands, and so on. So too, when one stipulatively defines terms, one may assign evidentially isolated roles to the definitions in the practice of description. This is not to deny that one can make predictions about how people within some group will move a bishop, treat a chieftain, or hold on to a definition sentence in the face of evidence. But stipulations are not empirical hypotheses about future behavior. They are prescriptions about what constitutes appropriate uses of expressions in a language. Moving a bishop in chess along the rank and file, for instance, does not falsify an empirical prediction but rather indicates a failure to understand or play by the rules. That such a prescriptive, evidentially isolated function exists for some statements is evident from an examination of common activities of playing games, exhibiting deferential attitudes toward leaders, and other practices. If Harman's third argument is to go through, then it appears that he must show how it is that all such apparently prescriptive statements are actually elements of predictive empirical theories.

His fourth objection includes the caveat that truth by virtue of meaning is an incoherent notion. Harman notes that it is difficult to see why the statements "Copper is a metal" and even "Copper is copper" are not made true by the fact that copper is a metal, and that copper is copper (or a more general fact concerning self-identities), respectively. Harman is correct to note that logical empiricists often thought that analytic sentences did not "answer to facts" for their truth in the way that nonanalytic sentences do. But a defender of analyticity need not make such claims. One might grant that the statement "Copper is copper" is made true by the fact that copper *is* copper. Furthermore, one can agree that copper would be copper whether or not anyone stipulated anything, although the fact that the sentence "Copper is copper" expresses what it does depends on human convention.

Perhaps of most importance in Harman's fourth argument is the claim that knowledge of truth by virtue of knowledge of meanings is impossible. The reason for the importance of this objection is that empiricists and others employed the notion of analyticity chiefly in order to solve an epistemological problem concerning the apparent a priori nature of our knowledge of logic and mathematics. If analyticity does not solve the basic problem for which it was invoked, then any further appeal to the notion threatens to be pointless, even if the notion proves to be coherent. Harman's objection can be put as follows. An empiricist might claim that what is distinctive about an analytic truth is that one cannot understand the claim without "seeing" that it is true. But "seeing" is a success term that begs the question. Even if it turns out that people by and large cannot fail to believe a sentence *S* as soon as they understand *S,* this could be due to a mere psychological compulsion, and so would have no epistemic (justificatory) relevance.

A possible reply on behalf of a defender of analyticity may be presented by returning to the meeting of the Game-Definers' Club. At this meeting, someone specifies a game for the first time and calls it Chess. Part of the specification includes the rule that there are to be bishops on the board at the beginning of any game-play sequence, that four miniature sculptures he points to are to be used as the bishops in the game to be played on that night, and that bishops move only diagonally if at all. Everyone agrees—well, almost everyone. After hearing the specification of the new game, the most recent Quinean infiltrator of the club, Skep, complains that no one in the room knows that bishops move diagonally, or that the particular piece pointed to is a chess bishop. Just saying that something is true does not make it so, Skep asserts, even if all of the club members assent to it. Further, Skep "shows us how to falsify" the claim that bishops move only diagonally by moving one of the pieces that club members agreed to treat as bishops along a file rather than diagonally. Club members might refrain (at first) from violent methods for the elimination of Quinean skepticism at such meetings and try to explain that the troublesome club member fails to understand the practice of stipulative definition. As for the concern that saying that something is so does not make it so, the appropriate response might be, "Yes it does, in the case of stipulative definition of novel terms." A Quinean may simply insist that stipulation does not guarantee truth and that no one has shown that it does. In response, one might grant that it has not been demonstrated that coherent stipulative definitions are true, any more than it has been proved that in chess, bishops move only diagonally, or that it is illegal to drive more than 70 miles per hour in certain areas. But it looks as though the Quinean must either deny that *any feature at all* can be guaranteed by conventional stipulation, including features such as legality, or explain why truth is relevantly different from these other features. The former denial seems hopelessly counterintuitive, whereas the latter seems of doubtful coherence, since if, e.g., the legality of *X* can be stipulated, the truth that *X* is legal seems guaranteed. Nevertheless, interesting issues may be raised here concerning what features of sentences can be said to be up to us, for even if one grants that one can make the case that no empirical evidence counts against a sentence *S* (and thereby nothing counts against the truth of the proposition expressed by *S*), this does not directly yield the truth of *S.*

What sorts of features of sentences can be determined by "arbitrary" decision? If one can decide what counts as evidence for or against a sentence, why not also grant that one can decide that a sentence expresses a truth? But can one even decide what counts for or against an arbitrary sentence? A completely satisfying account of analyticity will require plausible answers to these and related questions.

Does one *know* that stipulative definitions are true? Suppose, for simplicity, that one knows that the stipulative definition is not logically inconsistent with sets of other purportedly analytic statements and that the collection of stipulative definitions does not entail any paradigmatically empirical claim. There are several questions that could be considered, but let us focus on two: whether it is known that the sentence *S* is true and whether it is known that *p* is true, where *p* is the proposition expressed by *S.*

As regards *S,* there can be intelligible and even appropriate skepticism concerning, say, whether people within a certain group have all agreed to treat *S* as evidentially isolated, and thus skepticism as to whether *S* expresses an analytic proposition within the language of that group. Harman, Lycan, and other Quineans, however, apparently believe that even in a setting in which everyone agrees that *S* is evidentially isolated and in which it is common knowledge that everyone agrees, *S* still might, so far as anyone knows, be false. In the view suggested here on behalf of the advocate of analyticity, there is no intelligible room for skepticism concerning the truth of *S* for those present, any more than there is room for skepticism about how bishops move in the game of Chess at the Game-Definers' Club meeting.

When Quineans are asked for an account in which such sentences are falsified by evidence, they often provide an account that, to their opponents, looks like a story in which a different game is eventually called Chess, or in which a different concept is meant by the term *bachelor*. The Quineans then complain that the criteria employed to individuate games, concepts, or languages beg the question. Criteria of individuation, they claim, are themselves further claims open to empirical falsification, whereas the advocate of analyticity appeals to constitutive rules in the case of Chess and meaning postulates in the case of concepts or languages. So the advocate of analyticity who adopts the strategy suggested should present an account of individuation that coheres with his or her overall position. A natural proposal here would be to say that criteria of individuation are themselves analytic and stipulated (at least in the case of some terms, including new terms stipulatively defined, artificially extended languages, and some games). To be sure, the Quinean will object that this response assumes the notion being defended. But defenders of analyticity might reply that they need only defend the coherence of the notion of analyticity, and that it is the Quinean who has not supplied an independent argument against analyticity that a defender should feel any compulsion to accept. A related issue, a disagreement over whether the introduction of a new concept (a term as it is employed) requires justification, whether epistemic or pragmatic, is discussed a bit further in the final section.

As for knowledge that *p* is true, one might think that talk of having such knowledge makes sense in contrast to "mere" true belief, or perhaps mere justified true belief. But in the case of stipulative definitions, there is arguably no such contrast to be drawn. In such a conception of what is central to knowledge, one should either deny that one knows that *p* is true, where *p* is something stipulated, or warn that such claims are misleading. On the other hand, if one thinks that what is fundamental to having knowledge that *p* is true is the impossibility of being wrong about *p*, then if stipulation guarantees truth (of the sentence *S* stipulated, by guaranteeing that *S* expresses a true proposition), one might want to say that one knows that *p* is true, where *p* is the proposition expressed by *S* and where it is common knowledge that this is how *S* is being employed.

A number of the most important objections to the concept of analyticity (or the claim that some sentences express analytic truths) have been discussed thus far, along with one possible line of defense against the Quineans. Many other lines of objection and reply can be given (see Mates 1951; Martin 1952; Grice and Strawson 1956; Winnie 1975; and Boghossian 1996). The following discussion turns to some related questions.

Quine's Web of Belief

Quine grants that one can draw a distinction between sentences based upon their proximity to the center of a so-called web of belief. According to Quine's metaphor, altering one's assent behavior in regard to some sentences will "force" more drastic revisions in the rest of the web, whereas revising assent behaviors toward other sentences "near the periphery" will not have much effect on the rest of the web (see, e.g., Quine 1953a, 42–43). Thus for Quine, evidential isolation is a matter of degree, whereas according to the picture presented by the advocate of analyticity, it seems to be an all-or-nothing affair. Quineans think that the web metaphor captures everything sensible in the notion of analyticity while avoiding its pitfalls.

This picture has met with some significant challenges. In order to see why, one must look closely at the web metaphor that Quine employs. One problem noted by Laurence BonJour and others (BonJour 1998, 63ff; see also Wright 1980) is that Quine's talk of "connections" between elements of the web is misleading, in that in Quine's own view these connections are not normative or conceptual, but are instead causal or empirical. Further, for Quine, logical principles are themselves simply elements of the web. But if this is the case, it is unclear why any revision of the web is at any time required. As BonJour puts it:

> Thus the basis for any supposed incompatibility within any set of sentences (such as that which supposedly creates a need for revision in the face of experience) can apparently only be some further sentence in the system. . . . The upshot is that even the revision of one's epistemic or logical principles . . . turns out not to be necessary, since at some level there will inevitably fail to be a further sentence saying that the total set of sentences that includes those principles and that seems intuitively inconsistent really is inconsistent. This means that *any* non-observational sentence or set of sentences can always be retained. (BonJour 1998, 94)

It is not clear that the problem that BonJour raises can be restricted to nonobservational sentences. BonJour's argument is analogous to the famous "Achilles and the tortoise" argument of Lewis Carroll (1895). The tortoise asks Achilles why *B* should be inferred from *A* and *A* Ù *B*, and Achilles simply adds further sentences to the premise set at each stage, thereby failing to make any progress toward his goal. Similarly, BonJour argues that showing why a revision is required at some stage of inquiry cannot be accomplished by merely adding further sentences to the set supposedly requiring revision.

Another problem concerns the fact that Quine's metaphor of a web of belief allows for a loosely defined notion of greater or lesser degrees of the revisability of assent dispositions toward sentences. There are at least two notions of proximity to the center of the web to consider here. One is in terms of the pragmatic difficulties involved in giving up assent to a sentence *S*. The other is in terms of the "responsiveness" (or probabilities of change), to nerve firings, of assent dispositions toward *S.* One is a pragmatic notion and the other a ("merely") causal notion. Quine's view is that the former feature explains the latter. However, consider the classic philosophical example, "Bachelors are unmarried." Both Quine and Carnap will agree that this has a high degree of evidential isolation. As even supporters of Quine have acknowledged, it is false to say that giving up this sentence would require drastic revisions in our web of belief. Thus, the pragmatic difficulty of giving up such a sentence seems irrelevant to its low probability of being revised.

Putnam (1975), for example, thinks that Quine is dead wrong when he takes isolation to be a result of proximity to the center of the web. Putnam thinks that the only unrevisable sentences are near the periphery, at least in the pragmatic sense of not being connected with much else in the web. Putnam suggests that one call analytic those terms that have only a single criterion for their application, since there cannot arise any empirical pressure to revise the definition of a term if it is a "one-criterion" term, whereas if a term has two distinct criteria of application, then the possibility of an incompatibility between these criteria, and hence a breakdown in the application of the term, can arise. Putnam's notion of a one-criterion term as an account of analyticity, however, has been objected to on the basis that whether a term is one-criterion or not presupposes an analytic/synthetic distinction. For instance, should one count "Bachelors are unmarried men" and "Bachelors are unwed men" as two criteria or as two synonymous expressions of the same criterion? Clearly the latter seems more natural, and yet it presupposes a form of synonymy from which analyticity is definable, and thus does not provide a conceptually independent characterization of analyticity.

A complaint made against a number of responses to Quine's and Harman's arguments is that some, if not all, of them presuppose the distinction between analytic and synthetic propositions and for that reason are inefficacious. It is indeed difficult to respond to someone who denies the notions of meaning and synonymy without appealing to such notions, at least implicitly. However, in response to this worry the defender of analyticity can reasonably try to shift the onus of proof. One might suggest that the Quinean fails to correctly assess the status of an attack on a distinction. One sort of objection to a concept is that it is useless. Another is that the concept is incoherent or self-contradictory. From the point of view of a nonpragmatist, the Quinean claim that distinctions that are not explanatory or are otherwise useless are therefore nonexistent is puzzling. One may stipulate that "redchelors" are red-haired bachelors, for instance. Is there no such thing as a redchelor, or no distinction between redchelors and others, just because no one should have any interest in who is a redchelor, or in how many redchelors there are? It is one thing to claim that the distinction is useless for any explanatory purposes. It is another to say that the distinction is ill-defined, or alternatively that it is unreal. So far, it might be argued, the concept of analyticity has not been shown to be incoherent. Whether it is useless is also doubtful, especially in the clear case of stipulative definitions of new terms for the purposes of abbreviation. In response, the Quinean might grant much of this and yet propose that Quine's criticism has at the very least cast into doubt the pretensions of the logical empiricists, who thought that they could capture a wide variety of statements, such as those of logic and mathematics, with a single criterion.

It remains unclear whether adopting a Quinean practice (which disallows setting aside sentences as evidentially isolated in the absence of pragmatic justification) is somehow pragmatically deleterious compared with the practice of counting some sentences as isolated. To the extent that all that one is interested in is predicting future nerve-ending states from past ones, it may be that a Quinean practice works at least as well as a practice that allows for stipulative definitions to be permanently isolated from empirical refutation. On the other hand, if one can show that the practice of evidential isolation is even coherent, then this leaves open the possibility of so-called analytic theories of mathematical and other apparently a priori knowledge. This in itself might count as a pragmatic (albeit "theoretical") advantage of adopting a practice that allows for analyticity. Whether such a program can be carried through in detail, however, remains to be seen.

Cory Juhl
Eric Loomis

The authors acknowledge the helpful input of Samet Bagce, Hans-Johan Glock, Sahotra Sarkar, and David Sosa.

References

Ayer, Alfred J. (1946), *Language, Truth and Logic.* New York: Dover Publications.

Boghossian, Paul (1996), "Analyticity Reconsidered," *Noûs* 30: 360–391.

Bolzano, Bernhard (1973), *Theory of Science.* Translated by B. Terrell. Dordrecht, Netherlands: D. Reidel.

BonJour, Laurence (1998), *In Defense of Pure Reason.* Cambridge: Cambridge University Press.

Carroll, Lewis (1895), "What the Tortoise Said to Achilles," *Mind* 4: 278–280.

Carnap, Rudolf (1937), *The Logical Syntax of Language.* Translated by Amethe Smeaton. London: Routledge and Kegan Paul.

——— (1956), *Meaning and Necessity: A Study in Semantics and Modal Logic.* Chicago: University of Chicago Press.

——— (1990), "Quine on Analyticity," in R. Creath (ed.), *Dear Carnap, Dear Van: The Quine–Carnap Correspondence and Related Work.* Berkeley and Los Angeles: University of California Press, 427–432.

De Jong, Willem R. (1995), "Kant's Analytic Judgments and the Traditional Theory of Concepts," *Journal of History of Philosophy* 33: 613–641.

Frege, Gottlob (1974), *The Foundations of Arithmetic: A Logico-Mathematical Inquiry into the Concept of Number.* Translated by J. L. Austin. Oxford: Blackwell.

Friedman, Michael (1987), "Carnap's *Aufbau* Reconsidered," *Noûs* 21: 521–545.

Glock, Hans-Johann (1992), "Wittgenstein vs. Quine on Logical Necessity," in Souren Tehgrarian and Anthony Serafini (eds.), *Wittgenstein and Contemporary Philosophy.* Wakefield, NH: Longwood Academic, 154–186.

Grice, H. P., and P. F. Strawson (1956), "In Defense of a Dogma," *Philosophical Review* LXV: 141–158.

Harman, Gilbert (1967), "Quine on Meaning and Existence I," *Review of Metaphysics* 21: 124–151.

——— (1996), "Analyticity Regained?" *Nous* 30: 392–400.

Hintikka, Jaakko (1974), *Knowledge and the Known: Historical Perspectives in Epistemology.* Dordrecht, Netherlands: D. Reidel.

Kant, Immanuel (1965), *Critique of Pure Reason.* Translated by N. K. Smith. New York: St. Martin's Press.

——— (1974), *Logic.* Translated by R. S. Harriman and W. Schwartz. Indianapolis: Bobbs-Merrill.

Leibniz, Gottfried W. F. (1981), *New Essays Concerning Human Understanding.* Translated by P. Remnant and J. Bennett. New York: Cambridge University Press.

Lycan, William (1991), "Definition in a Quinean World," in J. Fetzer and D. Shatz (eds.), *Definitions and Definability: Philosophical Perspectives.* Dordrecht, Netherlands: Kluwer Academic Publishers, 111–131.

Martin, R. M. (1952), "On 'Analytic,'" *Philosophical Studies* 3: 65–73.

Mates, Benson (1951), "Analytic Sentences," *Philosophical Review* 60: 525–534.

Proust, Joelle (1986), *Questions of Form: Logic and the Analytic Proposition from Kant to Carnap.* Translated by A. A. Brenner. Minneapolis: University of Minnesota Press.

——— (1975), "The Analytic and the Synthetic," in *Philosophical Papers II: Mind, Language, and Reality*. Cambridge: Cambridge University Press, 33–69.

Quine, Willard Van (1953a), "Two Dogmas of Empiricism," in *From a Logical Point of View.* Cambridge: Harvard University Press, 21–46.

——— (1953b), "Reference and Modality," in *From a Logical Point of View.* Cambridge: Harvard University Press, 139–159.

Winnie, John (1975), "Theoretical Analyticity," in Jaakko Hintikka (ed.), *Rudolph Carnap, Logical Empiricist.* Dordrecht, Netherlands: D. Reidel, 149–159.

Wright, Crispin (1980), *Wittgenstein on the Foundations of Mathematics.* London: Duckworth.

See also **Carnap, Rudolf; Conventionalism; Explication; Logical Empiricism; Quine, Willard Van; Verificationism; Vienna Circle**

ANTHROPIC PRINCIPLE

An anthropic principle is a statement that there is a relation between the existence and nature of human life and the character of the physical universe. In the early 1970s, cosmologist Brandon Carter gave this name to a new family of principles, some of which he was also the first to formulate (Carter 1974; McMullin 1993). Cosmologists discussed these principles in response to the fine-tuning for life discovered in describing the evolution of the universe, and later also in the parameters of the Standard Model of particle physics (Collins and Hawking 1973; Barrow and Tipler 1986; Roush 2003). A universe is fine-tuned for *X* whenever (1) its parameters have values allowing for *X* in that universe, (2) that universe would not have (or be) *X* if the values of those parameters were slightly different, and (3) *X* is a significant, gross, or qualitative feature. (The values of the parameters in most cases currently have no explanation.) To say that a universe is fine-tuned in this way implies nothing about how the universe got the way it is—that an intelligent being tuned it would be an additional claim. That a universe is fine-tuned means only that key properties depend sensitively on the values of parameters.

The myriad examples of fine-tuning are often expressed in the form of counterfactual statements, for example:

1. If the ratio of the mass of the electron and the mass of the neutron were slightly different, then there would be no stable nuclei.
2. If the ratio of the strong to the electrical force were slightly different, then there would be few stable nuclei.
3. If any of the parameters affecting gravity were different, then the universe would have either lasted too short a time for galaxies to evolve or expanded too fast to allow clumping of matter into objects.

Since if there were no stable nuclei or material objects there would not be human life, such statements as these imply that if the universe were slightly otherwise than it is physically in any of a number of ways, then there would be no human beings to contemplate the fact.

Some, even if they would not put it in these terms, see this relationship as a spooky indication that human life was "meant to be," that the universe came to have the parameter values it has in order to make possible the existence of human beings, that the existence of human beings was a necessity antecedent to the determination of the physical features of the universe. These are the ideas behind the strong anthropic principle (SAP), which says that the universe must be such as to contain life (or human beings) in it at some point in its history. One must wonder, though, whether one should draw the same conclusion from the fact that human life depends on the existence of green plants, and the evolution of green plants depended on numerous contingencies. This analogy highlights the fact that the SAP class of responses to fine-tuning has involved arguments analogous to those of eighteenth-century models for the existence of God. In those arguments, the observation in the natural and especially biological world of an improbably well functioning phenomenon was supposed to make the existence of a designer-God more probable.

It is true that the physical universe is improbable on the assumption that its parameters were determined by direct chance, since fine-tuning implies that only a tiny proportion of possible universes possess the basic features in question; chance was unlikely to find them. (It is a matter of philosophical debate whether that makes the existence of God, or in some views multiple universes, more probable. See e.g., Earman 1987; Hacking 1987; White 2000.) However, direct chance is not the only possible hypothesis for explaining the values of parameters. Just as Darwin's concept of natural selection showed how it was possible for well-adapted organisms to arise without antecedent design, the next better physical theory may explain the values of the parameters in a way that makes no reference to a goal of providing for the existence of life (see Natural Selection). In the history of physics, there is precedent for a newer, deeper theory explaining previously unexplained values of parameters, as when, for example, the ideal gas constant, determined merely from observation in classical thermodynamics, got derived and explained in statistical thermodynamics. Indeed, physicists intend the next better theory to explain the values of the currently adjustable parameters of the Standard Model, and most see fine-tuning as an indication that a better physical theory than the Standard Model is needed. This is in keeping with a tradition in physics for eschewing frankly teleological explanation.

While SAP responses to fine-tuning try to use the fact that some basic physical characteristics of the universe are improbably fine-tuned to support a positive (and rather grand) thesis about the existence of a designer or of many universes, the weak anthropic principle (WAP) draws from the fact that for human-type observers, certain features of the physical universe are necessary in order for there to be a negative, or cautionary, lesson about what is presently known. The WAP says that what is observed may be restricted by the conditions necessary for the presence of humans as observers: For example, since humans are beings who could not have evolved in any place that lacked galaxies, there is reason to think that casual observation of the neighborhood humans find themselves in is not going to yield a representative sample of the cosmos with respect to the existence of galaxies. If there were regions of the universe that lacked galaxies, then humans probably would not have found them with this method of observation. Thus, instances of galaxies observed in this way would support the generalization that there are galaxies everywhere in the universe to a lesser degree than one would have expected if the instances had been a fair sample.

This type of argument, which found its first use by R. H. Dicke (1961) against the evidence P. A. M. Dirac (1937, 1938, 1961) had presented for a speculative cosmological hypothesis, is obviously an inference about bias and is generally endorsed by physicists and philosophers alike. There are disputes, however, about how best to model inferences about bias in evidence that is introduced by the method or process of obtaining the evidence. Some Bayesians think that the process through which evidence is produced is irrelevant to the degree to which the evidence supports a hypothesis and so, presumably, must reject the WAP style of argument (see Bayesianism). An error statistician will entirely approve of inferences about bias that exploit counterfactual dependencies that reveal the presence of error in the procedures used. Some Bayesians think that the bias introduced by the process of gathering evidence should be acknowledged and that this is best done by including an account of the process in the total evidence, a strategy whose rationale would be the principle of total evidence. Other Bayesian strategies for modeling WAP argumentation as legitimate may founder on the fact that the only analogy between WAP cases and ordinary cases of bias produced by evidence-gathering procedures lies in counterfactual statements that Bayesians resist incorporating into reasoning about evidence, such as: If there were regions without galaxies, then this method of observing would not have discovered them. These counterfactuals are not supported by the causal process of gathering the evidence but by background assumptions concerning the general possibility of that process.

Carter (1974) and others have claimed that anthropic principles represent a reaction against overwrought submission to the lesson learned from Copernicus when he proposed that the Earth was not at the center of the cosmos, or in other words that human beings were not special in the universe. While this diagnosis is clearly correct for the SAP, which does contemplate the specialness of human beings, it is misleading about the WAP. Indeed, reasoning that is associated with the WAP could be said to follow Copernicus, since it is analogous to the inference Copernicus made when he considered that the heavenly appearances might be the same regardless of whether the Earth were stationary and the stars moving or the Earth moving and the stars at rest. He inferred that evidence—generated by standing on the Earth—was inconclusive and therefore that the hypothesis that the Earth moves should not be scorned (Roush 2003).

SAP and WAP are misnamed with respect to each other, since WAP is not a weak form of SAP. WAP is not “weakly” anti-Copernican where SAP is “strongly” anti-Copernican, because WAP is not anti-Copernican at all. Also, WAP does not provide a way of weakly using the facts of fine-tuning to infer a speculative hypothesis about the universe, nor even does it use the facts of fine-tuning to infer a weaker speculative hypothesis; it provides a way of blocking such inferences. For these reasons, it would be better to refer to the SAP as the *metaphysical anthropic principle* (MAP), and to the WAP as the *epistemic anthropic principle* (EAP).

SHERRILYN ROUSH

References

Barrow, J. D., and F. J. Tipler (1986), *The Anthropic Cosmological Principle.* Oxford: Oxford University Press.

Carter, B. (1974), “Large Number Coincidences and the Anthropic Principle in Cosmology,” in M. S. Longair (ed.), *Confrontation of Cosmological Theories with Observational Data.* Boston: D. Reidel Co., 291–298.

Collins, C. B., and S. W. Hawking (1973), “Why Is the Universe Isotropic?” *Astrophysical Journal* 180: 317–334.

Dicke, R. H. (1961), “Dirac’s Cosmology and Mach’s Principle,” *Nature* 192: 440–441.

Dirac, P. A. M. (1937), “The Cosmological Constants,” *Nature* 139: 323.

——— (1938), “A New Basis for Cosmology,” *Proceedings of the Royal Society of London,* Series A 165: 199–208.

——— (1961), Response to “Dirac’s Cosmology and Mach’s Principle,” *Nature* 192: 441.

Earman, John (1987), “The SAP Also Rises: A Critical Examination of the Anthropic Principle,” *American Philosophical Quarterly* 24: 307–317.

Hacking, Ian (1987), “The Inverse Gambler’s Fallacy: The Argument from Design: The Anthropic Principle Applied to Wheeler Universes,” *Mind* 96: 331–340.

McMullin, Ernan (1993), “Indifference Principle and Anthropic Principle in Cosmology,” *Studies in the History and Philosophy of Science* 24: 359–389.

Roush, Sherrilyn (2003), “Copernicus, Kant, and the Anthropic Cosmological Principles,” *Studies in the History and Philosophy of Modern Physics* 34: 5–36.

White, Roger (2000), “Fine-Tuning and Multiple Universes,” *Nous* 34: 260–276.

See also **Particle Physics**

ANTI-REALISM

See **Instrumentalism; Realism; Social Constructionism**

APPROXIMATION

An approximation refers to either a *state* of being a nearly correct representation or the *process* of producing a nearly correct value. In the former sense a planet's orbit approximates an ellipse, while in the latter sense one approximates by ignoring terms or factors in order to facilitate computation. Examples of approximations in both senses are plentiful in the mathematically oriented physical, biological, and engineering sciences, including:

- Attempts to solve the general relativistic field equations without the Schwarzchild assumption of a nonrotating, spherically symmetric body of matter;
- The generalized three-body problem;
- Many problems in fluid mechanics;
- Attempts to solve the Schrödinger wave equation for atoms and molecules other than hydrogen;
- Multilocus/multiallele problems in population genetics and evolutionary biology; and
- Problems associated with the description of ecosystems and their processes.

Scientists often employ approximations to bypass the computational intractability or analytical complexity involved in the description of a problem. This can occur in two ways, which are often combined in multistep problem solutions. For example, in solving a three-body problem, one can write down the complete, 18-variable equation first and make approximations as the solution is generated. Here, an approximate solution to an exact equation is generated. Alternatively, one can assume that two of the three bodies dominate the interaction, treating the third body as a small perturbation on the other two bodies' motion. Here, the process of making the approximation is completed before the solution is attempted in order to generate an exact solution to an approximate equation. In this second sense, approximations resemble the processes of idealization and abstraction quite strongly. It is probably pointless to insist on a strict division among the three strategies, since many approximations are justified not by purely mathematical arguments but by arguments that some variables are unimportant for the particular class of problems being investigated. However, roughly speaking, one can say that approximations (of both types) are invoked upon contemplation of the complete description of the problem, whereas idealizations and abstractions are deliberate attempts to study only part of the complete problem.

Until relatively recently, philosophers have tended to view approximations as uninteresting or ephemeral categories of scientific activity. John Stuart Mill (1874, 416) claimed that "approximate generalizations" (Most *A*s are *B*s) were of little use in science "except as a stage on the road to something better." Further, philosophers have focused almost entirely on the *state* sense defined above. For Mill as well as the twentieth-century logical empiricists, this was a result of their focus on logic to the exclusion of the mathematical difficulties of solving equations. Discussions of "nearly correct" results are missing in Mill's analysis and also in those of many of the early logical positivists. Eventually, when it was pointed out that, for example, Kepler's laws were not deducible from Newtonian mechanics and gravitational theory (since Kepler's laws assert that the orbit of a planet is elliptical, but according to Newtonian theory a planet's orbit around the Sun will not be exactly elliptical due to the influence of the other planets), Hempel (1966) responded that there is a law entailed by Newtonian theory that Kepler's laws approximate. Similar remarks were made regarding the explanation of individual events (as opposed to laws). Schaffner (1967) provided a number of examples of approximate explanation in the context of the reduction of one theory or law to another. Subsequently, structuralist philosophers of science have developed a theory that assesses the amount of approximation between any two statements in a given language (Balzer, Ulises-Moulines, and Sneed 1987, Chap. 6). However, while scientifically important and interesting, discussion of the state sense of approximation appears philosophically limited. One can say that a value is or is not "close enough" to count as a correct prediction or explanation, but the question as to how close is close enough must be decided on pragmatic grounds supplied by the science available at the time. More importantly, it is difficult to address whether the process that produced the "nearly correct" value is justifiable.

Recently, more attention has been directed toward the process of making approximations. Analyses have shown that approximations are tied to a number of methodological and interpretive issues. Perhaps most basically, approximations raise the question of whether a theory actually accounts for a given datum, phenomenon, law, or other theory. For example, "the fact remains that the exponential decay law, for which there is so much empirical support in radioactive processes, is not a rigorous consequence of quantum mechanics but the result of somewhat delicate approximations" (Merzbacher, quoted in Cartwright 1983, 113). In the absence of more robust derivations from the underlying theory, one can quite seriously ask what kind of understanding one has of the phenomenon in question.

In addition, approximations sometimes produce intertheoretic relations. Philosophers have traditionally assumed that approximations frustrate intertheoretic reduction because they produce gaps in the derivation of one theory, law, or concept from another. Woody (2000) has analyzed the approximations involved in producing molecular orbital theory as a case of the approximations producing the intertheoretic connections. She notes that the complete, nonapproximate *ab initio* solution of the Schrödinger wave equation for molecules is rarely ever solved for because it leads to a solution in which the semantic information prized by chemists cannot be recovered. (Importantly, even this nonapproximate solution is based on an approximate representation of the molecule in the idealized Hamiltonian.) That is, it leads to a solution in which the information about bonds and other measurable properties is distributed throughout the terms of the solution and is not localized so that it can be recognized and connected with any measurable properties. In order to recover the chemical properties, only one step in the complete set of calculations is performed. Specifically, the spatial portion of the basis sets of the calculation are integrated to give atomic orbitals, the spherical and lobed graphical representations familiar from chemistry texts. So this involves an approximation to the full solution. The upshot is that the quantum mechanical description (or at least some portion of it) is connected with familiar chemical ideas via only an approximation. Without the approximation, the intertheoretic connection just does not appear.

The process of making approximations also sometimes creates terms that are given subsequent physical interpretations and sometimes eliminates terms that are later recovered and given interpretations. Questions then arise whether the entities that are created and eliminated are artifactual or not. The case of the atomic orbitals illustrates the issues surrounding the creation of entities. If one does not make the approximations, there is some justification for saying that such orbitals do not "exist" outside the model. This is because the complete solution produces orbitals that are not localized in the way the approximate solution pictures (cf. Scerri 1991). Other examples include (1) the construction of rigid, three-dimensional molecules via the Born-Oppenheimer approximation that treats nuclear and electronic motions separately; (2) potential energy surfaces in various physical sciences that arise due to the assumption that various kinds of energy functions are isolable from each other; and (3) pictures of fitness landscapes in evolutionary biology that give the impression that fitness values are static across wide ranges of environments.

Cartwright's (1983) discussion of the derivation of the exponential decay law and the discovery of the Lamb shift is a classic example of how approximations can eliminate terms that should be given a physical interpretation in some circumstances.

Typically, one is considering an excited atom in an electromagnetic field. The question is how to describe the process of decay from an excited to an unexcited state. The experimental data speak for an exponential law; the issue is to try to reproduce this law solely on theoretic grounds. (As yet, it cannot be done.) In one derivation of the decay law, scientists integrate certain slowly varying frequencies of the field first and the amplitude of the excited state second. The result is an equation that relates the amplitude of the excited state to the magnitude of the line broadening only. However, if the integrations are reversed (which can be done because the amplitude of the excited state is slowly varying compared with its rapid oscillations with the field), the resulting equation relates the amplitude to the line broadening and an additional term that describes a small displacement in the energy levels. That additional term is the Lamb shift, discovered by Willis Lamb and R. C. Retherford (1947). Thus, in this case the process of approximation affects what is believed to be going on in the world.

These interpretive issues give rise to a number of epistemological and methodological questions, especially with regard to the justifiability and the purpose of the approximation. As indicated by Merzbacher, noted above, approximations produce tenuous connections, so that the claim that a theory accounts for a given law or phenomenon is to some degree questionable. In such circumstances, how should one decide that the approximations are justified? Clearly, one important step toward answering this question is to delineate the features of the approximation that is being made. To this end, one can distinguish whether the approximation (1) is implicit or explicit; (2) is corrigible, incorrigible in practice, or incorrigible in principle; (3) has effects that may be estimable, not estimable in practice, or not estimable in principle; (4) is justified mathematically, with respect to some more foundational theory, a combination of both, or neither; and (5) is context dependent or independent and involves counterfactual assumptions (as when Galileo assumed what would happen to falling bodies in a vacuum). (For further discussion of these issues within the specific context of the reduction of one theory to another, see Sarkar 1998, chap. 3.) Once it is known how the approximation falls out with respect to these distinctions, one can begin to assess whether the approximation is warranted.

In addition to assessing whether an approximation is mathematically justified or justified with respect to some more fundamental theory, scientists typically assess approximations according to whether they track the set of causal distinctions drawn in the theory, whether the approximation works across a wide range of cases, and whether the approximation allows a plausible or coherent interpretation of the available data to be constructed (Laymon 1989, 1985; Ramsey 1990). Whether these considerations are exhaustive and whether any one or more of them is more basic remains unclear.

As a final methodological issue, consider the question of the purpose of the approximation. The examples thus far illustrate the purpose of producing a prediction or explanation that is superior to the one that can be given without the approximation. Yet, some analyses of approximations are post hoc justifications of idealizations already in use. Often, this happens by embedding the approximate analysis within a more complex theory so that the idealization can be interpreted as what results when certain terms are eliminated (Laymon 1991). Given different purposes for approximations, it is reasonable to expect that different kinds of justifications will be acceptable in the two situations. Whether this is so remains unanswered at present.

What has been said in this article probably represents only a small subset of the philosophical issues surrounding the use of approximations in the sciences. Given the ubiquity of approximations in many sciences and their varied modes of justification, much philosophical work remains to be completed.

JEFFRY L. RAMSEY

References

Balzer, W., C. Ulises-Moulines, and J. Sneed (1987), *An Architectonic for Science: The Structuralist Program.* Boston: D. Reidel.

Cartwright, N. (1983), *How the Laws of Physics Lie.* New York: Oxford University Press.

Hempel, C. (1966), *Philosophy of Natural Science.* Englewood Cliffs, NJ: Prentice-Hall.

Lamb, W., and Retherford, R. C. (1947), "Fine Structure of the Hydrogen Atom," *Physical Review* 72: 241.

Mill, J. S. (1874), *System of Logic,* 8th ed. New York: Harper and Brothers.

Laymon, R. (1989), "Applying Idealized Scientific Theories to Engineering," *Synthese* 81: 353–371.

——— (1985), "Idealizations and the Testing of Theories by Experimentation," in P. Achinstein and O. Hannaway (eds.), *Observation, Experiment and Hypothesis in Modern Physical Science.* Cambridge, MA: MIT Press, 147–173.

——— (1991), "Computer Simulations, Idealizations and Approximations," *Philosophy of Science* 2 (Proceedings): 519–534.

Ramsey (1990), "Beyond Numerical and Causal Accuracy: Expanding the Set of Justificational Criteria," *Philosophy of Science* 1 (Proceedings): 485–499.

Sarkar (1998), *Genetics and Reductionism.* New York: Cambridge University Press.

Schaffner, K. (1967), "Approaches to Reduction," *Philosophy of Science* 34: 137–147.

Scerri, E. (1991), "The Electronic Configuration Model, Quantum Mechanics and Reduction," *British Journal for the Philosophy of Science* 42: 309–325.

Woody, A. (2000), "Putting Quantum Mechanics to Work in Chemistry: The Power of Diagrammatic Representation," *Philosophy of Science* 67: S612–S627.

ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) aims at building machines that exhibit intelligent behaviors, understood as behaviors that normally require intelligence in humans. Examples include using language, solving puzzles, and playing board games. AI is also concerned with reproducing the prerequisites for those activities, such as pattern recognition and motor control. AI has typically been pursued using only computing machines. It has been influenced by the sciences of mind and brain as well as philosophical discussions about them, which it has influenced in turn. AI is a successful discipline, with many applications, ranging from guiding robots to teaching.

Origins

In 1936, Alan Turing offered a rigorous definition of computable functions in terms of a class of computing machines—now called Turing Machines—and argued persuasively that his machines could perform any computation (Turing [1936–7] 1965) (see Turing, Alan). In 1943, Warren McCulloch and Walter Pitts published their computational theory of mind, according to which mental processes are realized by neural computations (McCulloch and Pitts 1943). In the late 1940s, the first stored-program computers were built (von Neumann 1945). If Turing's thesis is correct, stored-program computers can perform any computation (until they run out of memory) and can reproduce mental processes. And whether or not mental processes are computational, it might be possible to program computers to exhibit intelligent behavior (Turing 1950).

Starting in the late 1940s, researchers attempted to put those ideas into practice. Some designed computing machines formed by networks of interactive units, which were intended to reproduce mental processes by mimicking neural mechanisms (Anderson and Rosenfeld 1988). This was the beginning of connectionism. Other researchers wrote computer programs that performed intelligent activities, such as playing checkers and proving theorems (Feigenbaum and Feldman 1963). Finally, some researchers built whole artificial agents, or robots. Sometimes the label "AI" is reserved for the writing of intelligent computer programs, but usually it also applies to connectionism as well as the construction of robots that exhibit intelligent behaviors (see Connectionism).

Artificial Intelligence and Computation

Although there is no room in this article for an overview of the field, this section introduces some fundamental concepts in AI (for further reference, standard AI textbooks include Russell and Norvig 1995 and Nilsson 1998).

Computation deals with strings of symbols. A symbol is a particular (e.g., *a*, *b*, *c*) that falls under a type (e.g., a letter of the English alphabet) and can be concatenated with other symbols to form strings (e.g., *abc*, *cba*, *bccaac*). Symbols are discrete in the sense that they fall under only finitely many types.

A computation is a mechanical process that generates new strings of symbols (outputs) from old strings of symbols (inputs plus strings held in memory) according to a fixed rule that applies to all strings and depends only on the old strings for its application. For example, a computational problem may be, for any string, to put all of the string's symbols in alphabetical order. The initial string of symbols is the input, whereas the alphabetized string of symbols is the desired output. The computation is the process by which any input is manipulated (sometimes together with strings in memory) to solve the computational problem for that input, yielding the desired output as a result. This is digital computation. There is also analog computation, in which real variables are manipulated. Real variables are magnitudes that are assumed to change over real time and take values of an uncountable number of types. In digital computation, accuracy can be increased by using longer strings of symbols, which are designed to be measured reliably. By contrast, accuracy in analog computation can be increased only by measuring real variables more precisely. But measuring real variables is always subject to error. Therefore, analog computation is not as flexible and precise as digital computation, and it has not found many AI applications.

Ordinary computations—like arithmetic additions—are sequences of formal operations on strings of symbols, i.e., operations that depend on only the symbols' types and how the symbols are concatenated. Examples of formal operations include deleting the first symbol in a string, appending a symbol to a string, and making a copy of a string. In order to define a computation that solves a computational problem (e.g., to put any string in alphabetical order), one must specify a sequence of formal operations that are guaranteed to generate the appropriate outcome (the alphabetized string) no matter what the input is. Such a specification is called an *algorithm.* (For an alternative computing scheme, see Connectionism. Connectionist computations manipulate strings of symbols based on which units of a connectionist network are connected to which, as well as their connection strengths.)

Executing algorithms takes both a number of steps and some storage space to hold intermediate results. The time and space—the resources—required to solve a computational problem with a specific algorithm grows with the size of the input and of the strings in memory: The same computation on a bigger input requires more steps and more storage space than on a smaller input. The rate at which needed resources grow is called computational *complexity.* Some computations are relatively simple: The required time and storage space grow linearly with the size of the input and the strings in memory. Many interesting computations are more complex: The required resources may grow as fast as the size of the input and strings in memory raised to some power. Other computations, including many AI computations, are prohibitively complex: Needed resources grow exponentially with the size of the input and strings in memory, so that even very large and fast computing machines may not have enough resources to complete computations on inputs and stored strings of moderate size.

When all the known algorithms for a computational problem have high complexity, or when no algorithm is known, it may be possible to specify sequences of operations that are not guaranteed to solve that computational problem for every input but will use a feasible amount of resources. These procedures, called *heuristics,* search for the desired output but they may or may not find it, or they may find an output that only approximates the desired result. Since most AI computations are very complex, AI makes heavy use of heuristics (Newell and Simon 1976).

Symbols have their name because they can be interpreted. For instance, the symbol '/' may represent the letter *a* or the number 1. A string of symbols may be interpreted in a way that depends on its component symbols and their concatenation. For example, assume that '/' represents 1. Then, under different interpretations, the string '///' may represent the number 3 (in unary notation), 7 (in binary notation), 111 (in decimal notation), etc. Interpreted strings of symbols are often called *representations.*

Much of AI is concerned with finding effective representations for the information—AI researchers call it "knowledge"—that intelligent agents (whether natural or artificial) are presumed to possess. If an artificial system has to behave intelligently, it needs to respond to its environment in an adaptive way. Because of this, most AI systems are guided by internal states that are interpreted as representations of their environment. Finding efficient ways to represent the environment is a difficult problem, but an even harder problem—known as the frame problem—is finding efficient ways to update representations in the face of environmental change. Some authors have argued that the frame problem is part of the more general problem of getting machines to learn from their environment, which many see as something that needs to be done in order to build intelligent machines. Much AI research and philosophical discussion have been devoted to these problems (Ford and Pylyshyn 1996).

Given an interpretation, a string of symbols may represent a formal operation on strings, in which case the string is called an *instruction* (e.g., "write an *a*"). A sequence of instructions, representing a sequence of operations, is called a *program* (e.g., "[1] append an *a* to the string in register 0025; [2] if register 0034 contains *ccccc*, stop computing; [3] append a *c* to the string in register 0034; [4] go back to instruction [1]"). Given a domain of objects, such as numbers, it may be useful to derive some new information about those objects, for instance their square roots. Given inputs representing numbers under some notation, there may be an algorithm or heuristic that generates outputs representing the square roots of the numbers represented by the inputs. That algorithm or heuristic can then be encoded into a program, and the computation generated by the program can be interpreted as operating on numbers. So, relative to a task defined over a domain of objects encoded by a notation, a program operating on that notation may represent an algorithm, a heuristic, or just any (perhaps useless) sequence of formal operations.

A stored-program computer is a machine designed to respond to instructions held in its memory by executing certain primitive operations on inputs and other strings held in memory, with the effect that certain outputs are produced. Because of this, a computer's instructions are naturally interpreted as representing the operations performed by the computer in response to them. Most stored-program computers execute only one instruction at a time (serial computers), but some execute many instructions at a time (parallel computers). All stored-program computers execute their instructions in one step, which requires the processing of many symbols at the same time. In this sense, most computers are parallel (Turing Machines are an exception), which is also the sense in which connectionist systems are parallel. Therefore, contrary to what is often implied, parallelism is not what distinguishes connectionist systems from stored-program computers.

Given that stored-program computers can compute any computable function (until they run out of memory) and that computations can be defined over interpreted strings of symbols, stored-program computers are very flexible tools for scientific investigation. Given strings of symbols interpreted as representations of a phenomenon, and a series of formal operations for manipulating those representations, computers can be used to model the phenomenon under investigation by computing its representations. In this way, computers have become a powerful tool for scientific modeling.

Typically, AI is the programming of stored-program computers to execute computations whose outputs are interpretable as intelligent behavior. Usually, these AI computations involve the manipulation of strings of symbols held in memory, which are interpreted as representations of the environment. Some AI research is devoted to building complete artificial agents (robots), which are usually guided by an internal computing machine hooked up to sensors and motor mechanisms. Some roboticists have downplayed the importance of representation, attempting to develop a nonrepresentational framework for robotics (Brooks 1997).

Philosophical Issues

The remainder of this article describes some of the debates within and about AI that are likely to interest philosophers of science, and some of the relations between the issues that arise therein.

Engineering Versus Science

Some say that AI is a mathematical discipline, concerned with formalisms for representing information and techniques for processing it. Others say it is a branch of engineering, aimed at constructing intelligent machines. Still others say it is a *bona fide* empirical science, whose subject matter is intelligent behavior by natural and artificial agents (Newell and Simon 1976) and whose experimental method consists of building, testing, and analyzing AI artifacts (Buchanan 1988). A few have argued that AI is a form of philosophical investigation that turns philosophical explications into computer programs (Glymour 1988) and whose main method is a priori task analysis (Dennett 1978).

These views need not be incompatible. As in any other science, work in AI can be more theoretical, more experimental, or more driven by applications, and it can be conducted following different styles. At least three kinds of AI research can be usefully distinguished (Bundy 1990):

- *Applied AI* builds products that display intelligent behavior. It is a form of software engineering that uses AI techniques.
- *Cognitive simulation* develops and applies AI techniques to model human or animal intelligence. It is constrained by the experimental results of psychology and neuroscience and is part of the science of mind.
- *Basic AI* develops and studies techniques with the potential to simulate intelligent behavior. In developing these techniques, different researchers follow different styles. Some proceed more a priori, by task analysis; others proceed more *a posteriori,* by looking at how humans solve problems. All are constrained by mathematical results in computability and complexity theory. Basic AI is a largely formal or mathematical discipline, but it often proceeds by exploring the capabilities of artifacts that embody certain formalisms and techniques. In obtaining results, it might be unfeasible to prove results about AI systems and techniques mathematically or by a priori argument. The only practical way to evaluate a design might be to build the system and see how it performs under various measures. In

> this respect, basic AI is experimental, though more in the sense in which engineering is experimental than in the sense in which experimental physics is.

To the extent that AI theories say what intelligent behaviors can be obtained by what means, they are relevant to our scientific and philosophical understanding of mind and intelligent behavior.

Strong Versus Weak

Strong AI holds that a computing machine with the appropriate functional organization (e.g., a stored-program computer with the appropriate program) has a mind that perceives, thinks, and intends like a human mind. Strong AI is often predicated on the computational theory of mind (CTM), which says that mental processes are computations (McCulloch and Pitts 1943; Putnam 1967; Newell and Simon 1976). Alternatively, strong AI can be grounded in instrumentalism about mentalistic language, according to which ascribing mental features to agents is not a matter of discovering some internal process but of using mentalistic predicates in a convenient way (Dennett 1978; McCarthy 1979) (see Instrumentalism).

Those who believe that ascribing genuine mental features is a matter of discovery, as opposed to interpretation, and that mental processes are not computations reject strong AI (Searle 1980). They agree that AI has useful applications, including computational models of the mind and brain, but they submit that these models have no genuine mental features, any more than computational models of other natural phenomena (e.g., the weather) have the properties of the systems they model (e.g., being humid or windy). Their view is called weak AI.

Some supporters of strong AI have replied that although performing computations may be insufficient for having a mind, performing the appropriate computations plus some other condition, such as being hooked up to the environment in appropriate ways, is enough for having at least some important mental features, such as intentionality (Pylyshyn 1984) and consciousness (Lycan 1987) (see Consciousness, Intentionality).

Hard Versus Soft

Not all attempts at writing intelligent computer programs purport to mimic human behavior. Some are based on techniques specifically developed by AI researchers in order to perform tasks that require intelligence in humans. Research in this tradition starts with a task, for example playing chess, and analyzes it into subtasks for which computational techniques can be developed. This approach is sometimes called hard AI, because it attempts to achieve results without regard for how humans and animals generate their behavior (McCarthy 1960).

Another tradition, sometimes called soft AI, attempts to build intelligent machines by mimicking the way humans and animals perform tasks. Soft AI—based on either computer programs or connectionist networks—is adopted by many cognitive psychologists as a modeling tool and is often seen as the basis for the interdisciplinary field of cognitive science (see Cognitive Science). In this guise, called *cognitive simulation,* it aims at mimicking human behavior as closely as possible. In contrast, hard AI does not aim at mimicking human behavior but at approaching, matching, and eventually outperforming humans at tasks that require intelligence.

Weak Versus Strong Equivalence

When two agents exhibit the same behavior under the same circumstances, they are weakly equivalent. When their identical behaviors are generated by the same internal processes, they are strongly equivalent (Fodor 1968). The distinction between weak and strong equivalence should not be confused with that between weak and strong AI. The latter distinction is about whether or not a machine weakly equivalent to a human has genuine mental features. If the answer is yes, then one may ask whether the machine is also strongly equivalent to a human. If the answer is no, then one may ask what the human has that the machine lacks, and whether a different kind of machine can have it too. The issue of strong equivalence takes a different form depending on the answer to the strong vs. weak AI question, but it arises in either case.

Those who originally developed the methodology of cognitive simulation endorsed CTM. They saw mental processes as computations and attempted to discover the computations performed by minds, sometimes by engaging in psychological research. When writing AI programs, they aimed at reproducing mental computations, that is, building machines that were strongly equivalent to humans.

Some authors think that strong equivalence is too much to hope for, because two agents whose behavior is empirically indistinguishable may nevertheless be performing different computations, and there is no empirical criterion for distinguishing between their internal processes (Anderson 1978). These authors still aim at cognitive simulation but see its purpose as to mimic intelligent behavior, without attempting to reproduce the internal processes generating that behavior in humans or animals. Their position is a form of instrumentalism about cognitive science theories, which aims at building machines weakly equivalent to humans and animals.

Those who take a realist position about cognitive scientific theories think that weak equivalence is an unsatisfactory goal for a science of mind. They argue that by reproducing in machines certain aspects of human behavior, such as the temporal duration of the process and the patterns of error, it should be possible to use cognitive simulation to study human internal computational processes, thereby striving for strong equivalence (Pylyshyn 1984).

Programs, Models, Theories, and Levels

Within cognitive simulation, there has been a dispute about the relationship between AI artifacts, models, and theories of intelligent behaviors. Computational models are common in many sciences, for instance in meteorology. Usually, these models are computations driven by programs and interpreted to represent some aspect of a phenomenon (e.g., the weather). The fact that the models perform computations is not modeling anything—computing is not a feature of the phenomenon but rather the means by which the model generates successive representations of the phenomenon.

When a cognitive simulation mimics human or animal behavior, there are two ways in which it can be seen as a model. Those who lean toward weak AI believe that appropriately interpreted AI artifacts are models in the same sense as are computational models of the weather: The model computes representations of the phenomenon without the phenomenon being computational. In contrast, those who incline toward strong AI, especially if they also believe CTM, hold that AI models are different from computational models in other disciplines in that when a model is correct, the computations themselves model the computational process by which humans and animals generate their behavior.

In AI (and cognitive science), a theory of an intelligent behavior should be specific enough that it can guide the design of machines exhibiting that behavior. It may be formulated at different levels of abstraction and detail:

1. A theory may specify a task, possibly by describing the characteristics that, given certain inputs, the outputs should have. For example, vision may consist of generating three-dimensional representations of physical objects from retinal stimuli. This is sometimes called the *computational level* (Marr 1982).
2. A theory may specify how to perform a task by giving a finite set of rules, regardless of whether those rules are part of the procedure through which an agent performs the task. For example, a theory of syntax may consist of rules for generating and recognizing grammatical sentences, regardless of how humans actually process linguistic items. This is sometimes called the *competence level* (Chomsky 1965).
3. A theory may specify a procedure by which an agent performs a task. This may be an algorithm or heuristic that operates on representations. This is sometimes called the *algorithmic level* (Marr 1982).
4. A theory may specify a mechanism (usually a program) that implements the representations and algorithm or heuristic of level 3. This corresponds to the building of a computational model and its interpretation. (In connectionist AI, levels 3 and 4 are the same, corresponding to the design and interpretation of a connectionist system for performing the task.)
5. A theory may specify the architectural constraints, such as the size of the memory, and how they affect the behavior of human and animal subjects and their errors in the task being performed. This is sometimes called the *performance level* (Chomsky 1965).
6. A theory may specify the physical components of the system and their mutual functional relations. This is sometimes called the *implementation level* (Marr 1982). It can be further subdivided into sublevels (Newell 1980).

The distinctions discussed in this aricle have arisen within computational approaches to AI and the mind, but they are largely independent of computational assumptions.

Although theories at levels 1–3 are usually proposed within a computational framework, they are not necessarily committed to CTM, because they do not specify whether the implementing mechanism is computational. Theories at level 4 may be interpreted either as describing the computations by which brains compute or as using computations to describe noncomputational cognitive processes (analogously to programs for weather forecasting). Levels 5 and 6 arise regardless of whether the mind is computational.

The distinction between strong and weak AI applies to any attempt to build intelligent machines. The question is whether a machine that appears intelligent has a genuine mind, and how to find evidence one way or the other. AI is committed to either dismissing this question as meaningless or answering it by forming hypotheses about the mind and testing them empirically. What constitutes empirical testing in this domain remains controversial.

The distinction between approaches that attempt to mimic natural agents (soft AI) and those that do not (hard AI) applies to any means of reproducing intelligence, whether computational or not.

Finally, within any attempt to mimic human intelligence, instrumentalists aim only at simulating behavior (weak equivalence), while realists attempt to reproduce internal processes (strong equivalence).

Gualtiero Piccinini

References

Anderson, J. A., and E. Rosenfeld (eds.) (1988), *Neurocomputing: Foundations of Research.* Cambridge, MA: MIT Press.

Anderson, J. R. (1978), "Arguments Concerning Representations for Mental Imagery," *Psychological Review* 85: 249–277.

Brooks, R. A. (1997), "Intelligence without Representation," in J. Haugeland (ed.), *Mind Design II.* Cambridge, MA: MIT Press: 395–420.

Buchanan, B. G. (1988), "Artificial Intelligence as an Experimental Science," in J. H. Fetzer (ed.), *Aspects of Artificial Intelligence.* Dordrecht, Netherlands: Kluwer, 209–250.

Bundy, A. (1990), "What Kind of Field Is AI?" in D. Partridge and Y. Wilks (eds.), *The Foundations of Artificial Intelligence: A Sourcebook.* Cambridge: Cambridge University Press: 215–222.

Chomsky, N. (1965), *Aspects of a Theory of Syntax.* Cambridge, MA: MIT Press.

Dennett, D. C. (1978), *Brainstorms.* Cambridge, MA: MIT Press.

Glymour, C. (1988), "Artificial Intelligence Is Philosophy," in J. H. Fetzer (ed.), *Aspects of Artificial Intelligence.* Dordrecht, Netherlands: Kluwer: 195–207.

Feigenbaum, E. A., and J. Feldman (1963), *Computers and Thought.* New York: McGraw-Hill.

Fodor, J. A. (1968), *Psychological Explanation.* New York: Random House.

Ford, K. M., and Z. W. Pylyshyn (1996), *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence.* Norwood, NJ: Ablex.

Lycan, W. (1987), *Consciousness.* Cambridge, MA: MIT Press.

Marr, D. (1982), *Vision.* New York: Freeman.

McCarthy, J. (1960), "Programs with Common Sense," in *Mechanisation of Thought Processes* (Proceedings of the Symposium at the National Physical Laboratory) 1, 77–84.

——— (1979), "Ascribing Mental Qualities to Machines," in M. Ringle (ed.), *Philosophical Perspectives in Artificial Intelligence.* Atlantic Highlands, NJ: Humanities Press, 161–195.

McCulloch, W. S., and W. H. Pitts (1943), "A Logical Calculus of the Ideas Immanent in Nervous Nets," *Bulletin of Mathematical Biophysics* 7: 115–133.

Newell, A. (1980), "Physical Symbol Systems," *Cognitive Science* 4: 135–183.

Newell, A., and H. A. Simon (1976), "Computer Science as an Empirical Enquiry: Symbols and Search," *Communications of the ACM* 19: 113–126.

Nilsson, N. J. (1998), *Artificial Intelligence: A New Synthesis.* San Francisco: Morgan Kauffman.

Putnam, H. (1967), "Psychological Predicates," in W. H. Capitan and D. D. Merill (eds.), *Art, Mind, and Religion.* Pittsburgh, PA: University of Pittsburgh Press, 37–48.

Pylyshyn, Z. W. (1984), *Computation and Cognition.* Cambridge, MA: MIT Press.

Russell, S. J., and P. Norvig (eds.) (1995), *Artificial Intelligence: A Modern Approach.* Englewood Cliffs, NJ: Prentice Hall.

Searle, J. R. (1980), "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3: 417–457.

Turing, A. ([1936–7] 1965), "On Computable Numbers, with an Application to the Entscheidungs problem," in M. Davis (ed.), *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions.* Hewlett, NY: Raven Press.

——— (1950), "Computing Machinery and Intelligence," *Mind* 59: 433–460.

von Neumann, J. (1945), *First Draft of a Report on the EDVAC.* Philadelphia, PA: Moore School of Electrical Engineering, University of Pennsylvania.

See also **Chomsky, Noam; Cognitive Science; Connectionism; Consciousness; Instrumentalism; Putnam, Hilary; Realism; Searle, John; Turing, Alan**

PHILOSOPHY OF ASTRONOMY

It has been claimed that inside every astronomer lies the heart of a cosmologist (Hoskin 1997, 108). This may have well been true when the universe was thought to be relatively small and simple, but as the understanding of the size and complexity of the universe has enlarged, especially in the latter half of the twentieth century, so with it has grown the establishment of subject areas whose degree of specialization may touch nary at all on cosmology—a subject that too has matured into a field all its own. Therefore, a philosophy of astronomy cannot be content simply to track the historical development of cosmological thought (see Munitz 1957). One also finds astronomers currently who disavow any self-identification with cosmology, and at times also theoretical astrophysics, at least in part because of cosmology's highly complicated theoretical nature, and because of cosmology's relative isolation from technology. No comprehensive philosophical account of astronomy up to the present day yet exists, and such an analysis should distinguish astronomical work as such from cosmology and theoretical astrophysics, since many of the important developments in astronomy have occurred as a result of technological change, without a theory in place to understand the new phenomena astronomers and engineers uncovered. Consequently, this article will use an historical approach, distinguishing phases of astronomy's theoretical development through the technologies in use. (For more details of the historical material presented here, see Lankford 1997.)

The Naked Eye

Artifacts such as Ireland's 3000 B.C.E. Newgrange Passage Tomb and Mesoamerica's Monte Alban and Xochicalco zenith tubes are evidence that human beings have long been aware of regular celestial cycles. Records of the Sun's, moon's, and stars' movements were kept by the Babylonians (1800 B.C.E.–75 C.E.) and Egyptians (starting perhaps as early as 3000 B.C.E.), who recognized a connection between time and celestial phenomena. By the second century B.C.E. the Greeks, informed by these records, devised geometrical models to create a unified image of what the substance and motion of the heavenly bodies were like. Classical Western-style philosophizing about the nature of heavens starts with the works of Plato and Aristotle. By all appearances, the Earth is fixed, while the domed ceiling of the sky revolves around it. These basic impressions coupled with physical arguments on the nature of matter and motion (cf. particularly Aristotle's *De caelo* and *Metaphysics*) led to the long-standing view that the Earth sits unmoving at the center of the universe, while all the celestial bodies—the Sun, moon, planets, and fixed stars, perfectly spherical bodies composed of a special unearthly substance—travel about in circular motion, carried along by a system of revolving crystalline spheres. These ideas remained fundamental to astronomy for nearly 2,000 years, ever modified to accommodate data gathered by the naked eye using a set of instruments that dominated the science for that entire period. The armillary sphere, a device for measuring position, was used by Eratosthenes around 204 B.C.E. and Hipparchus of Rhodes circa 150–125 B.C.E. The cross-staff, used in Alexandria around 284 B.C.E., measures separation between celestial objects. The quadrant and astrolabe, invented by Ptolemy circa 150 C.E., also measure position. With a fair degree of accuracy, one could chart the meanderings of the wandering stars, or planets, as they rise and fall in elevation, growing brighter and dimmer as their motion appears to accelerate, decelerate, stop, and go in reverse. In an Aristotelian view, one would adjust the number of spheres, each contributing its own element of motion, to explain planetary motion retroactively. Alternatively, around 200 B.C.E., Hipparchus proposed that the Earth was in the middle of the universe but slightly off-center, so that the stars, moon, and Sun traveled around it in eccentric orbits. Thus, depending upon whether the stars and Sun were closer or farther away from the Earth, the planets would appear brighter or dimmer, and the seasons would be longer or shorter

Ptolemy's *Almagest* ([137 C.E.] 1984) built upon the model of the universe Hipparchus introduced. Using the epicycle, eccentric orbits, and the equant point (the off-center position across from the Earth from where the planets appear to have uniform motion), Ptolemy's model was quite fruitful in generating predictions, although the metaphysics it endorsed went against the deep-seated belief that the Earth was at the very center of everything. The *Almagest* generated controversy and attempts to revive a purer geocentrism for nearly 1,500 years. For instance, Martianus Capella (365–440) argued that Mercury and Venus are visible only near the Sun at dusk or dawn because they circle the Sun (rather than the Earth) as the Sun orbits the Earth in a circle. (See Neugebauer 1969 for details of the history of astronomy in antiquity.) Between the fifth and tenth centuries C.E., Muslim astronomers contributed the bulk of observational records and improved instrumentation. Western European astronomy owed much to their colleagues in the East for introducing them in the tenth century not only to Greek classics like Aristotle but also to Ptolemy's *Almagest*, novel terminology (e.g., terms like "azimuth" and "zenith" and the names of several stars such as

Betelgeuse and Algol), and tools such as the sextant (invented by the Arabic astronomer Abu Abdullah al Battanti around 1000 C.E.). For the next several centuries, newly found texts were translated into Latin and disseminated across European centers of learning.

Dissatisfaction with Ptolemy's eccentric orbits and equant points was raised an order of magnitude around 1543, when Copernicus released his theory of a Sun-centered universe, where the Earth not only revolves around the Sun but also rotates on its own axis. Although not as exact for predictions as Ptolemy's model, Copernicus offered a less complicated geometry that made his theory desirable, even if controversial. The Greeks and the Bible were revisited for arguments against the motion of the Earth, although, by now, several counterarguments existed. The absence of stellar parallax—change in a star's apparent position due to the motion of the Earth around the Sun—was also considered telling evidence against heliocentrism. But breaking the Earth free from the center of the universe opened up the way for new ideas.

Tycho Brahe famously disagreed with Copernicus's theory, even though as an observer Brahe was unquestionably the better astronomer. In the late seventeenth century he improved upon the measurements and instrumentation gained from the Muslims, motivated by his 1572 observation of a nova in Cassiopeia, a startling event because the fixed stars were supposed to be unchangeable. In 1577, he measured a comet with such accuracy that he could determine its orbit to be within the region of the planets, indicating that the planets could not be carried about on material spheres. Brahe asked Johannes Kepler to join him in organizing his extensive data. Brahe also improved resolvability—the capacity to distinguish between two objects in the sky—to one arcminute (where there are 60 arcminutes to a degree, and 60 arcseconds to an arcminute).

Kepler, who had been educated in and was not averse to Copernicus's work, devoted his time to finding the most efficient geometrical model for celestial motion. No combination of simple circular orbits matching planets' motions east to west was able simultaneously to accommodate their changes in altitude. The models were off by as much as 8 arcminutes, now unacceptable because Brahe's measurements were more precise than that. Kepler fashioned his own geometrical account for planetary motion, publicly voiced in *Mysterium cosmographicum* (1596) and in *Astronomia nova* ([1609] 1992): Planetary orbits are ellipses with the Sun at one focus, and planetary motion is not uniform. (For more detail, see Pannekoek 1961.)

The Telescope

As Kepler disseminated his theory of planetary motion, Galileo was experimenting with the next generation's defining icon of astronomy: the optical telescope. Originating in the Netherlands but spreading rapidly throughout Europe, this device single-handedly transformed the conception of the nature of heavenly bodies and their relation to Earth. With an eightfold magnification, Galileo saw things never before imagined: that the moon and Saturn were not perfectly smooth spherical bodies, that Jupiter was orbited by its own moons, and that the planets appeared proportionately enlarged through magnification but the stars did not, indicating that the stars were extraordinarily far away. Received with skepticism at first, telescopes increased in popularity as confirmation of Galileo's findings grew. In 1656, Christiaan Huygens declared that Saturn's earlike appendages were rings that encircled the planet; others saw dark patches on the Sun (sunspots) and that Venus exhibited phases much like the Earth's moon.

Long-held beliefs about celestial bodies were severely strained, but jettisoning an Aristotelian organization of matter meant being without a natural explanation for why the planets orbit and why objects on the Earth do not fly off into space as the Earth rotates. In 1687, Isaac Newton proposed just such an explanation in his foundational work in physics with a theory of gravitation (see Classical Mechanics). The mathematical sophistication of his *Principia* (Newton [1687] 1999) made it inaccessible to many, but through colleagues such as Samuel Clark, Newton's word gradually spread.

The eighteenth and nineteenth centuries witnessed the invention of new telescope designs such as Newtonian, Cassegrain, and Gregorian focuses, and the micrometer. Mirrors gradually replaced lenses, and mirrors first made with polished speculum were replaced with lighter and easier-to-construct glass with silver reflective surfaces. In 1826, 9.5 inches was the largest diameter with which a good mirror could be made; by 1897 it was 40 inches (Pannekoek 1961). Between the seventeenth and nineteenth centuries, astronomical announcements began appearing in such journals as the *Philosophical Transactions of the Royal*

Society of London, which began publishing in 1665; the monthly *Astronomische Nachrichten* in 1823; *Monthly Notices of the Royal Astronomical Society* in 1827; and the *Astronomical Journal* in 1849. International collaboration was also becoming practical as the ease and rapidity of communication improved (the telegraph was introduced in 1838, the telephone in 1876). In 1887, 56 scientists from 19 nations met in Paris to begin a collaborative photographic sky atlas, the *Carte du ciel.* Sapping the energies of some observatories' staffs for several decades, the *Carte* was never completed. Not trivially, changes in scientific publishing helped speed the process of transmission and exchange of information. In November 1782, the star Algol was seen to change regularly in brightness, one night passing from what appeared to be of fourth magnitude to second in a matter of mere hours; it kicked off a flurry of projects to discover variable stars. William Herschel discovered Uranus in 1781. An asteroid belt between Mars and Jupiter was detected in 1801–1807. By 1846, astronomers were fully aware of how subtly Mercury's perihelion moves faster in longitude than Newton's theory of gravitation predicted.

By 1838, resolution had improved to the point that astronomers could finally detect stellar parallax, giving long-anticipated evidence of the Earth's orbit around the Sun. In 1785, William Herschel, assisted by his sister Caroline, published a general star count, categorizing stars with others nearby of similar magnitude, and included a map of how the stars are distributed across the sky. In 1785, Herschel was also the first to detect star clusters and, later, in 1790, what came to be called planetary nebulae. In 1802, he released a catalog of 2,500 nebulae. John Herschel followed in his father's line of work and took a 20-inch telescope to the Cape of Good Hope. In 1847, he released an all-sky survey (Pannekoek 1961).

No one knew exactly what the nebulae were; they appeared as wispy light spots, but were they gas clouds inside the galaxy or something else entirely? In 1845, William Parsons was able to make out, with a 6-foot reflecting telescope of his own construction, that the nebula M51 was spiral shaped. In 1852, Stephen Alexander suggested that the solar system's galaxy was possibly spiral as well, and he began conducting an analysis of stars starting with the closest and working outward. By the turn of the century it was understood that most stars lay in a flat disc, 8–10 times greater in diameter than in thickness, and with a radius of approximate 10–20 light years. But astronomy would have never passed beyond the stage of merely recording positional and luminosity measurements without the incorporation of spectroscopy.

The Spectrometer

As early as 1670, Newton taught that light from the Sun, commonly thought to be simple, was composed of several colors. In 1802, William Hyde Wollaston looked more closely and found seven dark bands in the Sun's spectrum, which he assumed were natural spaces between the colors, but 15 years later, Joseph Fraunhofer looked at the Sun's light still more closely with the first spectrometer set up with a telescope and found several hundred absorption lines. Gustav Kirchhoff and Robert Bunsen began interpreting these dark lines in 1858 as being due to elements in excited states, with each element having its unique set of lines. Spectroscopy, especially after early twentieth-century improvements in physics, provides a wealth of information about objects in space, their constitution, their density, and their physical conditions. Earlier in the nineteenth century, August Comte had declared that the chemical constitution of the Sun was inherently unknowable. But by 1862, A. J. Ångstrom identified hydrogen within the Sun, and 50 more elements were identified by the end of the decade. Studies of the Sun during solar eclipses revealed spectral lines not yet identified on Earth. One of them, called "helium," from the Greek word for the Sun, was later, in 1895, isolated in laboratories as a product of radioactive decay. In 1864, Giovan Battista Donati observed that comets' emission—presumed to be no more than reflected sunlight—had its own unique composition, containing carbon monoxide, hydrogen, methane, and ethylene. In conjunction with photographic techniques introduced in the nineteenth century, and the longer exposure and integration times photographic plates allowed, astronomers obtained spectra from ever fainter sources. The star Vega was daguerreotyped in 1850. In 1882, Henry Draper took a 137-minute exposure of the Orion Nebula that showed the entire nebula and faint stars in it. In 1879, William Huggins demonstrated that Vega had a spectrum in the ultraviolet.

With spectroscopy, astronomers can determine velocity of a star along the line of sight and its distance. Spectral lines exhibit Doppler shifting, already understood in the case of sound waves in 1842 to be a measure of motion toward and away from an observer, and the first stellar radial velocities were measured as early as 1890. It is also found on the basis of spectral features that stars fall into groups. Widely used today is the Harvard classification of spectral types, which began in the late 1880s. In 1911 and 1913, Ejnar Hertzsprung and Henry Norris Russell plotted the relationship of spectral type against known absolute magnitudes (or luminosity), creating the first of what have come to be called *H-R diagrams.* Ever since their introduction, the distance to a star can be determined by locating it on an H-R diagram on the basis of its observed spectrum and a reading of its absolute magnitude from its location. Since brightness is a function of distance, the difference between a star's absolute magnitude and its observed magnitude gives its distance.

Instead of spectral lines, Harlow Shapley and Henrietta Leavitt probed galactic distances using Cepheid variable stars. Cepheids cycle regularly in brightness, with a period related to its absolute magnitude. Shapley calculated the distance to nearby clusters of stars containing Cepheids and, assuming that star clusters are similar, compared the apparent magnitudes of the brightest stars in distant and nearby star clusters. By 1917, Shapley concluded that the remotest clusters were on the order of 200,000 light years away, and that the clusters were symmetrically concentrated around the region of Sagittarius. Shapley's proposal that the center of the galaxy was some 200,000–300,000 light years away was received with skepticism. If the solar system was so off-center from the galaxy, observers should see systematically red- and blue-shifted spectra from neighborhood stars circling around with the Earth—evidence that Jan Oort provided 10 years later.

By the end of the nineteenth century, astronomers' comprehension of the solar system had become quite reliably systematic. Outside the solar system it was entirely another matter: Data vastly exceeded explanatory power. Astronomers had a start on figuring the constitution of the Sun and other stars, but no good idea of how they worked. Shapley gave a picture of the Sun's place in the galaxy, but no one knew what the galaxy overall was like. As early as 1844, Bessel recognized that Sirius had an invisible companion as massive as the Sun and with the size of the Earth, but nothing in physics accounted for such density. Explanations for such phenomena would not be forthcoming until the early decades of the 1900s.

Meanwhile astronomers began thinking that the Milky Way may not be the only galaxy in the universe. As early as 1755, Immanuel Kant proposed that faint nebulous patches of light were more likely far away stellar systems than nearby diffusions of glowing gas. In 1913, V. M. Silpher's spectral analysis of M31, the Andromeda nebula, showed its radial velocity to be 300 km/s—quite extreme considering that normal stellar velocities tended around 20 km/s. In 1917, the spectra of 25 nebulae showed that 4 had velocities greater than 1000 km/s, although this conclusion remained hotly contested. Adriaan van Maanen argued that spiral nebulae were observed to be rotating—indicating that they must be small and relatively nearby (that is, within the galaxy), and even Shapley, the ingenious observer that he was, did not believe they were anything like external galaxies.

In 1923, Edwin Hubble examined spiral nebulae for signs of their being composed of stars—novae and Cepheid variables. That he found the requisite signs gave highly supportive evidence for the claim that M31 was an external stellar system, starkly opening up the realization there were other galaxies in the universe besides Earth's. By 1929, Hubble had calculated red shifts for 24 galaxies, finding that the velocity of a galaxy was proportional to its distance—the constant of proportionality now known as *Hubble's constant.* Why galaxies should be moving apart with such high velocity, and why the elements that are observed came to exist at all, remained a matter of speculation in the early twentieth century. Some theories were put forward, such as Georges Lemaître's 1931 primeval atom hypothesis, a close relative to George Gamow's 1948 theory of the "Big Bang" (see Berger 1984). Hermann Bondi, Fred Hoyle, and Thomas Gold in 1948 alternatively supported the thesis that the universe was constantly creating matter and maintaining itself in a steady state (Bondi and Gold 1948). The conflict between the Big Bang and Steady State cosmologies has had a significant influence on theory and experiment in astronomy in the latter half of the twentieth century.

Note that for the entire period of astronomy's existence so far discussed, observers had concerned themselves with only the small fraction of the electromagnetic spectrum available to the human eye. The next generation of astronomical research, beginning in the 1930s, was marked by scientists, quite often not starting out as astronomers, looking to the universe at other wavelengths.

Beyond the Optical

Radio waves from space were first found by Karl Jansky in the 1930s, but the study of radio astronomy did not take off until physicists and engineers, primarily in Britain and Australia, refocused their radar antennas after World War II. To general surprise, including their own, they often found intense signals where nothing optically interesting was seen. For several years they had to battle the common (but not universal) professional preconception that radio wavelengths would not show anything interesting about the universe. By the 1960s, radio instruments had improved to such an extent that optical astronomers were able to get help with optical identification, and radio astronomy was shown to have discovered some very interesting things.

Because radio waves are so long (ranging from 1 mm to 100 m), they travel relatively undisturbed by dust and gas through space. This means that radio waves originating extremely far away in space and time can be detected, and consequently radio astronomy played a central role in adjudicating between the Big Bang and Steady State cosmologies. Atoms and molecules also emit radiation at radio wavelengths: Atomic hydrogen was the first detected through radio waves in 1951 and was found to exist abundantly between the stars. Since 1951, radio astronomers have detected over 100 molecules, many of which are organic and complex. Atomic and molecular spectroscopy has revolutionized the understanding of the constitution of interstellar space, which had long been thought to be either empty or too vulnerable to cosmic rays for molecules to exist. These developments spawned the now mature interdisciplinary research field of astrochemistry (see Sullivan 1984 for more historical detail).

Starting in the 1930s, astronomical and military interests combined to develop finer infrared technology. Similar to radio, infrared wavelengths are long enough to be quite immune to dust blockage, although parts of its spectrum are blocked by the Earth's atmosphere, making high-altitude or satellite observations often preferable and sometimes necessary. Infrared astronomers are able to look at the youngest of stars developing deep inside their cocoons of dust and gas. Some mature stars, like Vega, were found to radiate more in infrared than one would expect for its spectral type. At least 50 extrasolar planetary systems have been detected through various instruments. Hundreds of galaxies have been detected as radiating over 95% of their total luminosity in the infrared (whereas the Milky Way radiates approximately 50%). Some of the more crucial measurements of the cosmic microwave background radiation have occurred at infrared frequencies.

Ultraviolet, x-ray, and gamma ray astronomy also came of age after the 1930s. All of these areas of study must collect their data above the Earth's atmosphere. As a result, spectroscopic and photometric surveys have been done initially using V-2 rockets and, later, satellite conveyors. These studies have aided the understanding of what the universe is like in its more energetic states. Ultraviolet spectroscopy provides evidence of the existence of molecular hydrogen existence throughout the universe, and provides a lower limit (10^6 °K) on the temperature of some of the gas in and around the galaxy. X-ray astronomy has recorded evidence of the high-energy features around black holes, and these were found to exist to an unexpected extent. At the shortest wavelengths, sudden bursts of extraordinarily intense gamma radiation have been detected, and with the assistance of instruments operating at other wavelengths, some of the sources have been isolated, although their nature is not yet completely understood. Astronomy, with the inclusion of techniques from physics, chemistry, and many types of technology, has radically transformed humanity's conception of itself and its place in the universe. It was once believed that Planet Earth was unique and at the center of everything. Now it is known that the Earth is far from the center and lies in only one of millions of galaxies in the universe, where neither solar systems nor carbon-rich molecules are unusual.

Astronomy has been dubbed a passive, observational enterprise. But the exploration of other wavelength regimes outside of the small fraction of the electromagnetic spectrum comprised by the optical range stretches the sense of "observation" far beyond any commonsense conception of the word (see Observation). Astronomers employ a complicated technological network on Earth and in space to interact with the causal nexus stemming from their objects of study. Designating astronomy as "passive" hardly seems a fit descriptor of the science. Although dated, it is philosophically common to construe scientific activity as a process of systematically testing theories against data, with an emphasis on theory. But much of the history of astronomy is marked by important periods when increasing quantities of data had no good theoretical explanation, and sometimes no explanation at all. By and large, theories of the universe and nature of the bodies occupying it have largely remained inert until new technologies powered change in sometimes highly unexpected directions. Consider astrobiology and dark matter as a couple of contemporary examples.

The issue of whether or not life exists elsewhere than Earth is on record for as long as any issue in philosophy, all decisions on the issue being able to rest on little more than a priori reasoning.

Recent technological developments have provided an unprecedented empirical basis for determining that there are indeed planets around other stars. Astrochemistry has revealed a universe replete with organic molecules. In 1986, what had been dubbed "exobiology" became the more institutionally organized disciple of astrobiology, focusing upon the detection of chemicals in space indicative of signatures of forms of life, such as with the detection of oxygen or methane from extrasolar planets, and perhaps the detection of extraterrestrial life itself, or its remains (on, e.g., Mars, Europa).

Michelle Sandell

References

Berger, A. (ed.) (1984), *The Big Bang and Georges Lemaître* Dordrecht, Holland: Reidel.

Bondi, H., and Gold, T. (1948), "The Steady State Theory of the Expanding Universe," *Monthly Notices of the Royal Astronomical Society* 108: 252–270.

Hoskin, Michael (1997), *The Cambridge Illustrated History of Astronomy*. Cambridge, UK: Cambridge University Press.

Kepler, J. ([1609] 1992), *The New Astronomy*. Cambridge, UK: Cambridge University Press.

Lankford, J. (ed.) (1997), *History of Astronomy: An Encyclopedia*. New York: Garland.

Munitz, Milton K. (ed) (1957), *Theories of the Universe*. Glencoe, IL: Free Press.

Neugebauer, Otto (1969), *The Exact Sciences in Antiquity* (2nd ed.). New York: Dover Publications.

Newton, I. ([1687] 1999), *The Principia: Mathematical Principles of Natural Philosophy* Berkeley and Los Angeles: University of California Press.

Ptolemy ([137 C.E.] 1984), *Almagest*. Berlin: Springer.

Pannekoek, A. (1961), *A History of Astronomy*. New York: Dover Publications.

Sullivan, W. T. (1984), *The Early Years of Radio Astronomy*. Cambridge, UK: Cambridge University Press.

See also **Classical Mechanics**

A. J. AYER

(29 October 1910–27 June 1989)

Alfred Jules Ayer attended Eton College and then Oxford, taking a first in classics in 1932. Impressed by Ludwig Wittgenstein's *Tractatus,* which he studied in late 1931, Ayer embarked on a career in philosophy. He was Lecturer (and then Fellow) at Christ College, and subsequently at Wadham College, Oxford. In 1946 he was elected Grote Professor of the Philosophy of Mind and Logic at University College, London, and in 1959 he became Wykeham Professor of Logic at Oxford. He was the author of more than 100 articles and several books about knowledge, experience, and language, among them *Language, Truth, and Logic* ([1936] 1946), *The Foundations of Empirical Knowledge* (1940), and *The Problem of Knowledge* (1956) (Rogers 1999).

Language, Truth, and Logic

Ayer's significance to the philosophy of science comes primarily as author of *Language, Truth, and Logic* (LTL), published in January of 1936. LTL was Ayer's first book and the most widely read early English discussion of the *wissenschaftliche Weltauffassung* (scientific world-conception) of the logical positivists of the Vienna Circle, whose meetings Ayer attended from December 1932 through March 1933 (see Logical Empricism). LTL is directly concerned less with science and the philosophical issues it raises than with philosophy and the form philosophy takes in light of the *verifiability criterion of significance,* a condition of meaningfulness applied to propositions (see Verificationism and Cognitive Significance). Philosophy in this relatively new form involved the dismissal of a swath of traditional philosophical problems (deemed "metaphysical" or "nonsensical") and the identification of what remained with the logical analysis of scientific claims. LTL's significance to the philosophy of science therefore arises less from the answers it offered to questions *within* the philosophy of science than from the vision of philosophy it popularized, a vision that made philosophy dependent upon science. Thus while LTL begins, famously, with the declaration that the "traditional disputes of philosophers are, for the most part, as unwarranted as they are unfruitful" ([1936] 1946, 33), it ends with an admonition:

> [P]hilosophy must develop into the logic of science . . . [which is] the activity of displaying the logical relationship of . . . hypotheses and defining the symbols which occur in them. . . . What we must recognize is that it is necessary for a philosopher to become a scientist, in this sense, if he is to make any substantial contribution towards the growth of human knowledge. (153)

In 1936, this view of philosophy was hardly novel, nor did Ayer claim it was. In LTL's original preface, Ayer credited his views to "the doctrines of Bertrand Russell and Wittgenstein, . . . themselves the logical outcome of the empiricism of Berkeley and David Hume" ([1936] 1946, 31). And among the members of the Vienna Circle, Ayer singled out as influential Rudolf Carnap, whom he met in London in 1934 when Carnap lectured on his recently published *Logische Syntax der Sprache* (Carnap 1934; Rogers 1999, 115).

Though Ayer's message was familiar to many, the verve with which LTL advanced this vision attracted readers who had ignored scientific philosophy, and it allowed LTL to orient (particularly British) discussion of scientific philosophy. The result was considerable attention for LTL and (particularly in Britain) the near identification of Ayer with logical positivism. For a time, Ayer was a (if not the) leading English proponent of scientific philosophy. It was a curious mantle to fall to someone who, in contrast to the leading members of the Vienna Circle, lacked (and would continue to lack) scientific knowledge or training of any sort (Rogers 1999, 129–130).

Behind Ayer's scientific conception of philosophy was the verifiability criterion, according to which, as Ayer ([1936] 1946) formulated it early in LTL, the

> question that must be asked about any putative statement of fact is, Would any observations be relevant to the determination of its truth or falsehood? And it is only if a negative answer is given to this . . . question that we conclude that the statement under consideration is nonsensical. (38)

Otherwise, the "putative" statement of fact is meaningful.

This criterion did not originate with Ayer, nor did he claim originality. Moreover, Ayer followed the logical empiricists in applying the criterion only to matters of fact, not mathematical or logical propositions. These were not verifiable, but neither were they meaningless. They were true because they were *analytic,* that is, their truth was a consequence of the meanings of their terms. And it was, moreover, thus that analytic (and only analytic) propositions could be known a priori. Thus Ayer adopted the familiar division of propositions into either synthetic *a posteriori* propositions (subject to the verifiability criterion and informative) or analytic a priori propositions (tautologous and uninformative). Any purported proposition falling into neither category was a meaningless *pseudo*-proposition (Ayer [1936] 1946, 31; cf. e.g., Carnap [1932] 1959).

Ayer's "modified" verificationism required neither that verification be certain nor that a proposition's verifiability be immediately decidable. In Ayer's view, a proposition is meaningful if evidence can be brought to bear on it *in principle.* "There are mountains on the farther side of the moon" was Ayer's example (borrowed from Moritz Schlick); this claim could not be tested in 1936, nor its truth or falsity established with certainty, but it is nevertheless significant or meaningful (Ayer [1936] 1946, 36). Even in such modified form, the criterion ultimately met with insurmountable criticism and came to be recognized as inadequate (see Cognitive Significance for further detail).

In the remaining chapters of LTL, Ayer applied the verifiability criterion to traditional philosophical problems, illustrating more than arguing for LTL's vision of a scientific philosophy. The range of philosophical issues across which Ayer wielded the criterion, combined with his efficient and unflinching (if ham-fisted) manner, remains remarkable. It is perhaps in range and vigor that Ayer and LTL can claim to have contributed to, rather than just echoed, logical empiricism's antimetaphysical project. In just its first two chapters, Ayer addresses Cartesian skepticism about knowledge of the world, monism versus pluralism, and the problem of induction—finding all of these to be "fictitious." Most notably, verifiability led infamously to emotivism, the view that "in every case in which one would commonly be said to be making an ethical judgment, the function of the relevant ethical word is purely 'emotive.' It is used to express feeling about certain objects, but not to make any assertion about them" ([1936] 1946, 108).

It is instructive to compare LTL with Carnap's *Überwindung der Metaphysik durch Logische Analyse der Sprache* (Carnap [1932] 1959), an influential essay Ayer read and cited in LTL. Nearly all of LTL's conceptual apparatus—verifiability, analyticity, opposition to traditional philosophy—are found in the *Überwindung*. However, where Carnap ([1932] 1959, 77) characterized philosophy not as making statements but as applying a *method*, Ayer ([1936] 1946, 57) regarded philosophy as consisting of tautological statements clarifying our language. And this is no small difference, for it reflects Carnap's recognition of the *significance* of metaphysics as an "expression of an attitude toward life" rather than a theory about the world. Metaphysicians were often guilty, claimed Carnap, of trying to express as a theory an attitude to be expressed by poetry or music; but metaphysics *itself* was not useless. For Ayer, though, metaphysics had *no* authority over life; its pseudo-propositions were grammatical confusions, not expressions of a legitimate "attitude" (cf. Rogers 1999, 97–98). Ayer's antimetaphysical bent ran deeper than Carnap's and was less tolerant, a fact that may explain logical empiricism's subsequent reputation for intolerance.

After *Language, Truth, and Logic*

The revised 1946 edition of LTL gave Ayer occasion to respond to criticism. While asserting that "the point of view which [LTL] expresses is substantially correct," he ceded ground on several fronts, most notably in recognizing basic empirical propositions, which "refer solely to the content of a single experience" and can be "verified conclusively" by it (Ayer [1936] 1946, 10; [1940] 1959). Ayer's epistemic foundationalism remained characteristic of his views and provoked others, especially J. L. Austin (Rogers 1999, 146–147).

Such adjustments to verificationism are often regarded as evidence of logical positivism's decline, resulting in replacement by a holism defended by Carl Hempel, Willard Van Quine, and Thomas Kuhn. Verificationism did wane, but the scientific philosophy Ayer popularized did not depend on verificationism; and, moreover, Hempel, Quine, and Kuhn retained elements of scientific philosophy (Friedman 1999; Hardcastle and Richardson 2003). And it was in these terms that Ayer gauged the influence of LTL, for example lamenting (in 1959) that "among British philosophers" there was little "desire to connect philosophy with science" (Ayer 1959, 8), but noting with pride that "in the United States a number of philosophers . . . conduct logical analysis in a systematic scientific spirit . . . [close] to the . . . ideal of the Vienna Circle" (7–8). To the considerable extent to which Ayer and LTL caused this, Ayer's influence upon the philosophy of science is significant.

GARY L. HARDCASTLE

References

Ayer, Alfred Jules ([1936] 1946), *Language, Truth, and Logic*. London: Victor Gollancz.

——— ([1940] 1959), "Verification and Experience," in *Logical Positivism*. New York: The Free Press, 228–243. Originally published in *Proceedings of the Aristotelian Society* 37: 137–156.

——— (1940), *The Foundations of Empirical Knowledge*. London: Macmillan Press, 1940.

——— (1956), *The Problem of Knowledge*. New York: St. Martin's Press.

——— (1959), *Logical Positivism*. New York: Free Press.

Carnap, Rudolf ([1932] 1959), "The Elimination of Metaphysics through Logical Analysis of Language," in A. J. Ayer (ed.), *Logical Positivism*. New York: Free Press, 60–81. Originally published as "Überwindung der Metaphysik durch der Logische Analyse der Sprache," *Erkenntnis* 2: 219–241.

——— ([1934] 1937), *Logical Syntax of Language*. London: Kegan Paul. Originally published as *Logische Syntax der Sprache*. Vienna: Springer.

Friedman, Michael (1999), *Reconsidering Logical Positivism*. Cambridge: Cambridge University Press.

Hardcastle, Gary L., and Richardson, Alan (eds.) (2003), *Logical Empiricism in North America*. Minneapolis: University of Minnesota Press.

Rogers, Ben (1999), *A. J. Ayer*. New York: Grove Press.

See also **Carnap, Rudolf; Cognitive Significance; Hempel, Carl Gustav; Kuhn, Thomas; Logical Empiricism; Quine, Willard Van; Verifiability**

B

BASIC SENTENCES

See **Popper, Karl; Protocol Sentences**

BAYESIANISM

Thomas Bayes's *Essay* ([1763] 1958) initiated a penetrating mathematical analysis of inductive reasoning based on his famous rule for updating a probability assignment (see equation 1b below) (see Induction, Problem of; Inductive Logic). Rediscovered and generalized by Laplace a decade later (equation 1d), it found widespread applications in astronomy, geodesy, demographics, jurisprudence, and medicine. Laplace used it, for example, to estimate the masses of the planets from astronomical data and to quantify the uncertainty of such estimates. Laplace also advanced the idea that optimal estimation must be defined relative to an error or loss function—as minimizing the expected error (Hald 1998, sec. 5.3; Jaynes 2003, 172–174).This was an important source of decision theory, whose revitalization at the hands of Ramsey, von Neumann, Wald, and Savage helped launch the modern Bayesian revival (see Decision Theory; Ramsey, Frank Plumpton; von Neumann, John). Bayesian decision theory has been extended in recent years to game theory by Harsanyi, Skyrms, and others. The focus of the present entry, however, is the distinctive and influential philosophy of science extracted from Laplace's rule (equation 1d). A survey of the issues that divide the Bayesian from rival philosophies is followed by a sketch of the mathematical analysis of inductive and predictive inference initiated by Bayes and Laplace and its extensions. There follows a section on minimal belief change, then some remarks on the alleged subjectivity of the prior probability inputs needed for a Bayesian inference and various attempts to "objectify" these inputs.

Bayesian Logic and Methodology

Cox (1946) derived the usual rules of probability, the *product rule,*

$$P(A \wedge B \mid I) = P(A \mid B \wedge I)P(B \mid I), \tag{1a}$$

and the negation rule, $P(A \mid I) = 1 - P(A \mid I)$, from a requirement of agreement with qualitative common sense and a consistency requirement that *two ways of performing a calculation permitted by the rules must yield the same result* (Jaynes 2003, chaps. 1–2). Using the equivalence of the conjunctions $A \wedge B$ and $B \wedge A$ and equation 1a, this requirement yields the symmetric form of the product rule,

$$P(A \mid B \wedge I)P(B \mid I) = P(B \mid A \wedge I)P(A \mid I)$$

of which Bayes's rule is, in more suggestive notation, the trivial variant:

$$P(H \mid D \wedge I) = P(H \mid I)P(D \mid H \wedge I)/P(D \mid I), \tag{1b}$$

where H is hypothesis, D is data, and I is the assumed information, which is included to note that probabilities depend as much on the background information as on the data.

The probability $P(D \mid H \wedge I)$ is called the sampling distribution when considered as a function of the data D, and the likelihood function *qua* function of the (variable) hypothesis H. The "most likely" hypothesis (or parameter value) is thus the one that maximizes the likelihood function, that is, the one that accords D the highest probability. It follows, then, from equation 1b that the hypothesis of a pair comparison that accords D the higher probability is *confirmed* (made more probable) and its rival *disconfirmed.* In fact, Laplace was led to the odds form of equation 1b, which expresses the updated odds as the product of the initial odds, by the likelihood ratio (LR), thus

$$\begin{aligned} &P(H \mid D \wedge I){:}P(K \mid D \wedge I) \\ &\quad = [P(H \mid I){:}P(K \mid I)] \times [P(D \mid H \wedge I){:}P(D \mid K \wedge I)] \end{aligned} \tag{1c}$$

as a quantitative sharpening of this qualitative condition. But he lacked a compelling reason for preferring equation 1c to alternative rules, for example, rules that multiply the initial odds by a positive power of the LR, and this led to much agonizing over the basis for Bayes's rule. It has even led some contemporary philosophers to claim that all such rules, agreeing as they do in their qualitative behavior, are on an equal footing. To appreciate the force of Cox's derivation of equation 1a is to recognize that *all these alternatives to Bayes's rule are inconsistent* (see Rosenkrantz 1992 for some illustrations). Moreover, the optimality theorem shows that Bayesian updating outperforms all rivals in maximizing one's expected score after sampling under any *proper scoring rule,* that is, a method of scoring forecasts that gives one no incentive to state degrees of prediction different from one's actual degrees of belief.

According to equation 1b, $P(H \mid D \wedge I)$ is directly proportional to $P(D \mid H \wedge I)$ and inversely proportional to $P(D \mid I)$. The latter is generally computed from the partitioning formula:

$$\begin{aligned} P(D \mid I) = {} & P(D \mid H_1 \wedge I)P(H_1 \mid I) + \ldots \\ & + P(D \mid H_n \wedge I)P(H_n \mid I) \end{aligned} \tag{2}$$

where $H_1, \ldots, H_n$ are mutually exclusive and jointly exhaustive in light of I. Given such a partition of hypotheses (the "live" possibilities from the perspective of I), Laplace recast equation 1b as

$$P(H_j \mid DI) = \frac{P(H_j \mid I)P(D \mid H_j \wedge I)}{\sum_{j=1}^{n} P(D \mid H_j \wedge I)P(H_j \mid I)} \cdot \tag{1d}$$

Hence, H_j is confirmed, so that $P(H_j \mid DI) > P(H_j \mid I)$, when $P(D \mid H_j I) > P(D \mid I)$, or when the probability that H_j accords D (in light of I) exceeds the weighted average (equation 2) of the likelihoods. In particular, this holds when $P(D \mid H_j I) = 1$ provided $P(D \mid I) < 1$, or: *Hypotheses are confirmed by their consequences and the more so as these are unexpected on the considered alternative hypotheses.*

Seen here as well are two further implications for the scientific method: first, the dependence of D's import for a hypothesis on the considered alternatives and, second, the tenet that evidence cannot genuinely disconfirm, much less rule out, a hypothesis merely because it assigns a low probability to the data or observed outcome. Rather, disconfirmation of H requires that some alternative accord D a higher probability. Thus, Bayesian inference is inherently *comparative;* it appraises hypotheses relative to a set of considered alternatives. At the same time, there is nothing to stop one from enlarging the "hypothesis space" when none of the considered alternatives appears consonant with the data (see Jaynes 2003, 99) or when predictions based on a given hypothesis space fail.

These illustrations already hint at the many canons of induction and scientific method that follow *in a sharper quantitative form* from equation 1. While critics of Bayesianism deny it, there is an abundance of evidence that working scientists are guided by these norms, many of them illustrated in the extensive writings of Polya on induction in mathematics. Then, too, there is the growing army of Jaynesians, Bayesian followers of Jaynes and Jeffreys, who apply Bayes's rule and its maximum entropy extension to an ever-expanding range of scientific problems, expressly endorse its methodological implications, and condemn violations thereof (Loredo 1990).

The Testing Approach

To bring out salient features of the Bayesian approach, it will be useful to survey some of the issues that divide it from its major rivals, which may be lumped together, notwithstanding minor variations, as the *testing approach.* In its most developed form (Giere 1983; Mayo 1996), it holds that a hypothesis h is confirmed—rendered more *trustworthy* as well as more *testworthy*—when and only when it passes a *severe* test, that is, one with a low probability of passing a false hypothesis. This approach grew out of the writings of Peirce (see Mayo 1996, chap. 12) and of Popper and, above all, out of Neyman and Pearson's approach to testing statistical hypotheses, embraced by most "orthodox" statisticians (see de Groot 1986, Chap. 7; Hodges and Lehmann 1970, Chap. 13, for accessible introductions). This approach directly opposes Bayesianism in denying any distinctive evidential relation between data and hypotheses. It relies on only "direct" probabilities of outcomes conditional on hypotheses and the assumed model of the experiment.

Consider a medical diagnostic test with the conditional probabilities given in Table 1.

Given that a patient tests positive (+), a Bayesian multiplies the prior odds of infection by the LR,

$$P(+ \mid h_0) : P(+ \mid h_1) = 94:2$$

to obtain the updated odds. (An LR of 49:1 is evidence of infection slightly stronger than the 32:1 LR a run of 5 heads accords the hypothesis that a coin is two-headed rather than fair.) Writing $\alpha = P(- \mid h_0)$ and $\beta = P(+ \mid h_1)$ for the probabilities of the two possible errors, *viz.,* false negatives and false positives, one could view the LR of $1 - \alpha{:}\beta$ as a plausible quantification of the (qualitative) characterization of a severe test as one with low probabilities, α and β, of passing a false hypothesis. The Bayesian approach based on LRs and the testing approach based on error probabilities do not differ appreciably in such cases. The differences show up when numerical outcomes like frequency or category counts come into play, such as in the cure rate of a new drug or phenotypic category counts in a genetic mating experiment.

Table 1 Conditional probabilities for infected and non-infected

	+	−
Infected (h_0)	0.94	0.06
Uninfected (h_1)	0.02	0.98

Bayesians look at the likelihood function, $L(\theta) = f(x \mid \theta)$, of the outcome x *actually observed* to estimate the unknown parameter, θ, or to compare hypotheses about it. On the other hand, testing theorists look at *sets* of outcomes, rejecting h_0 in favor of h_1 just in case the outcome x falls in a critical region R of the outcome space. Labeling the hypotheses of a pair comparison so that erroneously rejecting h_0 is the more serious of the two errors, the recommended Neyman-Pearson (NP) procedure is to fix the probability,

$$\alpha = P(X \in R \mid h_0)$$

of this type I error at an acceptable level, $\alpha \leq \alpha_0$, called the *size* of the test, and then choose among all tests of that size, α_0, one of maximal *power*, $1 - \beta$, where

$$\beta = P(X \notin R \mid h_1)$$

is the probability of the less serious type II error of accepting h_0 when h_1 is true. Naturally, there are questions about what it means to accept (or reject) a hypothesis (for a good discussion of which see Smith 1959, 297), but these are left aside in deference to more serious objections to the NP procedure.

Consider a test of $h_0{:}P = 0.5$ versus $h_1{:}P = 0.3$, where P is the success rate in n Bernoulli trials. The test for the best 5 percent of $n = 20$ trials has $R = [X \leq 5]$, i.e., rejects h_0 when 5 or fewer successes are observed. Thus, h_0 is accepted when $X = 6$, even though that is exactly the number of successes expected when $P = 0.3$. At $n = 900$ trials, $R = [X \leq 425]$ even though the boundary point, $X = 425$, is a lot closer to the 450 successes expected when $P = 0.5$ than the 270 expected when $P = 0.3$. In fact, the LR in *favor* of h_0 when $X = 425$ is

$$P(X = 425 \mid h_0) : P(X = 425 \mid h_1) = (5/3)^{425}(5/7)^{475} = 7.5 \times 10^{24}$$

and tends to ∞ at (or near) the boundary of R as $n \to \infty$. There is, then, a *recognizable* subset of R whose elements strongly favor the rejected hypothesis. The overall type I error rate of 5% is achieved by averaging the much higher than advertised 5% probability of being misled by outcomes near the boundary of R against the much lower probability of being misled by the elements of R farthest from the boundary. For testing theorists, however, the individual test result draws its meaning solely from the application of a generally reliable rule of rejection as attested by its *average* performance over many real or imagined repetitions of the experiment. This objection to the NP procedure was first lodged by a non-Bayesian, Fisher (1956, secs. 1 and 5 of chap. 6, especially 93).

Among the strongest methodological implications of equation 1 is the likelihood principle (LP), which counts two outcomes (of the same or different experiments) as *equivalent* if they give rise to the same likelihood function *up to a constant of proportionality.* Suppose, in the last example, that a second statistician elects to sample until the 6th success occurs and this happens, perchance, on the 20th trial. For this experiment, the likelihood function is:

$$L_1(P) = \binom{n-1}{5} P^6(1 - P^{n-6})$$

which, for $n = 20$, is proportional to the likelihood function for $r = 6$ successes in $n = 20$ trials:

$$L_2(P) = \binom{20}{6} P^6(1 - P)^{14}.$$

So the likelihood functions are proportional. In particular, both experiments yield the same LR of $(3/5)^6(7/5)^{14} = 5.2$ in favor of h_1. But for the second experiment $R = [n \geq 19]$ is the best 5% test and so h_0 is rejected when $n = 20$. In a literal sense, both statisticians observe the same result, 6 successes in 20 Bernoulli trials, yet one accepts h_0 while the other rejects it. NP theory is thus charged with the most obvious inconsistency—allowing opposite conclusions to be drawn from the same data (Royall 1997, Chap. 3). Testing theorists are open to the same charge of inconsistency at a higher ("meta") level (Jaynes 1983, 185), in as much as they concede the validity of equation 1 and so, by implication, of the LP, when the needed prior probabilities are "known" from frequency data, as when one knows the incidence of a disease or of a rare recessive trait. Is one to base an evaluation of a methodological principle on such contingencies as whether given prior probabilities are known or only partially known?

Indeed, it seems perfectly reasonable for an experimenter to stop sampling as soon as the incoming data are deemed sufficiently decisive. That is, after all, the idea behind Wald's extension of NP theory to so-called sequential analysis. Could it make a difference whether one planned beforehand to stop when the sample proportion of defectives exceeded B or fell below A or so decided upon observing this event? Moreover, this issue of "optional stopping" has an ethical dimension in that failure to terminate an experiment when sufficiently strong evidence has accumulated can expose experimental subjects to needless risk or even death (see Royall 1997, Sec. 4.6, for a chilling real-life example).

But what about a fraud who determines to go on sampling until some targeted null hypothesis is rejected? This is indeed possible using significance tests, for as was seen, the power of such a test approaches unity as the sample size increases. But using the likelihood to assess evidence, the probability of such deception is slight. When h_0 holds, the probability of obtaining with any finite sample an LR, L_1:$L_0 \geq$ k, in favor of h_1 against h_0 is less than $1/k$, for if S is the subset of outcomes for which the LR exceeds k, then there is the Smith-Birnbaum inequality:

$$P(L_1 : L_0 \geq k \mid h_0) = \textstyle\sum_{x \in S} P(x \mid h_0) \leq k^{-1} \textstyle\sum_{x \in S} P(x \mid h_1) \leq 1/k.$$

More generally, as Fisher also emphasized (1956, 96), "the infrequency with which ... decisive evidence is obtained should not be confused with the force, or cogency, of such evidence." In planning an experiment, one can compute *the probability of misleading evidence,* i.e., of an LR in favor of either hypothesis in excess of $L*$ when the other hypothesis is true, just as readily as one can compute probabilities (for an NP test) of rejecting h_0 when it is true or accepting it when it is false. These probabilities,

$$P(f(x \mid h_1) : f(x \mid h_0) > L^* \mid h_0)$$

and

$$P(f(x \mid h_1) : f(x \mid h_0) < 1/L^* \mid h_1),$$

which are governed by the Smith-Birnbaum inequality, tell the experimenter how large a sample to take in order to control adequately for the probability of misleading results. But this still leaves one free to break off sampling if sufficiently strong evidence turns up before many trials are observed. Thus, one can have the best of both worlds: (a) the use of likelihood to assess the import of the results of one's experiment and (b) control, in the planning stage, of the probability of obtaining weak or misleading evidence—the feature that made NP theory so attractive in the first place. Richard Royall (1997, Chap. 5) refers to this as the "new paradigm" of statistics.

One is free, in particular, to test new hypotheses against old data and to use the data as a guide to new models that can be viewed, in Bayesian terms, as supporting those data. Testing theorists take issue with this: first, in proscribing what Pearson branded "the dangerous practice of basing one's choice of a test . . . upon inspection of the observations"; and, second, in maintaining that "evidence predicted by a hypothesis counts more in its support than evidence that accords with a hypothesis constructed after the fact" (Mayo 1996, 251). The idea in both cases is that such tests are insufficiently severe.

Even orthodox statisticians routinely transgress, as when they check the assumptions of a model against the same data used to test the relevant null hypothesis (and, perhaps, base their choice of test on the departure from the relevant assumption indicated by those data), or when they test the hypothesis that two normal variances are equal before applying a *t* test (see Jaynes 1983, 157), or when they quote "critical" or "exact" significance levels (Lehmann 1959, 62) or test a complication of a model against the data that prompted that complication. Indeed, it is literally impossible to live within the confines of a strict predesignationism that requires that the sampling rule, the tested hypothesis, and the critical region (or rejection rule) all be stated prior to sampling; and the examples to show this come from the bible of orthodox testing (Lehmann 1959, 7). For Bayesians, the evidential relation is timeless, and no special virtue attaches to prediction. See Giere (1983, 274–276) and Mayo (1996, 252) for some of the history of this controversy in the philosophy of science and (new and old) references to such figures as Whewell, Mill, Jevons, Peirce, and Popper. Taking the extreme view, Popper insisted that a scientific theory cannot be genuinely confirmed ("corroborated") at all by fitting extant data or known effects but only by withstanding "sincere attempts" at refutation (see Corroboration; Popper, Karl Raimund). Critics were quick to point out the paradoxicality of Popper's attempt to confer greater objectivity on theory appraisal by appeal to a psychologistic notion. In responding to this criticism, Popper did what scientists often do (but he forbids): He amended his characterization of a severe test to read: "A theory is [severely] tested by applying it to ... cases for which it yields results different from those we should have expected without that theory, or in the light of other theories" (1995, 112). The problem for testing theorists is to detach this criterion from its transparently Bayesian provenance.

For Giere and Mayo, a severe test is one in which the theory has a low probability of passing if false, but the question is how to compute that probability without reference to alternative hypotheses. One knows how to compute such error probabilities in statistical contexts where they are given by the assumed model of the experiment. But what meaning can be attached to such probabilities in scientific contexts, as in Giere's (1983) example of the white spot that Fresnel's wave theory of diffraction predicted would appear at the center of the shadow cast by a circular disk? Merely to label such shadowy probabilities "propensities" does not confer on them any objective reality (see Probability).

In a study of the original sources bearing on the acceptance of eight major theories, Stephen Brush (1994, 140) flatly declares that in no case was the theory accepted, "primarily because of its successful prediction of novel facts." About quantum mechanics, he writes (136) that "confirmation of novel predictions played actually no role in its acceptance," that instead its advocates "argued that quantum mechanics accounted at least as well for the facts explained by the old theory, explained several anomalies that its predecessor had failed to resolve, and gave a simple method for doing calculations in place of a collection of *ad hoc* rules" (137). Brush even turns the tables in contending that retrodiction often counts more in a theory's favor than novel predictions, that for example, Einstein's general theory of relativity was more strongly supported by the previously known advance of the perihelion of Mercury than by the bending of light in the gravitational field of the Sun (138). The main reason he offers is that Mercury's perihelion advance was a *recognized anomaly.* The failure (up to 1919) to account for it in Newtonian terms was strong indication that no such explanation would be forthcoming, and so, in Bayesian terms, the effect was essentially "inexplicable" in other theories, while the general theory of relativity was able to account for it with quantitative precision.

Problems of Induction

James Bernoulli recognized that if one is to discover a population proportion *q* of some trait *Q* empirically—say, the proportion of 65-year-olds who survive 10 years or more—then the proportion q_n of Qs in a random sample of n (drawn with replacement) must approximate q arbitrarily well with "moral certainty" for sufficiently large samples. Bernoulli's elegant proof yields an upper bound on the least sample size, n_0, for which the sum C_n of the "central" binomial probabilities satisfying $|q_n - q| < \frac{1}{r+s}$ with $q:(1-q) = r:s$ exceeds $c(1 - C_n)$ whenever $n \geq n_0$. Notice, $\epsilon = \frac{1}{(r+s)}$ can be made arbitrarily small without changing the ratio, $r:s$ and that $C_n > c\,(1 - C_n)$ if and only if $C_n > \frac{c}{(1+c)}$. Bernoulli's rather loose bound was subsequently improved by his nephew, Nicholas, who interested de Moivre in the problem, leading the latter to his discovery of the normal approximation to the binomial.

The limitations of his approach notwithstanding, James Bernoulli had taken a major step toward quantifying uncertainty and deriving frequency implications from assumed probabilities. Yet, his actual goal of justifying and quantifying inductive inference eluded him and his followers, including de Moivre. Even though the statement that q_n lies within ϵ of q holds if and only if q lies within ϵ of q_n, it *does not follow* that these two statements have the same probability. The probability of the former is just a sum of binomial probabilities with q and n given, but there is no such simple way of computing the probability that an observed sample proportion lies within ϵ of the *unknown* population proportion. For all one knows, that sample may be quite deviant or unrepresentative.

While it is uncertain whether Bernoulli fell prey to this rather seductive fallacy, it is certain that Bayes, who had detected other subtle fallacies of this kind (Hald 1998, 134; Stigler 1986, 94–95), did not. He realized that a distinctively inductive inference is required. In fact, Bayes posed and solved a far more difficult problem than that of "inverting" Bernoulli's weak law of large numbers, *viz.,* given the number of Qs and non-Qs in a sample *of any size,* to find the probability that q "lies…between any two degrees of probability that can be named," i.e., in *any* subinterval of [0,1], when nothing is known about the population before sampling. His solution appeared in his *Essay,* published posthumously by his friend Price in 1763.

Bayes offered a subtle argument. He equated the uninformed state of prior knowledge ("ignorance" of q) with one in which all outcomes of n Bernoulli trials were equiprobable, whatever the value of n.That condition holds if the prior density, $w(q)$ of q is the *uniform density,* $w(q) = 1$, which assigns equal probability to intervals of equal length. Bayes tacitly assumed that conversely, the *only* density of q meeting this condition is the uniform density. The central moments of the uniform density are:

$$E(q^n) = \int_0^1 q^n w(q)dq = \frac{1}{n+1}$$

when $w(q) = 1$. Then, because the central moments uniquely determine a density that is concentrated on a finite interval (see de Groot 1986, Sec. 4.4), Bayes's assumption is proved correct. Later critics, among them Boole, Venn, and Fisher, all overlooked Bayes's criterion of ignorance, which is immune to the charge of inconsistency they leveled at "Bayes's postulate."

Bayes's solution of the "inverse problem" now required one more step, the extension of the partitioning formula, equation 2, to continuous parameters. With B = [k successes in n trials] and $A = [t_0 \leq q \leq t_1]$, the probability Bayes sought was $P(A \mid B \wedge I_0)$ computed as $P(A \wedge B \mid I_0)/P(B \mid I_0)$ with I_0 being the assumed model of Bernoulli trials and the "empty" state of knowledge about the parameter. Then, replacing the sum in equation 2 with an integral, Bayes found for $w(q) = 1$:

$$P(B \mid I_0) = \int_0^1 \binom{n}{k} q^k (1-q)^{n-k} w(q)dq = \frac{1}{n+1}.$$

He then found that:

$$P(A \mid B \wedge I_0) = \frac{(n+1)!}{k!(n-k)!} \int_{t_0}^{t_1} q^k (1-q)^{n-k} dq. \quad (3)$$

This method works, however, only for small samples, and so Bayes devoted the remainder of the *Essay* to approximating the solution of the more general case—a formidable undertaking (see Stigler 1986, 130ff; Hald 1998, Sec. 8.6). His ongoing work on this problem was the main cause of the delay in publishing the *Essay,* and not, as some have alleged, misgivings about his formalization of ignorance (Stigler 1986, 129).

Bayes's solution thus incorporated three highly original extensions of the probability theory he inherited from the Bernoullis and de Moivre: *first,* his rule (equation 1a) for updating a probability assignment; *second,* his novel criterion of ignorance leading to a uniform prior density of the unknown population proportion; and *third,* the extension of equation 2 to continuous parameters.

Price was fully cognizant of the relevance of the *Essay,* not only to Bernoulli's problem of justifying induction, but to the skeptical arguments Hume mounted against that possibility in his *Treatise of Human Nature* (1739) and even more emphatically in *Enquiry Concerning Human Understanding* (1748) (see Empiricism; Induction, Problem of). Hume contended that one's expectation of future successes following an unbroken string of successes is merely the product of "custom" or habit and lacks any rational foundation. In his *Four Dissertations* of 1767, Price writes:

> In [Bayes's essay] a method [is] shown of determining the exact probability of all conclusions founded on induction .... So far is it from being true, that the understanding is not the faculty which teaches us to rely on experience, that it is capable of determining, in all cases, what conclusions ought to be drawn from it, and what precise degree of confidence should be placed in it.

Possibly, Bayes, too, was motivated in part by the need to answer Hume, but all that is known of his motivation is what Price says in the introduction to the *Essay*. Price also contributed an appendix to the *Essay*, in which he drew attention to the special case of equation 3 in which $k = n$ and found that the probability was

$$(n+1)\int_{.5}^{1} q^n dq = \frac{2^{n+1}-1}{2^{n+1}}$$

that q lay between $t_0 = .5$ and $t_1 = 1$ He also showed that the probability tended to 1 and that q lay in a small interval around q_n. Price also moved into deeper waters with his suggestion that the inverse probability engine kicks in only after the first trial has revealed the existence of Qs (see Zabell 1997, 363–369, for more on Price's intriguing appendix).

It is curious that neither Bayes nor Price considered predictive probabilities per se. It was left to Laplace to take this next step, in 1774. He did this in a natural way by equating the probability of success on the next trial, following k successes in an observed sequence of n, with the mean value of the posterior density of q:

$$\begin{aligned} &P(X_{n+1} = 1 \mid B_n = k, I_0) \\ &= \frac{(n+1)!}{k!(n-k)!}\int_0^1 q^{k+1}(1-q)^{n-k}dq \\ &= \frac{(n+1)!}{k!(n-k)!}\frac{(k+1)!(n-k)!}{(n+2)!}. \end{aligned}$$

This simplifies to:

$$P(X_{n+1} = 1 \mid B_n = k, I_0) = \frac{k+1}{n+2}, \qquad (4)$$

or Laplace's *law of succession*, which specializes to $P(X_{n+1} = 1 \mid B_n = n, I_0) = \frac{n+1}{n+2}$ when $k = n$. Notice that equation 4 does not equate the probability of success on the next trial with the observed relative frequency of success, the *maximum likelihood estimate* of q. That rule would accord probability 1 to success on the next trial following an observed run of n successes even when $n = 1$. In Bayesian terms, that is tantamount to prior knowledge that the population is *homogeneous,* with all Qs or all non-Qs.

Laplace went beyond the derivation of equation 4 in later work (see Hald 1998, Chaps. 9, 10, 15). His aim was to obtain predictive probabilities of all sorts of events by "expecting" their sampling probabilities against the posterior distribution based on an observed outcome sequence. In particular, the probability, $P(c \mid m, a, n)$, of c Qs and $d = m - c$ non-Qs in the next m trials given a Qs and $b = n - a$ non-Qs observed in $n = a + b$ trials is:

$$\begin{aligned} P(c \mid m, a, n) = &\frac{(a+b+1)!}{a!b!}\binom{m}{c} \\ &\int_0^1 Q^c(1-Q)^d Q^a (1-Q)^b dq. \end{aligned} \qquad (5)$$

Laplace approximated equation 5 by

$$\begin{aligned} &P(c \mid m, a, n) \approx \\ &\binom{m}{c}\frac{(a+c)^{a+c+\frac{1}{2}}(b+d)^{b+d+\frac{1}{2}}n^{n+\frac{1}{2}}}{a^{a+\frac{1}{2}}b^{b+\frac{1}{2}}(n+m)^{n+m+\frac{1}{2}}}\frac{n+1}{n+m+1}. \end{aligned} \qquad (5a)$$

When c and d are small compared with a and b, so that $m \ll n$, this simplifies further to

$$P(c \mid m, a, n) \approx \binom{m}{c}(a/n)^c(b/n)^d, \qquad (5b)$$

the sampling probability, $\binom{m}{c}Q^c(1-Q)^d$, with the observed sample proportion, $Q_n = a/n$, in place of q.

He even showed that Q is asymptotically normally distributed about the observed sample proportion $h = q_n$, with variance hk/n, $k = 1 - h$ (Hald 1998, 169–170), hence that for any $\epsilon > 0$, the posterior probability

$$P_\epsilon = (n+1)\binom{n}{a}\int_{h-\epsilon}^{h+\epsilon} Q^a(1-Q)^b dQ \to 1$$

that $h - \epsilon \le q \le h + \epsilon$ approaches 1 as $n \to \infty$. This is the counterpart to the inverse probability of Bernoulli's weak law of large numbers. Finally, Laplace showed that the mode of equation 5 is the integer part, $\lfloor (m+1)Q_n \rfloor$, of $(m+1)Q_n$. Thus, the most probable sample frequency in a second sample is the one closest to that observed in the first sample. This is, arguably, the high point of the early Bayesian response to Hume, delivering a precise sense in which one can *reasonably* expect the future to "resemble" the past when nothing is known beyond what the observed random sample conveys.

It is astonishing that in two centuries of discussion of the problem of justifying induction, scarcely any mention has been made of these fundamental results of Bayes and Laplace. The subsequent development of inductive logic consists mainly of generalizations of Laplace's rule (equation 4) (see Inductive Logic). (For objections to this rule, most of them predicated on ignoring the conditions of its validity, *viz.,* independent trials and prior ignorance of q, see Fisher 1956, Chap. 2; Jaynes 2003, 563–85.)

Consider, first, random sampling *without* replacement. Jaynes (2003) has given this rather shopworn topic a new lease on life and illustrated, at the same time, how to teach probability and statistics as a unified whole (Chaps. 4 and 6). Does equation 4 generalize to this case? Given an urn containing R red and $W = N - R$ white balls, write R_j [red on jth trial] for the event of drawing a red ball on trial j, $j = 1, \ldots, N$. Then the probability of red on trial 1 is $P(R_1 \mid B) = R/N$, writing B for the assumed background knowledge, while

$$P(R_2 \mid B) = P(R_2 \mid R_1 \wedge B)P(R_1|B) + P(R_2|W_1 \wedge B)P(W_1|B)$$
$$= \frac{R-1}{N-1}\frac{R}{N} + \frac{R}{N-1}\frac{N-R}{N} = \frac{R}{N},$$

the same as $P(R_1 \mid B)$. By mathematical induction, drawing red (respectively, white) has the same probability on *any* trial, prior, of course, to sampling.

Moreover, by the product rule, and using $P(R_k \mid B) = P(R_j \mid B)$:

$$P(R_j \mid R_k \wedge B) = P(R_k \mid R_j \wedge B) \qquad (6)$$

Thus, knowledge that a red ball was drawn on a later trial has the same effect on the probability of red as knowledge that a red ball was drawn on an earlier trial. (This also poses a barrier to any propensity interpretation of these conditional probabilities [Jaynes 2003, Sec. 3.2].) (see Probability)

Next:

$$P(R_1 \wedge W_2 \mid B) = \frac{R}{N}\frac{N-R}{N-1} = \frac{N-R}{N}\frac{R}{N-1}$$
$$= P(W_1 \wedge R_2 \mid B),$$

and an obvious extension of this shows that each sequence containing r red and w white has probability

$$\frac{R^{\underline{r}}(N-R)^{\underline{w}}}{N^{\underline{n}}} = \frac{R!(N-R)!(N-n)!}{(R-r)!(N-R-w)!N!} \qquad (7)$$

with $x^{\underline{k}} = x(x-1)\ldots(x-k+1)$, irrespective of the *order* in which the red and white balls are drawn. Such sequences of trials are called *exchangeable.* It follows, just as when sampling with replacement, that the probability $h(r \mid N, R, n)$ of drawing r red and w white balls in $n = r + w$ trials is obtained by multiplying equation 7 by $\binom{n}{r}$ to yield:

$$h(r \mid N, R, n) = \frac{n!}{(n-r)!r!}\frac{R!}{(R-r)!}\frac{(N-R)!}{(N-R-w)!}$$
$$\frac{(N-n)!}{N!} = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}. \qquad (8)$$

In this derivation of the familiar *hypergeometric* distribution, equation 8 brings out the many parallels between random sampling with and without replacement; above all, their common exchangeability.

Having dealt with the "direct" probabilities that arise in this connection, Jaynes (2003, ch. 6) then turns to the inverse problem, where $D = (n, r)$ is given (the data) and the population parameters (N, R) are both unknown. This is a far richer and more challenging problem than its binomial counterpart—Bayes's problem—since it involves two unknown parameters. (Indeed, it may well lie beyond the capabilities of orthodox statistics.) Here the import of the data is inextricable from the prior information, which may take many forms, and, in addition, if interest centers on R or the population proportion, R/N, N then enters as a well-named "nuisance" parameter, a real stumbling block for both orthodox and likelihood methods.

The joint posterior distribution, $P(N \wedge R \mid DI)$, may be written using the product rule as:

$$P(N \wedge R \mid DI) = P(N \mid I)p(R \mid N \wedge I)\frac{P(D \mid N \wedge R \wedge I)}{P(D \mid I)}.$$

Hence, the marginal distribution of N is given either by

$$P(N \mid D \wedge I) = \sum_{R=0}^{N} P(N \wedge R \mid D \wedge I)$$
$$= P(N \mid I)\sum P(R \mid NI)\frac{P(D \mid N \wedge R \wedge I)}{P(D \mid I)}$$

or directly from equation 1 by

$$P(N \mid D \wedge I) = P(N \mid I)\frac{P(D \mid N \wedge I)}{P(D \mid I)}.$$

Now it would be natural to assume that $D = (n, r)$ can do no more than eliminate values of N less than n, leaving the relative probabilities of those greater than n unchanged. The general condition on $p(R \mid N \wedge I)$ that the data say no more about N than to exclude values less than n is that

$$P(D \mid N \wedge I) = \sum_{R=0}^{N} P(D \mid N \wedge R \wedge I)P(R \mid N \wedge I) = f(n, r)$$

where $f(r, n)$ may depend on the data but not on N, or using equation 8, that

$$\sum \binom{R}{r}\binom{N-R}{n-r} P(R \mid N \wedge I) = f(n, r)\binom{N}{n}. \tag{9}$$

Thus, it took Bayes's rule to ferret out this condition, one that unaided intuition would never have discovered.

As the condition is commonly met, Jaynes (2003) next turns his attention to the factor $p(R \mid DNI)$ in the joint posterior distribution of N and R, $P(N \wedge R \mid D \wedge I) = p(N \mid D \wedge I)p(R \mid D \wedge N \wedge I)$. By Bayes's rule (equation 1),

$$P(R \mid D \wedge N \wedge I) = P(R \mid N \wedge I)\frac{P(D \mid N \wedge R \wedge I)}{P(D \mid I)}.$$

To begin with, assume that I_0 is the state in which nothing is known about R beyond $0 \leq R \leq N$. Then the prior is uniform over this range, that is $p(R \mid N \wedge I_0) = \frac{1}{N+1}$. The posterior distribution of R is then

$$P(R \mid D \wedge N \wedge I_0) = \binom{N+1}{n+1}^{-1}\binom{R}{r}\binom{N-R}{n-r}. \tag{10}$$

From this, it is easy to show that Jaynes's condition, equation 9, is satisfied. For predictive purposes, what is needed is the posterior mean:

$$\langle R \rangle = E(R \mid D \wedge N \wedge I_0) = \sum_{R=0}^{N} RP(R \mid D \wedge N \wedge I_0).$$

Using equation 10, after much algebraic manipulation:

$$\langle R \rangle + 1 = (N+2)\frac{r+1}{n+2}, \tag{11}$$

which for large N, n, and r is close to the mode, $R' = (N+1)r/n$, of equation 10. Moreover, the expected fraction $\langle F \rangle$ of red balls left after the sample (n, r) is

$$\langle F \rangle = \frac{\langle R \rangle - r}{N-n} = \frac{r+1}{n+2}. \tag{11a}$$

Finally, the probability of drawing red on the next trial, given (n, r), is obtained by averaging the probability, $(R - r)/(N - n)$, of drawing red from the depleted urn against the posterior distribution:

$$P(R \mid DNI_0) = \sum_{R=0}^{N} \frac{R-r}{N-n}\binom{N+1}{n+1}^{-1}\binom{R}{r}\binom{N-R}{n-r}$$
$$= \frac{1}{N-n}[\langle R \rangle + 1 - (r+1)] = \langle F \rangle$$

whence

$$P(R_{n+1} \mid DNI_0) = \frac{r+1}{n+2}, \tag{11b}$$

the same result Laplace obtained for sampling with replacement. The rediscovery of equation 11b by Broad in 1918 and the surprise that it did not depend on N sparked a revival of interest in the mathematical analysis of inductive reasoning by Jeffreys, Johnson, Keynes, and Ramsey.

Jaynes goes on to consider other priors for sampling an urn, but the basic conclusion is already apparent. The import of the data depends on the prior information; the two are inextricable. Hence, different priors may or may not lead to different inferences from the same data.

A path to the treatment of more substantial prior knowledge begins by imagining that this prior information comes from a pilot study. In the binomial case Bayes treated, a pilot sample issuing in a successes and b failures could be pooled with a subsequent experiment issuing in r successes and $s = n - r$ failures to yield a posterior beta density,

$$f_\beta(q \mid a+r+1, b+s+1)$$
$$= B(a+r+1, b+s+1)^{-1}q^{a+r}(1-q)^{b+s}$$

starting from Bayes's uniform prior, which is $f_\beta(q \mid 1, 1)$. And this is, of course, the same posterior density obtained from the beta prior

$$f_\beta(q \mid a+1, b+1) = \frac{(a+b+1)!}{a!b!}q^a(1-q)^b,$$

which is the posterior density obtained from the pilot sample. (Jaynes has dubbed this property of Bayesian inference "chain consistency.")

Given, therefore, r successes in n Bernoulli trials, the posterior density, $f_\beta(q \mid a+r, b+n-r)$, for a beta prior, $f_\beta(q \mid a, b)$, yields a posterior mean or predictive probability of

$$\frac{a+r}{a+b+n}$$

and, in the case $a = b$ of a symmetric beta prior, with mean, mode, and median all equal to 1/2, this becomes

$$P(R_{n+1} \mid n, r) = E_{\beta}(Q \mid a + r, b + n - r) = \frac{a + r}{2a + n}$$

or, putting $\lambda = 2a$,

$$P(R_{n+1} \mid n, r) = \frac{r + \frac{\lambda}{2}}{n + \lambda}. \qquad (12)$$

Again, one gets a weighted average,

$$\left[n\frac{r}{n} + \lambda\frac{1}{2} \right] / (n + \lambda)$$

of the sample proportion and the prior expectation of 1/2, the weights $n/(n + \lambda)$ and $\lambda\ /\ (n + \lambda)$ reflecting the relative weights of the sample information and the prior information. (Here the separation is clean.) Thus, one arrives at a whole new continuum of rules of succession, which Carnap (1952) dubbed the *λ – continuum*. Laplace's rule is included as the special case $\lambda = 2$. Unfortunately, Carnap, who sought the holy grail of a universally applicable rule of succession, wrote as if one must choose a single value of λ for life, as a function of logical and personal considerations, like risk averseness, while the derivation from symmetric beta priors shows that different choices of λ merely correspond to different states of prior knowledge of q, or to different beliefs about the uniformity of the relevant population.

The uniqueness of the beta prior, $f_{\beta}(q \mid a, a)$, that corresponds to the *λ – rule* with $\lambda = 2a$ follows from the uniqueness part of de Finetti's celebrated *representation theorem*. Recall, binary random variables, $X_1, \ldots, X_n$, are *exchangeable* if their joint distribution is permutation invariant:

$$P(X_1 = e_1, \ldots, X_n = e_n) = P(X_1 = e_{\sigma(1)}, \ldots, X_n = e_{\sigma(n)})$$

for any permutation σ of {1, 2,..., n}. An infinite sequence $X_1, \ldots, X_n, \ldots$ is exchangeable if every finite subsequence of length n is exchangeable for all $n \geq 1$. This nails down the idea that the probability of any finite (binary) outcome sequence depends only on the number of 1's it contains and not the particular trial numbers on which they occur. This is manifestly true of binomial outcome sequences, as well as probability mixtures of exchangeable sequences. De Finetti's theorem is a strong converse, affirming that every infinite exchangeable sequence is a mixture of binomials, so that if $S_n = X_1 + \ldots + X_n$, then there is a *unique* distribution F of q such that for all n and k,

$$P(S_n = k) = \int_0^1 \binom{n}{k} q^k (1 - q)^{n-k} dF(q). \qquad (13)$$

Notice, one and the same F works for all n and k. If F is continuous (admits a density), equation 13 becomes:

$$P(S_n = k) = \int_0^1 \binom{n}{k} q^k (1 - q)^{n-k} f(q) dq. \qquad (13a)$$

An immediate corollary is that if $P(S_n = k) = \frac{1}{n+1}$ for all n and k, then the uniform density, for which

$$\frac{1}{n+1} = \int_0^1 \binom{n}{k} q^k (1 - q)^{n-k} dq$$

holds for all n, and k must be the *only* one for which Bayes's criterion of ignorance holds. Next, assuming that only the trials to which the *λ – rules* (equation 12) apply are exchangeable, it follows that the corresponding beta density is the only one that yields the *λ – rule*, being the "mixer" in equation 12a. To see this, note first that for any exchangeable sequence satisfying equation 12, $P(R_1) = P(W_1) = 1/2$.

Then if, say, $\lambda = 4$, so that $a = 2$, de Finetti's theorem affirms that for a unique distribution, F,

$$P(W_1 \wedge R_2) = \int_0^1 q(1 - q) dF(q).$$

But

$$\begin{aligned} P(W_1 \wedge R_2) &= P(W_1)P(R_2 \mid W_1) \\ &= \frac{1}{2} \int_0^1 q f_{\beta}(q \mid 2 + 0, 2 + 1) dq, \end{aligned}$$

which simplifies to

$$\int_0^1 q(1 - q) f_{\beta}(q \mid 2, 2) dq.$$

An extension of this argument shows that for any n and k, a binary exchangeable sequence of length n and k 1's has probability

$$\int_0^1 q^k (1 - q)^{n-k} f_{\beta}(q \mid 2, 2) dq.$$

Hence, the unique mixer of equation 13a can only be $f_{\beta}(q \mid 2, 2)$.

For $K > 2$ colors, the $\lambda - rules$ generalize to

$$P(X_{n+1} = i \mid (n_1, \ldots, n_K)) = \frac{n_i + \frac{\lambda}{K}}{n + \lambda} \qquad (12a)$$

with n_i the number of trials on which color i is drawn and $\sum_{i=1}^{K} n_i = 1$.

The corresponding prior densities—or mixers in the extension of de Finetti's theorem to this case—are the symmetric Dirichlet priors:

$$\frac{\Gamma(Ka)}{\Gamma(a)} p_1^{a-1} p_2^{a-1} \ldots p_K^{a-1}$$

where p_i is the population proportion of color i and $\sum p_i = 1$.

Again, this correspondence is one to one.

Exchangeability in the multicolored case means that two outcome sequences with the same vector, $(n_1, \ldots, n_k)$, of category counts are equiprobable. Thus, it makes no difference in which trials the different colors are drawn. Johnson generalized Bayes's criterion for $K = 2$ colors to the requirement that every "ordered K – partition" $(n_1, \ldots, n_k)$ of n is equiprobable, a much stronger condition than exchangeability. Later, he weakened it to require that the probability of color i in trial $n + 1$ depends only on n_i and n, not on the frequencies with which other colors are drawn, i.e.,

$$P(X_{n+1} = 1 \mid X_1 = i_1, \ldots, X_n = i_n) = f(n_i, n). \quad (14)$$

He then showed that $f(n_i, n)$ is given by equation 12a, provided $K > 2$. Carnap rediscovered all of this two decades later. Johnson's derivation of equation 12a was a milestone, for it casts into sharp relief the question of when and whether a symmetric Dirichlet prior adequately represents one's prior knowledge of the relevant categories. Readers unfamiliar with this material should try hard to think of convincing cases in which the number of times other categories occur does matter.

A severe limitation of the λ–*rules* is that no finite sample can raise the probability of a generalization that affirms the nonemptiness of a specified subset of the categories above zero if the population is infinite. Hintikka and Niiniluoto (1976) discovered that such confirmation becomes possible if Johnson's postulate (equation 14) is relaxed to permit dependence on the number, c, of kinds or colors observed. Their "representative function," $f(n', n, c)$, like Johnson's $f(n', n)$, is linear in n':

$$f(n', n, c) = \mu(n, c) \frac{n' + \frac{\lambda}{K}}{n + K - c + \lambda} \qquad (15)$$

where $\mu(n, c)$ does not depend on n' and $\lambda = \lambda(K)$ is given by

$$\lambda = \frac{Kf(1, K+1, K)}{1 - Kf(1, K+1, K)} - 1. \qquad (15a)$$

Thus, the probability that a color whose sample proportion is n'/n will turn up next increases with both n' and the number c of colors in the sample.

Indeed, the λ–*rules* enter the new family as a limiting case—the extreme of caution in generalizing. They satisfy

$$f(1, K-1) = \frac{1 + \frac{\lambda}{K}}{K - 1 + \lambda}$$

and Hintikka and Niiniluoto show that when

$$f(1, K-1, K-1) > \frac{1 + \frac{\lambda}{K}}{K - 1 + \lambda},$$

all predictions of the new system are more optimistic:

$$f(n', n, K-1) > \frac{n' + \frac{\lambda}{K}}{n + \lambda}.$$

More to the point, the posterior probability of the *constituent*, $C^{(K-1)}$, which affirms the nonemptiness of all but one of the possible kinds when all of these have occurred, is greater than 0 if and only if

$$f(1, n, K-1) > \frac{1 + \frac{\lambda}{K}}{n + \lambda} = f(1, n)$$

provided λ is a multiple of K. More generally, they show that when the parameters, $f(0, c, c)$, of the new system are chosen more "optimistically" than their Carnapian counterparts, the posterior probability of the simplest constituent, $C^{(c)}$, compatible with the sample approaches 1. The longer those "missing" kinds remain unsampled, the higher the probability that they do not exist.

This result, though qualitative, was an important clue to Rosenkrantz's Bayesian account of simplicity (Rosenkrantz 1977, Chap. 5), which says, roughly, that simpler theories are prized not because they are more probable a priori but because they are *more confirmable by conforming data.* A *simpler* theory is one that "fits" a smaller proportion of the possible outcomes of a relevant class of experiments or imposes stronger constraints on phenomena. If a theory T_2 *effectively* accommodates all of the possibilities that T_1 accommodates, along with many others besides, as with an ellipse of positive eccentricity (a *proper* ellipse) versus a circle, or a proper quadratic versus a linear polynomial, then T_1 is more strongly confirmed than a theory T_2 if what is observed is among the values or states of affairs allowed by T_1. Any account of simplicity and confirmation that failed to deliver this result would be a nonstarter. The main arguments Copernicus marshaled on behalf of heliocentrism nicely illustrate this Bayesian rationale (Rosenkrantz 1977, Chap. 7). But, of course, a simpler theory may be perceived as "too simple by half," as implausible—depending on the background knowledge. Few, if any, of Mendel's contemporaries would have credited (or did credit) his theory that inheritance is particulate and governed by simple combinatorial rules, much less Darwin's suggestion that all living species evolved from a single ancestral form.

There are other ways of modifying the Johnson-Carnap system. Exchangeable models often arise when the considered individuals are indiscernible, like electrons or copies of the same gene. By treating the categories as interchangeable, one arrives at a stronger concept, partition exchangeability, which requires the joint probability distribution of $X_1,\ldots,X_n$ to be invariant under both permutations of the indices (the trial numbers) and the category indices, 1, 2,$\ldots$,K. For example, for a die ($K = 6$), the probability of an outcome sequence of n flips would then depend only on the "frequencies of the frequencies," that is, on the a_r = the number of category counts, n_i, equal to r, so that, for example, the sequences 1,6,2,2,4,3,2,1,4,6 and 4,6,3,3,5,2,3,4,5,6 get the same probability, since they have the same *partition vector:*

$$(a_0, a_1, \ldots, a_K) = (1, 1, 3, 1, 0, \ldots, 0)$$

with $a_4 = a_5 = \ldots = 0$. Thus, one face is missing ($a_0 = 0$), one face turns up just once, three turn up twice, and one thrice, and *which* faces have these frequencies is immaterial. Notice, this definition would apply just as well if the number of possible categories were infinite or even indefinite (in which case a_0 is omitted).

Turing seems to have been the first to emphasize the relevance of these "abundancies," and after World War II, his statistical assistant, Good, published papers fleshing out this idea (see Good 1965, Chap. 8; Turing, Alan). If many kinds occur with roughly equal frequencies in a large sample with none predominant, then it seems likely that the relevant population contains many kinds (e.g., species of beetles), and one would expect to discover new ones. From this point of view, the relaxation of Johnson's postulate advanced by Hintikka and Niiniluoto was but a first step in limiting consideration to a_0 (through $c = n - a_0$), the first abundance (Zabell 1992, 218). "Ultimately," Zabell writes, "it is only partition exchangeability that captures the notion of complete ignorance about categories; any further restriction on a probability beyond that of category symmetry necessarily involves some assumption about the categories" (ibid.).

Zabell (1997) developed an imposing new system of inductive logic on this basis, in which the predictive probability of observing a species that has occurred n_i times in a sample of n depends only on n_i and n, as in Johnson and Carnap, but, in addition, the probability of observing a new species depends only on n and the number c of species instantiated in that sample, as in Hintikka and Niiniluoto. In addition, the trials are assumed to form an infinite exchangeable partition.

It would seem that these developments are far from Bayes and Price. However, in his appendix to Bayes's *Essay,* Price invited consideration of a die of an *indeterminate* number of sides; and de Morgan, an ardent Laplacian, also wrestled with the open-ended case where the possible species are not known in advance (Zabell 1992, 208–210). De Morgan constructed a simple urn model in which a ball is drawn at random from an urn containing a black "mutator" and t other balls of distinct colors and then replaced together with a ball of a new color if black is drawn and a ball of the same color otherwise. This led him to a rule of succession that may be considered an ancestral form of Zabell's rule. Finally, the results comprising Laplace's justification of induction were extended from Bernoulli sequences to exchangeable sequences by de Finetti ([1937] 1964, Chap. 3; [1938] 1980, 195–197).

Minimal Belief Change

Knowing nothing whatever about the horses in a three-way race, one's state of knowledge is unchanged by relabeling the entries. Hence, *mere consistency* demands that equal probabilities be assigned a given entry in these equivalent states, and the only distribution of probabilities invariant under all permutations of the horses' numbers or labels is, of course, the uniform distribution. In this manner, Jaynes (2003) has reinvented Laplace's hoary principle of indifference between events or "possibilities" as one of indifference (or equivalence) between problems. Namely, in two equivalent formulations of a problem, one must assign a given proposition the same probability. And two versions of a problem that fill in details left unspecified in the statement of the problem are *ipso facto* equivalent (Jaynes 1983, 144).

Suppose, next, that one of the entries is scratched. Provided no further information is supplied, it would be quite arbitrary to change one's relative odds on the remaining entries. If one's initial probabilities were p_1, p_2, and p_3, and horse 3 (H_3) drops out, the revised probabilities

$$\frac{P_1}{P_1+P_2} \qquad \frac{P_2}{P_1+P_2} \qquad 0$$

are merely the initial ones *renormalized.* One's partial beliefs undergo the *minimal* change dictated by the new information.

Suppose, instead, that one learns the relative frequency with which horse 2 (H_2) finished ahead of horse 1 (H_1) in a large sample of past races both entered. If nothing further is learned, one should be led to a constraint of the form

$$P_2 = rP_1 \qquad (*)$$

with $r > 0$. Call $r = \mathrm{o}(H_2 : H_1)$ the *revealed* odds for H_2 versus H_1, where H_i is the proposition that horse i will win tonight's race. Frequentists and Bayesians would agree, presumably, up to this point.

The question now is this: How should one revise one's probabilities on H_1 given $(*)$? Notice, one cannot conditionalize on $(*)$, for one cannot assign $(*)$ probabilities conditional on H_1. Still, one can hew to the principle of moving the prior "just enough" to satisfy $(*)$. So the question becomes: What effect should $(*)$ have on H_3?

Intuitions about this are somewhat conflicted. One might think that merely shifting the relative probabilities of H_1 and H_2 should have no bearing on H_3. But what if $r = 10^{10}$, whose effect would be to virtually eliminate H_1? Eliminating H_1 would increase the probabilities of the other two horses by mere renormalization, and by the same factor. A continuity argument then applies to say that $P(H_3)$ should increase *whatever* the value of r, approaching its maximum increase as $r \to \infty$. At the same time, one feels that $P(H_3)$ should not increase by as much as $P(H_2)$ when $r > 1$. That is about as far as unaided intuition can take one; it cannot quantify these qualitative relations.

Given conflicting intuitions, the only rational recourse is to seek compelling general principles capable of at least narrowing the range of choices. Such a narrowing is achieved in an interesting paper by van Fraassen, Hughes, and Harman (1986).

Rewrite the initial probability vector, (P_1, P_2, P_3), as (1, S, T), where $S = P_2{:}P_1$ and $t = P_3{:}P_1$ are the initial odds on H_2 and H_3 against H_1. Then $P_1 = 1/(1+S+T)$, $p_2 = S/(1+S+T)$, and $P_3 = T/(1+S+T)$. Clearly, the initial odds should not be altered if $R = S$, merely reinforcing the bettor's initial odds. Writing the updated odds vector as (1, R, $g(S, R, T)$), van Fraassen et al. (1986) developed an argument that entails $g(S, R, T) = T\gamma(S, R)$. Thus, the updating assumes the form

$$(1, S, T) \to (1, R, T\gamma(S,R)).$$

Next, van Fraassen et al. (1986) laid down a number of conditions on γ:

(i) $\gamma(S, R) = 1$ if $R = S$;
(ii) $\gamma(S, R)$ is continuous and $\lim_{r=\infty}\gamma(s,r) = r/s$;
(iii) γ should satisfy the functional equation, $\frac{S}{R}\gamma(S, R) = \gamma(\frac{1}{S}, \frac{1}{R})$.

To arrive at (iii), they ask what difference would it make if one's research had disclosed $P(H_1) = rP(H_2)$ instead of the other way around? "Really none," they answer. "It is the same problem as before," just as if the hypotheses have been relabeled (458). This is just Jaynes's principle. If the relabeling is carried out consistently, then the probabilities of H_3 in these two equivalent versions are:

$$\frac{T\gamma(S,R)}{1+R+T\gamma(S,R)} \quad \text{and} \quad \frac{\frac{T}{S}\gamma(\frac{1}{S},\frac{1}{R})}{1+\frac{1}{R}+\frac{T}{S}\gamma(\frac{1}{S},\frac{1}{R})}$$

since the initial and updated odds are now $1/s$ and $1/r$ for H_2 over H_1 and t/s for H_3 over H_1. After a little simple algebra, equating these two expressions for $P(H_3)$ then yields the functional equation of (iii).

Van Fraassen et al. (1986) introduce three rules that satisfy their conditions, whose representative functions are:

$$\mathrm{MUD} : \gamma(S,R) = \max(1, r/s);$$

$$\mathrm{MRE} : \gamma(S,R) = \left(\frac{R}{S}\right)^{\frac{R}{R+1}}; \text{ and}$$

$$\mathrm{MTP} : \gamma(S,R) = \left(\frac{1+R}{1+\sqrt{RS}}\right)^2.$$

It is easy to verify that (i)–(iii) hold for all three rules. MRE makes the probability of H_3 grow at an intermediate rate: faster than MTP but slower than MUD. The burden of van Fraassen et al. (1986) is to argue that (i)–(iii) exhaust "the symmetries of the problem," hence the conditions one can reasonably impose; and since these three rules all meet their conditions, "the problem does not admit a unique solution" (453). They further support this conclusion with two sorts of empirical comparison and find that MRE "is not the best on either count" (453).

Their argument is quite persuasive. But is the discouraging conclusion they draw inescapable? The first thing one notices about MTP is that it is the special case, $m = 2$, of a continuum of rules:

$$\gamma_m(s,r) = \left(\frac{1+r^{2/m}}{1+(rs)^{1/m}}\right)^m$$

with $m > 0$ (real), all of which satisfy (i)–(iii). By parity of reasoning, van Fraassen et al. (1986) must concede that all of these rules are on an equal footing with MTP, MRE, and MUD, or any others satisfying (i)–(iii). Now it is easy to see that $\gamma_m(S,R) \to 1$ as $m \to \infty$. Thus, the "flat rule,"

$$(1,S,T) \to (1,R,T)$$

which leaves the probability of H_3 unchanged, is obtained as a limiting case of MTP. Recall, however, that the flat rule is the very one that van Fraassen et al. (1986) ruled out on grounds of continuity. Hence, one may exclude MTP on the very same grounds.

Consider, next, the case $R < S$ where the initial odds, $S = o(H_2 : H_1)$, overshoot the revealed odds. In this case, MUD produces

$$(1,S,T) \to (1,R,T)$$

leaving the odds, $o(H_3 : H_1)$, unchanged, even when $R < 1$. Again, when $S = T < R$, MUD yields:

$$(1,S,T) \to (1,R,R)$$

which raises the odds on H_3 against H_1 as much as H_2 does. This, too, represents extreme inductive behavior. MRE approaches this result as the limit of $r \to \infty$.

The constraint $(*)$ is a very special case of a linear constraint, $\sum a_k P_k = 0$, or its continuous counterpart. Let an abstract rule of belief change operate on such a constraint, C, and a predistribution, P°, to produce a postdistribution, $P = P^\circ \circ \mathrm{C}$.

As linear constraints are satisfied by mixtures, $\alpha P + (1-\alpha)R$, of distributions P and R that satisfy them, one can best view the constraint C as a *convex* set of distributions. Members of this set will be called "class-C" distributions.

Next, impose conditions on such a rule all of which can be interpreted as saying that *two ways of doing a calculation must agree.* The conditions are that the results should (1) be unique; (2) not depend on one's choice of a coordinate system; (3) preserve independence in the prior when the constraint implies no dependence; and (4) yield the same conditional distribution on a subset whether one applies the relevant constraint to the prior on that subset or condition the postdistribution of the entire system to that subset.

Shore and Johnson (1980) show that the one and only rule that satisfies these broader consistency conditions goes by minimizing the deviation from P° among all class-C distributions, with the deviation between distributions P and Q given by the *cross entropy:*

$$H(P,Q) = \sum p_i \ln(p_i/q_i). \tag{16}$$

Applied to the horse race problem, this rule of minimizing cross entropy (*MINXENT*) becomes MRE. Using the inequality, $\ln x \le x - 1$, with equality if and only if $x = 1$, one shows that $H(P,Q) \ge 0$ with equality if and only if $P = Q$. Then using the convexity of $g(x) = x \ln x$, one shows that $H(P,Q)$ is convex in its first argument:

$$H(\alpha P + (1-\alpha)R, Q) \le \alpha H(P,Q) + (1-\alpha)H(R,Q) \tag{17}$$

with strict inequality if $P \ne R$ and $0 < \alpha < 1$.

Now suppose there are distinct class-C distributions P, R, which both minimize the distance to Q. Then:

$$H(P,Q) = H(R,Q) = \alpha H(P,Q)$$
$$+(1-\alpha)H(R,Q) > H(\alpha P + (1-\alpha)R, Q)$$

by convexity. Thus, the $\alpha -$ *mixture* of P and R has a smaller deviation from Q than either P or R. This contradiction establishes *uniqueness*. Moreover, a "nearest" distribution to P° among class-C distributions *exists,* provided C is closed under limits. Notice, too, that any belief rule that goes by minimizing a function, $I(x,y)$, satisfying $I(x,y) \ge 0$ with equality if and only if $x = y$ will have the *redundancy property:* $P \wedge C = P$ if $P \in C$ (i.e., if P already satisfies the constraint).

Minimizing cross entropy with respect to a uniform distribution,

$$H(P,U) = \sum P_i \ln(P_i/n^{-1}) = \sum P_i \ln P_i + \ln n$$

is equivalent to maximizing the (Shannon) entropy,

$$H(P) = -\sum P_i \ln P_i \tag{18}$$

a measure of the uncertainty embodied in the distribution P. The rule of maximizing the entropy (*MAXENT*), as mentioned in the introductory remarks, has also vastly expanded the arsenal of Bayesian statistics and modeling. To illustrate the third axiom governing independence, consider the alternative rule that goes by minimizing the repeat rate (RR), $\sum P_i^2$.

Like entropy, it is a continuous function of $P1$ that assumes its extreme values of $1/n$ and 1 at the extremes of uniformity and concentration. The RR is often considered a good approximation to (negative) entropy, and Table 2 shows how

Table 2 Maximum entropy and repeat rate

	1	2	3	4	5	6
Maxent	.103	.123	.146	.174	.207	.247
Repeat rate	.095	.124	.152	.181	.209	.238

closely the distribution of a die of mean 4 obtained by minimizing $\sum p_i^2$ approximates the maxent distribution obtained by maximizing the entropy.

A superficial examination might lead one to suppose that the RR rule performs about as well as MAXENT, just as van Fraassen et al. (1986) concluded that their rules performed as well as MINXENT. However, it can be shown that RR is inconsistent.

Consider, too, the more general family of rules that minimize a Csiszar divergence:

$$H_f(P, Q) = \sum q_i f(p_i/q_i)$$

with f being a convex function. This family includes MINXENT as the special case $f(x) = x \ln x$, as well as the chi-squared rule that minimizes $\sum \frac{(p_i - q_i)^2}{q_i}$, given by $f(x) = (x-1)^2$.

There is much at stake here for Bayesian subjectivists, who wish to deny that prior information can ever single out one distribution of probabilities as uniquely reasonable. But if one grants that MINXENT is a uniquely reasonable way of modifying an initial distribution in the light of linear constraints, then its special case, MAXENT, singles out a prior satisfying given mean value constraints as uniquely reasonable. Alive to this threat, subjectivists have denied any special status to MINXENT or MAXENT, as in the paper by van Fraassen et al. (1986).

The upshot of Shore and Johnson's derivation is to validate Jaynes's 1957 conjecture that "deductions made from any other information measure will eventually lead to contradictions" (Jaynes 1983, 9). It places MINXENT on a par with Bayesian conditionalization, given Cox's demonstration that there is no other consistent way to update a discrete prior. In addition, MINXENT yields Bayes's rule as a special case (Williams 1980).

MAXENT also has a frequency connection (Jaynes 1983, 51–52). Of the k^N outcome sequences of N trials, the number that yields category counts $(n_1, \ldots, n_k)$ with $\sum n_i = N$ is given by the multinomial coefficient:

$$W = \frac{N!}{n_1! \ldots n_k!}.$$

Using Stirling's approximation to the factorial, one easily proves that

$$N^{-1} \ln W \rightarrow H(f_1, \ldots, f_k). \tag{19}$$

Hence, the MAXENT distribution is realized by the most outcomes. In fact, the peak is enormously sharp. Just how sharp is quantified by Jaynes's concentration theorem (1983, 322), which allows one to compute the fraction of possible outcome sequences whose category frequencies, f_i, have entropy in the range $H_{max} - \Delta H \leq H(f_1, \ldots, f_k) \leq H_{max}$, where H_{max} is the entropy of the maxent distribution.

Representing Prior Knowledge

A crucial part of the answer Efron (1986) offers to his question "Why isn't everyone a Bayesian?" is: "Frequentists have seized the high ground of objectivity." Orthodoxy has deemed the prior probability inputs needed for Bayesian inference as of no interest for science unless they are grounded in frequency data. Bayesian objectivists partially agree, as when Jaynes writes (1983, 117):

> Nevertheless, the author must agree with the conclusions of orthodox statisticians that the notion of personal probability belongs to the field of psychology and has no place in applied statistics. Or, to state the matter more constructively, objectivity requires that a statistical analysis should make use, not of anybody's personal opinions, but rather the specific factual data on which those opinions are based.

But that "factual data" need not be frequency data. It might comprise empirically given distributional constraints or the role a parameter plays in the data distribution. At any rate, it is clear that if the much-heralded "Bayesian revolution" is ever to reach fulfillment, the stain of subjectivism must be removed. For the attempts of Bayesian subjectivists to sweeten the pill are themselves rather hard to swallow.

The first coat of sugar is to draw a distinction between the "public" aspect of data analysis, the data distribution, $f(x \mid \theta)$, and the "personal" element, *viz.*, the investigator's beliefs about θ before sampling (Edwards, Lindman, and Savage 1965, 526). Subjectivism differs from orthodox statistics, in this view, only in its wish to *formally* incorporate prior beliefs into the final appraisal of the evidence. But, it is not always possible to cleanly separate the import of the data from one's prior beliefs. Inferences about a population mean are colored by one's beliefs about the population variance. Even Bayesians who sail under the flag of subjectivism routinely handle such so-called nuisance parameters by integrating them out of a joint posterior density, using either a prior chosen to represent sparse prior knowledge or one that is *minimally informative about the parameter(s) of interest.* (Priors based on these two principles usually turn out to be the same or nearly so.) Non-Bayesians, eschewing the formal inclusion of prior information, have no resources for dealing with the problem of nuisance parameters (Royall 1997, Chap. 7), and certainly not a single resource that has won anything like universal acceptance.

The second coat of sugar is the claim that application of Bayes's rule to the accumulating data will bring initially divergent opinions into near coincidence. "This approximate merging of initially divergent opinions is, we think, one reason why empirical research is called 'objective'" (Edwards et al. 1965, 523). Opinions will converge, however, only if the parties at variance, Abe and Babe, share the same prior information. For their posterior probabilities satisfy

$$P(H \mid D \wedge I) = P(H \mid I)\frac{P(D \mid H \wedge I)}{P(D \mid I)}$$

with $I = I_A$ for Abe and $I = I_B$ for Babe, and if, as often happens, $I_A \neq I_B$, it does not follow that

$$|P(H \mid D \wedge I_A) - P(H \mid D \wedge I_B| < | P(H \mid I_A) - P(H \mid I_B) | .$$

Given the dependence of evidential import on prior opinion, one instantly senses trolls lurking under this placid surface.

Jaynes (2003, Sec. 5.3) reveals their presence. He notes that opinions may diverge when the parties distrust each other's sources of information (which he sees as a major cause of "polarization")—e.g., let H be the proposition that a drug is safe, and D that a well-known pundit has pronounced it unsafe. Abe considers the pundit reliable, Babe considers him a fraud. Both assign $P(D \mid \bar{H}) = 1$ but differ in their likelihoods: $P(D \mid HI_A) = 0.99$ and $P(D \mid HI_B) = 0.01$. Hence, if both assign H a fairly high prior probability, say, 0.9, their posterior probabilities of 0.899 and 0.083 are now far apart instead of identical.

More surprising is that Abe and Babe may diverge even when they are in total agreement about the pundit's reliability, assigning $P(D \mid H \wedge I_A) = P(D \mid H \wedge I_B) = a$ and $P(D \mid H \wedge I_A) = P(D \mid H \wedge I_B) = b$.

Given priors of $x = P(H \mid I_A)$ and $y = P(H \mid I_B)$, their posterior probabilities are:

$$P(H \mid D \wedge I_A) = \frac{ax}{ax + b(1-x)} \quad \text{and}$$

$$P(H \mid D \wedge I_B) = \frac{ay}{ay + b(1-y)}.$$

The necessary and sufficient condition for divergence works out to be

$$ab > [ax + b(1-x)][ay + b(1-y)],$$

which is easily satisfiable, by $a = 1/4$, $b = 3/4$, $x = 7/8$, and $y = 1/3$. Thus opinions may be driven further apart even when the parties place the same construction on the evidence.

Subjectivists view probabilities as partial beliefs to be elicited by introspection or betting behavior, but they require that they be "coherent" (i.e., consistent with the probability calculus). But if the subjectivist theory "is just a theory of consistency, plain and simple" (Zabell 1997, 365), how can subjectivists *consistently* disavow Jaynes's principle of equivalence or the consistency requirement that two calculations permitted by the rules must agree? The former underwrites the derivation of *uninformed* and the latter that of *informed* probability distributions, be they prior distributions or sampling distributions.

Thus, MAXENT, which derives from the latter requirement, leads to informed priors when the prior information takes the form of empirically given distributional constraints. On one hand, this can lead to *consensus priors* for experts who agree on those constraints, or, at worst, to a narrowing of the field and a clearer identification of the remaining areas of disagreement. On the other hand, the empirical success of many of the most ubiquitous probability models, like exponential decay or the normal (Gaussian) law of errors, is best explained by their derivation as maxent distributions satisfying commonly given constraints (see Jaynes 2003, Sec. 7.6). Arguably, this far better explains why empirical research is called objective.

Given scanty prior information, all but the most extreme subjectivists concede that some probability distributions are less faithful representations of that state than others. However, they insist that the "empty state" of knowledge is an illusory abstraction devoid of meaning (Lindley 1965, 18).

What lies behind this is the belief that earlier attempts to represent "total ignorance" all founder on the alleged "arbitrariness of parameterization" (see Hald 1998, sec. 15.6, for the tangled history of this objection, as well as Zabell 1988, esp. Sec. 6). For example, suppose one assigns a uniform distribution to the volume (V) of a liquid known to lie between only 1 and 2, but then, being equally ignorant of $D = V^{-1}$, one assigns it a uniform density on the corresponding interval $[\frac{1}{2}, 1]$. Since equal intervals of V, like $[1, \frac{3}{2}]$ and $[\frac{3}{2}, 1]$, correspond to unequal intervals of D, namely, $[\frac{1}{2}, \frac{2}{3}]$ and $[\frac{2}{3}, 1]$, there is a contradiction.

But as Jeffreys first pointed out, a log-uniform distribution of V with density

$$p(V)dV = V^{-1}$$

is the *same* distribution of V^k, for any power of V. For V is log-uniformly distributed in $[a,b]$, so that

$$P(c \leq V \leq d) = \frac{\ln d - \ln c}{\ln b - \ln a}$$

if and only if $\ln V$ is uniformly distributed in $[\ln a, \ln b]$, whence the term "log-uniform," and then $\ln V^k = k \ln V$ is uniformly distributed in $[k \ln a, k \ln b] = [\ln a^k, \ln b^k]$, so that V^k is log-uniformly distributed in $[a^k, b^k]$. Jaynes provided the justification Jeffreys only hinted at by noting that a log-uniform distribution is the only one invariant under changes of scale, while the uniform distribution is the only one that is translation invariant (Jaynes 1983, 126). Therefore, if all one knows about μ and θ is that μ is a *location parameter* and θ a *scale parameter* of the data distribution, $f(x \mid \mu, \theta)$, which means that the latter can be expressed in the form

$$f(x \mid \mu, \theta) = h\left(\frac{x - \mu}{\theta}\right).$$

Then *mere consistency* forces one to assign both μ and $\ln\theta$ a uniform density on $(-\infty, \infty)$. In practice, of course, one does know the (albeit vague) limits between which they lie, and so, in cases where the data are also scanty and it matters, one truncates these improper uniform densities to make them *proper*. Indeed, Jaynes recommends that Bayesians make a habit of such truncation as a kind of safety device (2003, 487).

For teaching purposes, however, nothing can match the mathematical simplicity of the improper Jeffreys priors that lead in a few lines of routine calculation to a joint posterior density, $p(\mu, \theta \mid DI)$, of the mean and variance of a normal population given a random sample (Lindley 1965, Sec. 5.4). Then by "marginalization," i.e., integrating with respect to θ, one is led to the posterior density of the mean (and, similarly, to one for the variance). The orthodox ("sampling theory") approach arrives, though much more laboriously, at interval estimates for μ that are numerically indistinguishable from their Bayesian counterparts owing to the mathematical quirk that the nuisance parameter can (in this special case of sampling a normal population) be eliminated (see de Groot 1986, Secs. 7.3–7.4). The same numerical agreement obtains for Bayesian "credence intervals" and orthodox "confidence intervals" of a binomial p, despite their radically different interpretation (Jaynes 1983, 170–171). Hence, orthodox acolytes of "performance characteristics" are hardly in a position to reject these Bayesian solutions on grounds of their inferior performance. Indeed, use of the Jeffreys priors realizes R. A. Fisher's ideal of "allowing the data to speak for themselves."

The Bayesian solution extends to the two-sample problem (Lindley 1965, Secs. 6.3–6.4), but the orthodox solution extends only if the two variances are known or are known to be equal. Behrens and Fisher gave a solution for the case where the two variances are known to be unequal that has never found widespread acceptance in the orthodox community but that follows rather easily from Bayes's rule. Moreover, in a definitive Bayesian treatment of this whole nexus of problems, Bretthorst *smoothly* extends the orthodox solutions of these two cases (variances known to be equal, known to be unequal) to a weighted average of their corresponding posterior densities, the weights being, of course, the respective posterior probabilities of being in each case (Bretthorst, 1993). Hence, a continuity argument comes into play here as well.

This "Bayes equivalence" of orthodox methods fails, however, whenever the latter are not based on *sufficient statistics,* functions of the data that yield the same posterior probability as the raw data (Lindley 1965, Sec. 5.5; de Groot 1986, Sec. 6.7). Such orthodox solutions become unavoidable when the given data distribution admits no nontrivial sufficient statistics, as in the famous example of the Cauchy distribution.In such cases, Bayes's rule (equation 1) will automatically pick the best interval estimate of a parameter for the sample actually observed, while the orthodox statistician must average over all samples that *might* be observed. The result is that the confidence coefficients attached to orthodox interval estimates are systematically misleading, and one will be able to define "good" and "bad" classes of samples in which the actual probability of including the true value of the parameter is better or worse than indicated (see Jaynes 1983, Chap. 9, examples 5, 6; Jaynes 2003, Sec. 17.1).

There is also pathology here. One may view orthodox methods—confidence intervals, unbiased estimators, chi-squared goodness-of-fit tests, etc.—as assorted ad hoc *approximations* to their Bayesian counterparts, joined by no unifying theoretical thread. When the approximation is satisfactory, one may expect the orthodox solution to perform about as well as the Bayesian, but where it is not satisfactory, orthodox solutions either fail to exist or yield absurd results. For examples, see Jaynes (1983, Chap. 9; 2003, Chap. 17). Necessary conditions under which orthodox solutions will closely approximate Bayes solutions are: (i) that they be based on sufficient statistics, (ii) that no nuisance parameters enter, and (iii) that prior information be sparse.

One could sum up the salient differences between the two approaches as follows. First, Bayesian methods solve the standard problems of statistics more simply, for they avoid the (often difficult) steps of choosing a suitable statistic (as test statistic or estimator) and finding its sampling distribution (or an approximation to it). Second, Bayesians are able to pose and solve problems involving nuisance parameters that lie wholly beyond the reach of orthodox theory. Above all, MINXENT and MAXENT have enormously expanded the powers of statistical inference and probability modeling. Third, Bayesianism offers a unified approach to all the problems of scientific method, inference, and decision. In particular, in place of an assortment of ad hocs, it offers a unified approach to the "modeler's dilemma" of deciding when the improved accuracy that normally accompanies a complication of theory is enough to compensate for the loss of simplicity.

All of these virtues are characteristic of any alleged new paradigm in the process of supplanting an old and entrenched theory in those upheavals termed "scientific revolutions" (see Kuhn, Thomas; Scientific Revolutions). By viewing orthodox solutions as approximate Bayes solutions, the Bayesian approach is also able to delineate the conditions of their validity—another very characteristic feature of a new paradigm.

There is yet another arena in which the mettle of Bayesianism can be tested, for, like MAXENT, symmetry has empirical implications when it is made to yield a data distribution. By utilizing this connection to the real world, two more stock objections to Laplace's principle of indifference are transformed into further triumphs of Jaynes's principle.

Naive application of Laplace's principle leads one to assign equal probabilities to the hypotheses, h_j, that j is the first significant digit in a table of numerical data, like the areas of the world's largest islands or lakes. But empirical investigation reveals that the probabilities decrease from $j = 1$ to $j = 9$. Nothing was said about the scale units, however, and the implied scale invariance leads to a log-uniform distribution:

$$p_j = P(h_j) = \log_{10}(j+1) - \log_{10} j = \log_{10}(1 + j^{-1}) \tag{20}$$

where $j = 1, 2, \ldots, 9$. Thus, $P_1 = \log_{10} 2 = 0.301$, $P_2 = \log_{10} 3 - 0.301 = 0.176$, $P_9 = 1 - \log_{10} 9 = 0.046$.

Benford discovered this now-famous "law of first digits" in 1938 but failed to explain it. Its derivation as the unique scale-invariant distribution explains why it works for ratio-scaled data, but Benford found that it also works for populations of towns or for street addresses. The explanation lies in Hill's recent discovery that "base invariance implies Benford's law" (Hill 1995). That is, equation 20 is the only distribution invariant under change of the base $b > 1$ of the number system employed. Any other distribution would yield different frequencies when the scale or base is changed. Hill even derives a far-reaching generalization of equation 20 that applies to blocks of $d > 1$ digits, hence, by marginalization, to 2*nd*, 3*rd*,…, as well as to first digits. About this example one may ask: What is the relevant "chance mechanism" that produces equation 20? What frequency data have led to it?

Bertrand's chord paradox asks for the probability that a "random chord" of a circle of radius R will exceed the side, $s = \sqrt{3}R$, of the inscribed equilateral triangle. Depending on how one defines a random chord, different answers result, and Bertrand himself seems to have attached no greater significance to the example than that "la question est mal posée." Jaynes, however, has given the problem a physical embodiment in which broomstraws are dropped onto a circular target from a height great enough to preclude skill. Nothing having been said about the exact size and location of the circle, the implied translation and scale invariance uniquely determine a density:

$$f(r, \theta) = \frac{1}{2\pi r R} \tag{21}$$

for the center (r, θ) of the chord in polar coordinates. And with $L = 2\sqrt{R^2 - r^2}$ the length of a chord whose center is at (r, θ), the relative length $x = L/2R$ of a chord has the induced density

$$p(x)dx = \frac{x}{\sqrt{1 - x^2}}. \tag{21a}$$

Finally, since $L = \sqrt{3}R$ is the side length of the inscribed triangle, the probability sought is:

$$\int_{\sqrt{3}/2}^{1} p(x)dx = \frac{1}{2}\int_{0}^{1/4} u^{-1/2} du = \frac{1}{2}$$

with $u = 1 - x^2$.

All of these predictions of equation 21 can be put to the test (see Jaynes 1983, 143, for one such test and its outcome). In particular, equation 21 shows to which "hypothesis space" a uniform distribution should be assigned in order to get an empirically correct result, namely, to the linear distance between the centers of chord and circle. There is no claim, however, to be able to derive empirically correct distributions a priori, much less to conjure them out of "ignorance." All that has been shown is that any other distribution must violate one of the posited invariances. If, for example, the target circle is slightly displaced in the grid of lines determined by a rain of broomstraws, then the proportion of "hits" predicted by that distribution will be different for the two circles. But if, as Jaynes argues (1983, 142), the straws are tossed in a manner that precludes even the skill needed to make them fall across the circle, then, surely, the thrower will lack the microscopic control needed to produce different distributions on two circles that differ slightly in size or location. Hence, one is tempted to view equation 21 as the "objective" distribution for this experiment.

Have these arguments then answered all of the arguments mounted against the possibility of objectively representing information or states of knowledge in the language of probability? The method indicated, group invariance, applies as readily to data distributions, as in the last two examples, as to prior distributions. (And, as Jaynes remarks, "One man's data distribution is another man's prior.")

Attention has been focused on invariance under a suitable group of transformations to the exclusion of all the many other approaches because this method is, in Jaynes's formulation of it, so closely tied to the consistency principle that "in two situations where we have the same state of knowledge, we must assign the same probabilities." This requirement may be said to underwrite all sound applications of symmetry to probability, answering, in effect, the main question addressed in Zabell (1988). *Exactly those symmetry arguments are sound that rest on Jaynes's reinvented principle of indifference.*

When the empirical probability distributions such symmetry arguments yield prove inaccurate, that may be taken as indication that some symmetry-breaking element is at work, just as the failure of a MAXENT distribution indicates the presence of some additional constraint forcing the data into a proper subset of the otherwise allowed possibility space. Hence, Jaynes concludes (2003, 326), "We learn most when our predictions fail," a theme also emphasized by Jeffreys. But to learn from our mistakes, we need to be sure that those failures are not mere artifacts of poor inductive reasoning; hence, the relevant inferences "must be our *best* inferences, which make full use of all the knowledge we have."

In conclusion, mention must be made of a nest of paradoxes of continuous probability, which, according to Jaynes, are mass-produced in accordance with the following simple prescription:

1. Start with a mathematically well defined problem involving a finite set or a discrete or normalizable distribution, where the correct solution is evident.
2. Pass to a limit without specifying how that limit is approached.
3. Ask a question whose answer depends on how that limit is approached.

He adds that "as long as we look only at the limit, and not the limiting process, the source of the error is concealed from view" (485). Under this head, Jaynes defuses the nonconglomerability paradoxes, the Borel-Kolmogorov paradox (for which see de Groot 1986, Sec. 3.10), and the marginalization paradoxes of Dawid, Stone, and Zidek aimed at discrediting improper priors. Jaynes proposes to block all such paradoxes, which have more to do with the ambiguities surrounding continuous probability than with prior distributions per se, by adopting a "finite sets policy" in which probabilities on infinite sets are introduced only as well-defined limits of probabilities on finite sets.

There has been discussion of alternative approaches to representing prior states of partial knowledge—what Jaynes has called "that great neglected half of probability theory." It might also be called the new epistemology. Notable contributors, apart from Jeffreys and Jaynes, include Box and Tiao, Bernardo, Novick, and Hall, and Lindley, Hartigan, and Zellner (see Zellner and Min 1993). Apart from the satisfaction of seeing that various approaches all lead, in many cases, to the same prior, like the Jeffreys log-uniform prior, one may expect the different methods to generalize in different ways when applied to harder, multiparameter problems. In any case, further work along these lines will undoubtedly contribute importantly to attempts to model expert opinion and heuristic reasoning in artificial intelligence or to arrive at consensus priors for policy decisions.

ROGER ROSENKRANTZ

References

Bayes, Thomas ([1763] 1958), "An Essay Towards Solving a Problem in the Doctrine of Chance." Originally published in the *Philosophical Transactions of the Royal Society* 53: 370–418. *Reprinted in Biometrika* 45: 293–315.

Bretthorst, G. L. (1993), "On the Difference in Means," in W. T. Grandy and P. W. Milonni (eds.), *Physics and Probabiity: Essays in Honor of Edwin T. Jaynes*. Cambridge: Cambridge University Press, 177–194.

Brush, Stephen (1994), "Dynamics of Theory Change: The Role of Predictions," in *Philosophy of Science* 2 (Proceedings): 133–145.

Carnap, Rudolf (1952), *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

Cox, R. T. (1946), "Probability, Frequency, and Reasonable Expectation," *American Journal of Physics* 17: 1–13. Expanded in Cox (1961), *The Algebra of Probable Inference*, Baltimore: Johns Hopkins University Press.

Dale, A. I. (1999), *A History of Inverse Probability*. New York: Springer-Verlag.

Diaconis, Persi, and David Freedman (1980), "De Finetti's Generalizations of Exchangeability," in Richard C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, vol. 2. Berkeley and Los Angeles: University of California Press.

de Finetti, B. ([1937] 1964), "La prevision: Ses lois logiques, ses sources subjectives." Translated into English in Henry Kyburg and Howard Smokler (eds.), *Studies in Subjective Probability*. New York: Wiley. Originally published in *Annales de l'Institut Henri Poincare* 7: 1–38.

——— ([1938] 1980), "Sur la condition d'equivalence partielle." Translated by Paul Benacceraf and Richard Jeffrey in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*. Berkeley and Los Angeles: University of California Press, 193–205. Originally published in *Actualites scientifiques et industrielles* 739 (Colloque Geneve d'Octobre 1937 sur la Theorie des Probabilites, 6tieme partie).

——— (1972), *Probability, Induction and Statistics*. New York: Wiley.

de Groot, Morris (1986), *Probability and Statistics*, 2nd ed. Reading, MA: Addison-Wesley.

Edwards, W., H. Lindman, and L. J. Savage (1965), "Bayesian Statistical Inference for Psychological Research," in R. Duncan Luce, R. Bush, and Eugene Galanter (eds.), *Readings in Mathematical Psychology*, vol. 2. New York: Wiley.

Efron, Bradley (1986), "Why Isn't Everyone a Bayesian?" *American Statistician* 40: 1–4.

Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

Giere, R. N. (1983), "Testing Theoretical Hypotheses," in John Earman (ed.), *Testing Scientific Theories*. Minneapolis: University of Minnesota Press.

Good, I. J. (1965), *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. Cambridge, MA: MIT Press.

——— (1983), *Good Thinking: The Foundations of Probability and its Applications*. Minneapolis: University of Minnesota Press.

Hald, A. (1998), *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley.

Hill, T. R. (1995), "Base Invariance Implies Benford's Law," *Proceedings of the American Mathematical Society* 123: 887–895.

Hintikka, Jaakko, and Ilka Niiniluoto ([1976] 1980), "An Axiomatic Foundation for the Logic of Inductive Generalization," in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*. Berkeley and Los Angeles: University of California Press, 157–182. Originally published in M. Przelecki, K. Szaniawski, and R. Wojcicki (eds.) (1976), *Formal Methods in the Methodology of Empirical Sciences*. Dordrecht, Netherlands: D. Reidel.

Hodges, J. L., and E. L. Lehmann (1970), *Basic Concepts of Probability and Statistics*, 2nd ed. San Francisco: Holden-Day.

Jaynes, E. T. (1983), *Papers on Probability, Statistics and Statistical Physics*. Edited by R. D. Rosenkrantz. Dordrecht, Netherlands: D. Reidel.

——— (2003), *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Jeffreys, H. ([1939] 1961), *Theory of Probability*, 3rd ed. Oxford: Clarendon Press. Jeffrey, Richard (ed.) (1980), *Studies in Inductive Logic and Probability*, vol. 2. Berkeley and Los Angeles: University of California Press.

Kullback, S., *Information Theory and Statistics*, New York: John Wiley.

Lehmann, E. L. (1959), *Testing Statistical Hypotheses*. New York: John Wiley.

Lindley, Dennis (1965), *Introduction to Probability and Statistics*, pt. 2: *Inference.* Cambridge: Cambridge University Press.

Loredo, T. J. (1990), "From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics," in Paul Fougere (ed.), *Maximum Entropy and Bayesian Methods*, 81–142. Dordrecht, Netherlands: Kluwer Academic Publishers.

Mayo, Deborah G. (1996), *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Popper, Karl (1995), *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper Torchbooks.

Rosenkrantz, R. D. (1977), *Inference Method and Decision*. Dordrecht, Netherlands: D. Reidel.

——— (1992), "The Justification of Induction," *Philosophy of Science* 59: 527–539.

Royall, Richard, M. (1997), *Statistical Evidence: A Likelihood Paradigm*. London: Chapman-Hall.

Shore, J. E., and R. W. Johnson (1980), "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross Entropy," *IEEE Transactions on Information Theory* IT-26: 26–37.

Smith, C. A. B. (1959), "Some Comments on the Statistical Methods Used in Linkage Investigations," *American Journal of Genetics* 11: 289–304.

Stigler, Stephen (1986), *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge, MA: Belknap Press of Harvard University, 1986.

van Fraassen, B., R. I. G. Hughes, and G. Harman (1986), "Discussion: A Problem for Relative Information Minimizers," *British Journal for the Philosophy of Science* 34: 453–475.

Zabell, S. L. (1988), "Symmetry and its Discontents," in Brian Skyrms and W. L. Harper (eds.), *Causation, Chance and Credence.* Dordrecht, Netherlands: Kluwer Academic Publishers, 155–190.

——— (1992), "Predicting the Unpredictable," *Synthese* 90: 205–232.

——— (1997), "The Continuum of Inductive Methods Revisited," in John Earman and John D. Norton (eds.), *The Cosmos of Science*. Pittsburgh: University of Pittsburgh Press, 351–385.

Zellner, Arnold, and Chung-ki Min (1993), "Bayesian Analysis, Model Selection and Prediction," in W. T. Grandy and P. W. Milonni (eds.), *Physics and Probability: Essays in Honor of Edwin T. Jaynes.* Cambridge: Cambridge University Press.

See also **Carnap, Rudolf; Confirmation Theory; Decision Theory; Epistemology; Induction, Problem of; Inductive Logic; Probability**

BEHAVIORISM

Behaviorism is regarded properly as a formal approach to psychology. The first to articulate systematically the tenets of the behaviorist approach was John B. Watson. In his essay "Psychology as a Behaviorist Views It," Watson (1913) attacked the predominant tendency to define psychology as the study of consciousness. His major target was Edward Titchener's structural psychology. Titchener (1898) had advanced psychology as the study of constituent elements of conscious states. This form of psychology favored introspection as the primary means to study the mind. To a lesser extent, Watson was also critical of functional psychology. Adherents of this view, such as James Angell (1907), placed emphasis on the biological significance of conscious processes. Critical of introspection and of the more general tendency to link the validity of psychological data to consciousness, Watson (1913) argued that psychology should be regarded as a purely objective experimental branch of natural science. He defined psychology as the prediction and control of behavior, and explicitly aligned psychology with the methods of physics and chemistry. He discarded conscious states and processes as the objects of observation and replaced them with stimulus–response connections.

Watson's views did not develop in isolation. (For a complete discussion of precursors to Watson, see O'Donnell 1985.) In the fist decade of the twentieth century, the Russian physiologist Ivan Pavlov advanced objectivity in psychology through research on the conditioned reflex. Pavlov demonstrated that a response that normally follows one particular stimulus could be produced by a different stimulus if the two stimuli were to occur together over a period of time. This technique was known as classical conditioning. Pavlov demonstrated the technique by inducing salivation in dogs through ringing a bell that had previously accompanied food. (A collection of English translations of Pavlov's major papers can be found in Pavlov 1955.) Pavlov's notion of classical conditioning became central to Watson's work.

After the publication of Watson's 1913 essay, behaviorism quickly became the mainstream approach in American psychology. The movement enjoyed immense popularity well into the 1950s through the work of such figures as Clark Hull (1943) and, more importantly, B. F. Skinner (1953). Skinner developed the notion of operant conditioning, which holds that behavior is shaped by its results, either positive or negative. Operant conditioning differed from the classical conditioning of Pavlov and Watson and was closely connected to the work of Edward Thorndike. In the early 1900s, Thorndike had formulated an approach known as connectionism. This approach held that learning consisted of connecting situations and responses, as opposed to connecting ideas. A central element in Thorndike's (1905) psychology was his *law of effect.* In essence, the law of effect held that the frequency with which a behavior occurred was related to the tendency the behavior had to produce positive or negative results. For instance, the likelihood that an animal will push on a lever is increased if doing so produces food, and is decreased if doing so produces pain. When incorporated into Skinner's concept of operant conditioning, the law of effect explained how new patterns of behavior emerge.

Skinner (1957) argued that higher cognitive processes, such as language, also could be treated as complex forms of operant behavior. In Skinner's view, verbal forms of expression differ from nonverbal forms only with respect to the contingencies that affect them. Even consciousness could be explained with reference to operant conditioning. According to Skinner, consciousness emerges as "stimulus control" designed to permit discriminations regarding one's own responding. Consciousness had often been considered a "private" or "first-person" event, but when treated as stimulus control, it was open to an operant analysis because it was placed in a functional relationship with the entire context of antecedent and consequent stimulation.

Behaviorism had profound philosophical implications. Prior to the rise of behaviorism, dualistic theories of mind enjoyed considerable popularity. In the philosophy of Descartes, for example, mind was interpreted as being of an essence that was distinct from matter. While philosophers such as Kant challenged such dualism, the views of behaviorists ultimately dealt dualism the most severe blow. Watson's attack on structural psychology, and on consciousness in particular, magnified the fact that consciousness was unobservable and, hence, outside the confines of science. Watson (1913) believed that mentality should be studied by focusing only on observable manifestations of the mind; namely, behavior. Skinner (1953) also argued against defining the mind in a way that located it in an unobservable realm. According to him, the goal of psychology was to describe the laws by which stimuli and behavior were connected, and the ways in which such connections were affected by changes in the physical environment. The laws governing these connections and their modifications were regarded by Skinner as being on a par with the laws of motion.

Skinner (1953) specifically argued that psychology could treat behavior as a function of the immediate physical environment and the environmental history. To say that behavior is a function of environmental history means that the way someone will behave in a given situation is determined by the physical stimuli to which that person has been exposed in the past. The implication of Skinner's view is that psychologists could eliminate entirely any reference to hidden entities and internal causes, leaving the concept of 'mind' to nonscientific forms of investigation. According to Skinner, this implication applied even to neural explanations. Neural factors could be eliminated, since they perform no function other than to describe behavior itself. To emphasize this point, Skinner presented what has become known as the theoretician's dilemma. If observable events are connected, no theory about internal mental states is needed because psychologists can predict one event from another without any reference to the theory. If the events are not connected, then the theory is not needed because it does not help make a prediction. (See Hempel 1958 for an early philosophical analysis of this problem.)

Behaviorist psychology became aligned closely with two interrelated philosophical movements. The first movement was operationalism. The goal of operationalism was to render the language and terminology of science more objective and precise and to rid science of those problems that were not physically demonstrable. The chief proponent of operationalism, Percy Bridgman (1927), held that a theoretical construct should be defined in terms of the physical procedures and operations used to study it. The implication for psychology is that a psychological construct is the same as the set of operations or procedures by which it is measured and used in experimentation. Since behaviorism eschewed consciousness and embraced instead observable activity, it corresponded neatly to operationalism (see Bridgman, Percy).

The second philosophical movement with which behaviorism was aligned was logical positivism. The positivist doctrine rested on the *verifiability theory of meaning*. The verifiability theory held that the meaningfulness of any scientific question was determined by asking whether there was observational evidence that could be collected to answer the question. Through the application of the verifiability theory, positivists held that the task of philosophy was to analyze the meaning of scientific statements in terms of the evidence that would confirm or disconfirm them. (A collection of essays on logical positivism can be found in Ayer 1959. See Logical Empiricism, Verifiability).

Behaviorism provided the positivists with the means for applying their project to psychology. Positivists argued that all descriptions of the mind were confirmed or disconfirmed solely by facts about a person's behavior in a given environment. According to the verifiability theory of meaning, these facts constitute the *meaning* of any psychological statement. This theory about the meanings of psychological statements became known as logical behaviorism. Carl Hempel (1949) defended logical behaviorism in his article "The Logical Analysis of Psychology." He argued that psychological statements that were meaningful, that is, verifiable, could be translated into statements that did not involve psychological concepts, but rather physical concepts. For example, statements about feelings of depression are meaningful because they can be translated into descriptions about the person's behavior and physical body. Hempel's position implied that meaningful statements of psychology were physicalistic statements and that proper psychology was an integral part of physics (see Hempel, Carl Gustav; Physicalism).

By the late 1950s, behaviorism faced serious criticisms. One major criticism corresponded to verbal behavior. Contrary to what Skinner (1957) had argued, it was not obvious that verbal behavior could be treated simply in terms of operant conditioning. Noam Chomsky (1959) argued that linguistic abilities could be explained only on the assumption that language was the result of complex mental processes that analyzed sentences into their grammatical and semantic components. In supporting this position, Chomsky pointed out that behaviorists wanted to claim that *all* behavior was a product of laws formulated in terms of responses to environmental stimuli. Yet, the only laws behaviorists were able to demonstrate arose in controlled experiments, usually with animals. In real situations in which natural language was used, Chomsky noted that psychologists could not know what the stimuli and environmental history were until a subject responded (see Chomsky, Noam).

Chomsky's (1959) own position assumed that stimuli must be described in terms of how the subject perceived them, rather than by objective physical characteristics. He was concerned that Skinner had failed to offer a characterization of the stimuli that permitted the connection of verbal behavior to objective physical properties of the environment. Chomsky argued that how a person behaved depended not only on the physical character of the stimulus but also on what was going on in the person's mind at the time the stimulus was presented. Chomsky consequently believed that psychologists could correctly predict what people would say only by considering their utterances to be the result of complex internal states of mind. This was particularly true of cases in which humans produced novel sentences. Many of the sentences that humans utter have never been produced before; hence there is no way that prior reinforcement would explain how they are learned.

Though some philosophers, such as Quine (1960), continued to defend it, eventually behaviorism fell out of favor (see Quine, Willard Van). It was replaced by an approach known as cognitivism. This approach placed greater emphasis on internal mental processing. Even though behaviorism is no longer a mainstream approach in psychology, recent scholarship (e.g., Thyer 1999) has attempted to highlight some potentially significant aspects of Skinner's views that may have been overlooked. This attempt rests on a distinction between Skinner's work and earlier behaviorism. In the earlier period, behaviorism was regarded as "methodological." Figures such as Watson and Hull considered behavior as important because it offered psychology epistemologically valid grounds for speaking about causal entities from a nonphysical dimension. Observable behaviors were means through which purely mental, or conscious, events could be studied scientifically. Skinner's behaviorism, on the other hand, was "radical." His approach treated behavior as a subject matter in its own right. Behavior was regarded as the interaction of organism and environment. Radical behaviorism held that mental events were appropriately regarded as aspects of the overall context, not as causes in a nonphysical dimension. Radical behaviorism consequently did not reduce behavior to physiological mechanisms. In comparison with methodological behaviorism, radical behaviorism may very well have unique consequences for major philosophical topics, including mind–body dualism, free will, and determinism.

ANDREW BACKE

References

Angell, James R. (1907), "The Province of Functional Psychology," *Psychological Review* 14: 61–91.

Ayer, A. J. (ed.) (1959), *Logical Positivism*. New York: Free Press.

Bridgman, Percy W. (1927), *The Logic of Modern Physics*. New York: Macmillan.

Chomsky, Noam (1959), "Review of *Verbal Behavior*, by B. F. Skinner," *Language* 35: 26–58.

Hempel, Carl G. (1949), "The Logical Analysis of Psychology," in H. Feigl and W. Sellars (eds.), *Readings in Philosophical Analysis*. New York: Appleton-Century-Crofts, 373–384.

——— (1958), "The Theoretician's Dilemma: A Study in the Logic of Theory Construction," *Minnesota Studies in the Philosophy of Science* 2: 37–98.

Hull, Clark L. (1943), *Principles of Behavior*. New York: Appleton.

O'Donnell, John M. (1985), *The Origins of Behaviorism: American Psychology, 1870–1920*. New York: NYU Press.

Pavlov, Ivan P. (1955), *Selected Works*. Moscow: Foreign Languages Publishing House.

Quine, Willard Van (1960), *Word and Object*. Cambridge, MA: MIT Press.

Skinner, Burrhus Frederic (1953), *Science and Human Behavior*. New York: Macmillan & Co.

——— (1957), *Verbal Behavior*. New York: Appleton.

Thorndike, Edward L. (1905), *The Elements of Psychology*. New York: A. G. Seiler.

Thyer, Bruce A. (ed.) (1999), *The Philosophical Legacy of Behaviorism*. London: Kluwer Academic Publishers.

Titchener, Edward B. (1898), "The Postulates of a Structural Psychology," *Philosophical Review* 7: 449–465.

Watson, John B. (1913), "Psychology as a Behaviorist Views It," *Psychological Review* 20: 158–177.

See also **Cognitive Science; Logical Empiricism; Mind–Body Problem; Physicalism; Psychology, Philosophy of; Verifiability**

BIOLOGICAL INFORMATION

Information is invoked by biologists in numerous contexts. Animal behaviorists examine the signaling between two organisms or attempt to delimit the structure of the internal map that guides an organism's behavior. Neurobiologists refer to the information passed along neurons and across synapses in brains and nervous systems. The way in which information terminology is used in these contexts has not so far been the main critical focus of philosophers of science. Philosophers of mind discuss animals' representation systems, such as bees' internal maps, and also focus on the way in which brains operate, with a view to shedding light on traditional problems in the philosophy of mind. In contrast, the focus of much discussion in philosophy of biology is the notion of information invoked to explain heredity and development: genetic information. The focus of this article will be on this latter form of biological information.

The ideas that genes are bearers of information and that they contain programs that guide organisms' development are pervasive ones, so much so in biology that they may seem hardly worth examining or questioning. Consulting any biology textbook will reveal that genes contain information in the form of DNA sequences and that this information provides instructions for the production of phenotypes. In contrast, an examination of the philosophical literature on biological information reveals that there are very few philosophers of biology who promote unqualified versions of either of these ideas. To understand how this situation has arisen requires first looking at the role that informational concepts play in biology.

Preliminaries

The two processes that are most relevant to the present context are evolution and development. There was much progress in conceptualizing evolutionary change when it was characterized in terms of changing gene frequencies in the 1930s and 1940s. Many evolutionary biologists discuss evolution entirely from a genetic perspective (see Evolution). After genes were established as the relevant heritable material, the next step was to conceptualize it in terms of molecular structure (see Genetics). In 1953 the structure of DNA was discovered and with this discovery came a mechanism for accounting for the duplication of heritable material and its transmission from one generation to the next. What the discovery of the structure of DNA also ushered in was a research focus for the developing field of molecular biology. An important part of this field is directed at uncovering aspects of organisms' development (see Developmental Biology).

Theory in developmental biology has often diverged from theory in evolutionary biology. Developmental biologists have periodically challenged views and approaches in evolutionary biology, including evolutionary biologists' focus on the gene. With the new techniques in molecular biology came the hope for a unified approach to evolution and development. In this approach, molecular evolutionary biology and molecular developmental biology would work consistently side by side (see Molecular Biology). The processes of development and evolution could be understood from a unified molecular perspective if the component of heredity in evolution were understood to be the passing on of DNA from one generation to the next and the component in development to be the production of proteins from DNA. In this picture, genes were discrete strands of DNA and each was responsible for the production of a particular polypeptide.

The linear structure of DNA and RNA reveals a role that a concept of information can play in understanding heredity and development. The bases in DNA and RNA can be helpfully construed as letters in an alphabet, and the relation between the triplets of letters in the DNA and the resulting polypeptide chain can be construed as a coding relation. So, the DNA contains the code for the polypeptide. Rather than *causing* the production of the relevant protein, the DNA sequence contains the *code* for it.

So, rather than genes being discrete strands of DNA passed on from one generation to the next, they can now be characterized as containing information that is transmitted across generations, and this information is the code for a particular polypeptide. What is relevantly transmitted across generations is the *information* in the DNA, encoded in the unique sequence of bases. Development can now be conceptualized as the faithful transmission of information from DNA to RNA, via complementary base pairs, and then the passing on of that information into the linear structure of the protein, via the coding relation between triplets of base pairs and specific amino acids. Molecular biologists have introduced terminology that is consistent with this approach: The information in DNA is "replicated" in cell division, "transcribed" from DNA to RNA, and "translated" from RNA into proteins.

Although the process of development includes every part of the life cycle of any particular organism, leading to the whole collection of the organism's phenotypic traits, the discussion that follows focuses on the part of the developmental process operating within cells that starts with the separation of DNA strands and concludes with the production of a protein. In some discussions, genetic information is presented as containing instructions for the production for phenotypic traits such as eyes, but these extensions of the concept present many additional problems to those reviewed below (Godfrey-Smith 2000).

The Pervasive Informational Gene Concept: History and Current Practice

In his provocative *What Is Life?* of 1944, the physicist Erwin Shrödinger said "these chromosomes . . . contain in some kind of code-script the entire pattern of the individual's future development and of its functioning in the mature state" (Shrödinger 1944, 20). He went on to explain his terminology:

> In calling the structure of the chromosome fibers a code-script we mean that the all-penetrating mind, once conceived by Laplace, to which every causal connection lay immediately open, could tell from their structure whether the egg would develop, under suitable conditions, into a black cock or into a speckled hen, into a fly or a maize plant, a rhododendron, a beetle, a mouse or a woman. (20–21)

As Morange (1998) put it, Shrödinger saw "genes merely as containers of information, as a code that determines the formation of the individual" (75). Shrödinger's proposals were made before the discovery of the structure of DNA. What is important is that his words were read by many of those who were instrumental in the development of molecular biology.

As Sarkar (1996) points out, Watson and Crick were the first to use the term "information" in the context of discussions of the genetic code:

> The phosphate-sugar backbone of our model is completely regular, but any sequence of the pairs of bases can fit into the structure. It follows that in a long molecule many different permutations are possible, and it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information. (Watson and Crick 1953, 964)

Subsequently, Jacob and Monod also played roles in sustaining Shrödinger's language of the code, helping to reinforce the use of information language in the new field of molecular biology (Keller 2000). By the early 1960s this terminology was established there.

The informational gene concept also became pervasive in the work of theoretical evolutionary biologists. Perhaps the most influential formulation of the concept of heredity in terms of information was that of the evolutionary theorist George Williams. In his influential *Adaptation and Natural Selection,* Williams claims:

> In evolutionary theory, a gene could be defined as any hereditary information for which there is a favorable or unfavorable selection bias equal to several or many times the rate of endogenous change. (Williams 1966, 25)

And, later:

> A gene is not a DNA molecule; it is the transcribable information coded by the molecule. (Williams 1992, 11)

It should now be clear that information terminology is pervasive in disciplines of biology, and also at least somewhat clear why this is the case. There were some historical reasons for adopting the terminology, and there is some utility to the informational concepts. There are, however, some problems associated with construing genes informationally. Many of these problems have been introduced by philosophers of biology, but there has also been much discussion of the informational gene concept within biology.

Problems of the Informational Gene Concept

In several of his recent writings, the evolutionary biologist John Maynard Smith has invited philosophers to join the discussion about the informational gene concept. For example, he says that "given the role that ideas drawn from a study of human communication have played, and continue to play, in biology it is strange that so little attention has been paid to them by philosophers of biology. I think that it is a topic that would reward serious study" (Maynard Smith 2000, 192). While not addressing the concept of genetic information directly, philosophers of biology have been attending to these issues indirectly for some time in working on central problems in the philosophy of biology. For example, the notion of genes as information has played an important role in discussions of reductionism, units of selection, the replicator/interactor distinction, gene/environment interactions, and nativism (see Innate/Acquired Distinction, Population Genetics, Reductionism). Recently, philosophers' focus has turned more explicitly to the informational gene concept. Several philosophers are now engaged in the project of developing a general notion of information that fits best with biologists' aims when they invoke genetic information.

The informational definition of the gene introduced above says that genes contain information that is passed on from one generation to the next, information that codes for particular proteins and polypeptides. As Sterelny and Griffiths (1999) put it: "The classical molecular gene concept is a stretch of DNA that codes for a single polypeptide chain" (132). Genes, in this view, contain information about the phenotype, the protein that is expressed. While most biologists believe that genes contain information about the relevant phenotype, probably no one believes that the information in the genes is sufficient to produce the relevant phenotypes. Even those most routinely chastised for being genetic determinists understand that the information in the gene is expressed only with the aid of a whole host of cellular machinery. As a result, the standard view is that genes contain the relevant or important information guiding the development of the organism. All other cellular machinery merely assists in the expression of the information. One way to put this idea is that genes introduce information to the developmental process, while all other mechanisms make merely a causal contribution to development.

One move that philosophers (and some biologists) have made is to characterize the process of passing on the information in the gene by using terms from information theory. Information theory holds that

> an event carries information about another event to the extent that it is causally related to it in a systematic fashion. Information is thus said to be conveyed over a "channel" connecting the "sender" [or "signal"] with the "receiver" when a change in the receiver is causally related to a change in the sender. (Gray 2001, 190)

In this view information is reduced to causal covariance or systematic causal dependence. Philosophers of biology refer to this characterization of genetic information as the "causal" view. Sterelny and Griffiths (1999) illustrate how the causal information concept could work in the context of molecular biology:

> The idea of information as systematic causal dependence can be used to explain how genes convey developmental information. The genome is the signal and the rest of the developmental matrix provides channel conditions under which the life cycle of the organism contains (receives) information about the genome. (102)

It has been argued that the causal view suffers from serious problems. Sterelny and Griffiths (1999) point out that "it is a fundamental fact of information theory that the role of signal source and channel condition can be reversed" (102) as the signal/channel distinction is simply a matter of causal covariance. Further, the signal/channel distinction is a function of observers' interests. For example, one could choose to hold the developmental history of an organism constant, and from this perspective the organism's phenotype would carry information about its genotype. But if it is instead chosen to "hold all developmental factors other than (say) nutrient quantity constant, the amount of nutrition available to the organism will covary with, and hence also carry information about its phenotype" (102). The causal information concept is lacking, because it cannot distinguish the genes as the singular bearers of important or relevant information. Rather, in this view, genes are just one source of information; aspects of the organism's environment and cellular material also contain information. This position is called the parity thesis (Griffiths and Gray 1994). The parity thesis exposes the need for another information concept that elevates genes alone to the status of information bearers.

Alternative concepts of information have been examined in attempts to respond to this situation; one is referred to variously as intentional, semantic, or teleosemantic information. This notion of information has been defended most forcefully recently by Maynard Smith, and also by philosophers Daniel Dennett (1995) and Kim Sterelny (2000). The term *teleosemantics* is borrowed from "the philosophical program of reducing meaning to biological function (teleology) and then reducing biological function to" natural selection. (A good survey of relations between the philosophy of mind and genetic information concepts is provided in Godfrey-Smith 1999.) This view is articulated in the philosophy of mind as the thesis that a mental state token, such as a sentence, has the biological function of representing a particular state of the world and that this function arose as a result of selection.

Applying this view to the current problem results in the following: "A gene contains information about the developmental outcomes that it was selected to produce" (Sterelny and Griffiths 1999, 105). Maynard Smith puts the view as: "DNA contains information that has been programmed by natural selection" (Maynard Smith 2000, 190). Here the information in the gene is analogous to a sentence in the head. The gene contains information as a result not just of relevantly causal covariance with the phenotype, but of having the function of producing the relevant phenotype. Defenders of this view claim that this function allows for the information to stay the same even if the channel conditions change, in which case the information in the gene has simply been misinterpreted. This concept could solve the problem of rendering the genes as the sole information bearers, as "if other developmental causes do not contain [teleosemantic] information and genes do, then genes do indeed play a unique role in development" (Sterelny and Griffiths 1999, 104).

Although the teleosemantic view shows promise, the debate has not ended here. The teleosemantic view opens up a possibility: If a developmental cause—part of the cellular machinery, for example—is found to be heritable and performs the function of producing a particular developmental outcome, then, by definition, it also contains teleosemantic information. Many, including Sarkar (1996, 2000), Griffiths and Gray (1994), Gray (2001), Keller (2000), Sterelny (2000), have argued that indeed there are such mechanisms. These authors draw various conclusions from the demonstrated presence of mechanisms that are not genes, are heritable, and perform the function of producing a specific developmental outcome. Developmental systems theorists such as Griffiths and Gray take these findings to show that teleosemantic information succumbs to the parity thesis also. They go on to argue that no concept of information will distinguish genes as a special contributor to development. Genes are just fellow travelers alongside cellular machinery and the environment in shaping developmental outcomes. Others such as Sarkar and Keller are more cautious and hold out for a concept of information that can distinguish genes as a distinct kind of information bearer. On the other side, Maynard Smith and others have attempted to refine the notion of teleosemantic information to preserve a biological distinction that seems to be important: "The most fundamental distinction in biology is between nucleic acids, with their role as carriers of information, and proteins, which generate the phenotype" (Maynard Smith and Szathmary 1995, 61).

Three coherent options present themselves to answer the question, Where is biological information found?

1. Information is present in DNA and other nucleotide sequences. Other cellular mechanisms contain no information.
2. Information is present in DNA, other nucleotide sequences, and other cellular mechanisms (for example, cytoplasmic or extracellular proteins), and in many other media—for example, the embryonic environment or components of an organism's wider environment.
3. DNA and other nucleotide sequences do not contain information, nor do any other cellular mechanisms.

These options can be read either ontologically or heuristically. The ontological reading of (1) is that there is a certain kind of information that is present only in DNA and other nucleotide sequences. As a result, any workable concept of information is constrained. The concept adopted cannot be consistent with information of the relevant sort existing in any other media that are causally responsible for an organism's development. The heuristic reading of (1) is that viewing information as present in DNA and other nucleotides is the most reliable guide to good answers in research in developmental molecular biology. The philosophical discussion presented above focuses on developing or challenging accounts of information that are consistent with an ontological reading of (1). For example, Maynard Smith and others, such as Dennett, are defenders of an ontological version of (1).

Many assume that (2) makes sense ontologically only if one adopts a causal information concept, but some of the discussion already referred to indicates that other developmentally relevant media can be construed as containing teleosemantic information. Defenders of the developmental systems theory approach hold a version of (2), as does Sarkar (1996).

Only Waters (2000) seems to have provided a sustained defense of (3). Maynard Smith argues that to construe all processes of development in causal terms without recourse to the concept of genetic information is to relegate them to the hopelessly complex and to implicitly argue that no systematic explanations will be forthcoming (see e.g., Maynard Smith 1998, 5–6). Waters differs, arguing that informational talk in biology is misleading and can entirely be coherently substituted for by causal talk. Waters also argues that it is the intent of most practicing biologists to provide a causal account of development rather than one that invokes information.

In conclusion, philosophers are actively cooperating with theoretical biologists to develop fruitful concepts of information that help make sense of the information terminology widely used in biology. These discussions are as yet inconclusive, and as a result this is a potentially fertile area for future work.

STEPHEN M. DOWNES

The author acknowledges the helpful input of Sahotra Sarkar, University of Texas and Lindley Darden, University of Maryland.

References

Dennett, D. C. (1995), *Darwin's Dangerous Idea*. New York: Simon and Schuster.
Godfrey-Smith, P. (1999), "Genes and Codes: Lessons from the Philosophy of Mind?" in V. G. Hardcastle (ed.), *Where Biology Meets Psychology: Philosophical Essays*. Cambridge, MA: MIT Press: 305–332.
——— (2000), "On the Theoretical Role of 'Genetic Coding,'" *Philosophy of Science* 67: 26–44.
Gray, R. D. (2001), "Selfish Genes or Developmental Systems?" in R. S. Singh, C. B. Krimbas D. B. Paul, and J. Beatty (eds.), *Thinking About Evolution: Historical, Philosophical, and Political Perspectives*. Cambridge: Cambridge University Press, 184–207.
Griffiths, P. E., and R. D. Gray (1994), "Developmental Systems and Evolutionary Explanation," *Journal of Philosophy* 91: 277–304.
Keller, E. F. (2000), "Decoding the Genetic Program: Or, Some Circular Logic in the Logic of Circularity," in P. J. Beurton, R. Falk, and H. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution*. Cambridge: Cambridge University Press, 159–177.
Maynard Smith, J. (1998), *Shaping Life*. New Haven, CT: Yale University Press.
——— (2000), "The Concept of Information in Biology," *Philosophy of Science* 67: 177–194.
Maynard Smith, J., and E. Szathmary (1995), *The Major Transitions in Evolution*. Oxford: Oxford University Press.
Morange, M. (1998), *A History of Molecular Biology*. Cambridge, MA: Harvard University Press.
Sarkar, S. (1996), "Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology," in S. Sarkar (ed.), *The Philosophy and History of Molecular Biology: New Perspectives*. Dordrecht, Netherlands: Kluwer, 187–231.
——— (2000). "Information in Genetics and Developmental Biology," *Philosophy of Science* 67: 208–213.
Schrödinger, E. (1944), *What Is Life? The Physical Aspects of the Living Cell*. Cambridge: Cambridge University Press.
Sterelny, K. (2000), "The 'Genetic Program' Program: A Commentary on Maynard Smith on Information in Biology," *Philosophy of Science* 67: 195–201.
Sterelny, K., and P. E. Griffiths (1999), *Sex and Death*. Chicago: University of Chicago Press.
Waters, K. (2000), "Molecules Made Biological," *Revue Internationale de Philosophie* 4: 539–564.
Watson, J. D., and F. H. C. Crick (1953), "Genetical Implications of the Structure of Deoxyribonucleic Acid," *Nature* 171: 964–967.
Williams, G. C. (1966), *Adaptation and Natural Selection*. Princeton, NJ: Princeton University Press.
——— (1992), *Natural Selection: Domains, Levels and Challenges*. New York: Oxford University Press.

See also **Evolution; Genetics; Molecular Biology; Population Genetics**

PHILOSOPHY OF BIOLOGY

The philosophy of biology has existed as a distinct subdiscipline within the philosophy of science for about 30 years. The rapid growth of the field has mirrored that of the biological sciences in the same period. Today the discipline is well represented in the leading journals in philosophy of science, as well as in several specialist journals. There have been two generations of textbooks (see Conclusion), and the subject is regularly taught at the undergraduate as well as the graduate level. The current high profile of the biological sciences and the obvious philosophical issues that arise in fields as diverse as molecular genetics and conservation biology suggest that the philosophy of biology will remain an exciting field of enquiry for the foreseeable future.

Three Kinds of Philosophy of Biology

Philosophers have engaged with biological science in three quite distinct ways. Some have looked to biology to test general theses in philosophy of science (see Laws of Nature). Others have engaged with conceptual puzzles that arise within biology itself. Finally, philosophers have looked to biological science for answers to distinctively philosophical questions in such fields as ethics, the philosophy of mind, and epistemology (see Evolutionary Epistemology).

The debate that marked the beginning of contemporary philosophy of biology exemplified the first of these three approaches, the use of biological science as a testing ground for claims in general philosophy of science. In the late 1960s, Kenneth C. Schaffner applied the logical empiricist model of theory reduction to the relationship between classical, Mendelian genetics and the new molecular genetics (Schaffner 1967, 1969; Hull 1974). While the failure of this attempt in its initial form reinforced the near-consensus in the 1970s and 1980s that the special sciences are autonomous from the more fundamental sciences, it also led the formulation of increasingly more adequate models of theory reduction (Schaffner 1993; Sarkar 1998) (see Reductionism).

Another important early debate showed philosophy engaging biology in the second way, by confronting a conceptual puzzle within biology itself. The concept of reproductive fitness is at the heart of evolutionary theory, but its status has always been problematic (see Fitness). It has proved surprisingly hard for biologists to avoid the criticism that natural selection explains the reproductive success of organisms by citing their fitness, while defining their fitness in terms of their reproductive success (the so-called tautology problem). Philosophical analysis of this problem begins by noting that fitness is a supervenient property of organisms: The fitness of each particular organism is a consequence of some specific set of physical characteristics of the organism and its particular environment, but two organisms may have identical levels of fitness in virtue of very different sets of physical characteristics (Rosenberg 1978) (see Supervenience). The most common solution to the tautology problem is to argue that this supervenient property is a propensity—a probability distribution over possible numbers of offspring (Mills and Beatty 1979). Thus, although fitness is defined in terms of reproductive success, it is not a tautology that the fittest organisms have the most offspring. Fitness merely allows us to make fallible predictions about numbers of offspring that become more reliable as the size of the population tends to infinity. It remains unclear, however, whether it is possible to specify a probability distribution or set of distributions that can play all the roles actually played by fitnesses in population biology (Rosenberg and Bouchard 2002).

The third way in which philosophy has engaged with biology is by tracing out the wider ethical, epistemological, and metaphysical implications of biological findings. This has sometimes occurred in response to philosophical claims issuing from within biology itself. For example, some proponents of sociobiology—the application to humans of the models developed in behavioral ecology in the 1960s—suggested that the conventional social sciences could be reduced to or replaced by behavioral biology (see Evolutionary Psychology). Others claimed that certain aspects of human behavior result from strongly entrenched aspects of human biology and thus that public policy must be designed to work with and around such behavior rather than seeking to eradicate it. These claims were evaluated by leading philosophers of biology like Michael Ruse (1979), Alexander Rosenberg (1980), and Philip Kitcher (1985).

On other occasions, rather than responding to philosophical claims issuing from within biology, philosophers have actively sought from biology answers to questions arising in their own discipline that may not be of particular interest to working biologists. The extensive literature on biological teleology is a case in point (see Function). After a brief flurry of interest around the time of the modern synthesis (see Evolution), during which the term "teleonomy" was introduced to denote the specifically evolutionary interpretation of teleological language (Pittendrigh 1958), the ideas of function and goal directedness were regarded as relatively unproblematic by evolutionary biologists, and there was little felt need for any further theoretical elaboration of these notions. In the 1970s, however, philosophers started to look to biology to provide a solid, scientific basis for normative concepts, such as illness or malfunction (Wimsatt 1972; Wright 1973). These discussions eventually converged on an analysis of teleological language fundamentally similar to the view associated with the modern synthesis, although elaborated in far greater detail. According to the *etiological theory of function,* the functions of a trait are those activities in virtue of which the trait was selected (Brandon 1981; Millikan 1984; Neander 1991, 1995). Despite continued disputes over the scope and power of the etiological theory amongst philosophers of biology (Ariew, Cummins, and Perlman 2002), the idea of "etiological" or "proper" function has become part of the conceptual toolkit of philosophy in general and of the philosophy of language and of mind in particular.

These three approaches to doing philosophy of biology are exemplified in different combinations in philosophical discussion of the several biological disciplines.

The Philosophy of Evolutionary Biology

Evolutionary theory has been used as a case study in support of views of the structure of scientific theories in general, an approach that conforms to the "testing ground" conception of philosophy of biology described above (see Evolution). The example is most often thought to favor the "semantic view" of theories (Lloyd 1988) (see Theories). Most philosophical writing about evolutionary theory, however, is concerned with conceptual puzzles that arise inside the theory itself, and the work often resembles theoretical biology as much as pure philosophy of science. Elliott Sober's classic study *The Nature of Selection: Evolutionary Theory in Philosophical Focus* (Sober 1984) marks the point at which most nonspecialists became aware of the philosophy of biology as a major new field. In this work Sober analyzed the structure of selective explanations via an analogy with the composition of forces in dynamics, treating the actual change in gene frequencies over time as the result of several different "forces," such as selection, drift, and mutation (see Natural Selection). Sober's book also introduced the widely used distinction between "selection for" and "selection of." Traits that are causally connected to reproductive success and can therefore be used to *explain* reproductive success are said to be selected *for* (or to be "targets" of selection). In contrast, there is selection *of* traits that do not have this property but nevertheless are statistically associated with reproductive success, usually because they are linked in some way to traits that *do* have the property. For example, when two DNA segments are "linked" in the classical sense of being close to one another on the same chromosome, they have a high probability of being inherited together (see Genetics). If only one of the two segments has any effect on the phenotype, it is the presence of this segment alone that explains the success of both. There is selection *for* the causally active segment but only selection *of* its passive companion.

Robert Brandon's (1990) analysis of the concept of the environment is, similarly, of as much interest to biologists as to philosophers. Several biological authors have criticized the idea that the "environment," in the sense in which organisms are adapted to it, can be described independently of the organisms themselves. Brandon defines three different notions of 'environment,' all of which are needed to make sense of the role of environment in natural selection. All organisms in a particular region of space and time share the "external environment," but to understand the particular selective forces acting on one lineage of organisms it is necessary to pick out a specific "ecological environment" consisting of those environmental parameters whose value affects the reproductive output of members of the lineage. The ecological environment of a fly will be quite different from that of a tree, even if they occupy the same external environment. Finally, the "selective environment" is that part of the ecological environment that *differentially* affects the reproductive output of variant forms in the evolving lineage. It is this last that contains the sources of adaptive evolutionary pressures on the lineage.

Part of the early philosophical interest in selective explanation arose due to philosophical interest in sociobiology. Sociobiology was widely criticized for its "adaptationism," or an exclusive focus on selection to the exclusion of other evolutionary factors (see Adaptation and Adaptationism). This gave rise to several important papers on the concept of "optimality" in evolutionary modeling (Dupré 1987). Philosophers have now distinguished several distinct strands of the adaptationism debate, and many of the remaining issues are clearly empirical rather than conceptual, as is made clear in the latest collection of papers on this issue (Orzack and Sober 2001).

The sociobiology debate, and related discussion of the idea that the fundamental unit of evolution is the individual Mendelian allele (Dawkins 1976), also drove the explosion of philosophical work on the "units of selection" question in the 1980s (Brandon and Burian 1984). Philosophical work on the units-of-selection question has tended to favor some form of pluralism, according to which there may be units of selection at several levels within the hierarchy of biological organization—DNA segments, chromosomes, cells, organisms, and groups of organisms. Arguably, philosophers made a significant contribution to the rehabilitation of some forms of "group selection" in evolutionary biology itself, following two decades of neglect (Sober and Wilson 1998).

More recently, a heated debate has developed over the ontological status of the probabilities used in population biology (see Evolution). On the one hand, the best models of the evolutionary process assign organisms a certain probability of reproducing (fitness) and make probabilistic predictions about the evolutionary trajectory of populations. On the other hand, the actual process of evolution is the aggregation of the lives of many individual organisms, and those organisms lived, died, and reproduced in accordance with deterministic, macro-level physical laws. Hence, it has been argued, the evolutionary process itself is deterministic, a vast soap opera in which each member of the cast has an eventful history determined by particular causes; and the probabilities in evolutionary models are introduced because one cannot follow the process in all its detail (Rosenberg 1994; Walsh 2000). If correct, this argument has some interesting implications. It would seem to follow, for example, that there is no real distinction in nature between the process of drift and the process of natural selection. Robert Brandon and Scott Carson (1996) have strongly rejected this view, insisting that evolution is a genuinely indeterministic process and that the probabilistic properties ascribed to organisms by evolutionary models should be accepted in the same light as the ineliminable explanatory posits of other highly successful theories.

The Philosophy of Systematic Biology

Philosophical discussion of systematics was a response to a "scientific revolution" in that discipline in the 1960s and 1970s, which saw the discipline first transformed by the application of quantitative methods and then increasingly dominated by the "cladistic" approach, which rejects the view that systematics should sort organisms into a hierarchy of groups representing a roughly similar amount of diversity, and argues that its sole aim should be to represent evolutionary relationships between groups of organisms (phylogeny). Ideas from the philosophy of science were used to argue for both transformations, and the philosopher David L. Hull (1988) was an active participant throughout this whole period. Another major treatment of cladism was by Sober (1988).

The best-known topic in the philosophy of systematics was introduced by the biologist Michael Ghiselin (1974), when he suggested that traditional systematics was fundamentally mistaken about the ontological status of biological species (see also Hull 1976). Species, it was argued, are not natural kinds of organisms in the way that chemical elements are natural kinds of matter. Instead, they are historical particulars like families or nations (see Individuality). However, the view that species are historical particulars leaves other important questions about species unsolved and raises new problems of its own. As many as 20 different so-called species concepts are represented in the current biological literature, and their merits, interrelations, and mutual consistency or inconsistency have been a major topic of philosophical discussion (the papers collected in Ereshefsky [1992] provide a good introduction to these debates) (see also Species).

The philosophy of systematics has influenced general philosophy of science and, indeed, metaphysics, through its challenge to one of the two classical examples of a "natural kind," *viz.,* biological species. The result has been a substantial reevaluation of what is meant by a natural kind, whether there are natural kinds, and whether traditional views about the nature of science that rely on the idea of natural kinds must be rejected (Wilkerson 1993; Dupré 1993; Wilson 1999).

The Philosophy of Molecular Biology

As mentioned above, one of the first topics to be discussed in the philosophy of biology was the reduction of Mendelian to molecular genetics. The initial debate between Schaffner and Hull was followed by the "anti-reductionist consensus," embodied in Philip Kitcher's (1984) classic paper *1953 and All That: A Tale of Two Sciences.* The reductionist position was revived in a series of important papers by Kenneth Waters (1990, 1994) and debate over the cognitive relationship between these two theories continues today, although the question is not now framed as a simple choice between reduction and irreducibility. For example, William Wimsatt has tried to understand 'reduction' not as a judgment on the fate of a theory, but as one amongst several strategies that scientists can deploy when trying to unravel complex systems (see Reductionism). The philosophical interest lies in understanding the strengths and weaknesses of this strategy (Wimsatt 1976, 1980). Lindley Darden, Schaffner, and others have argued that explanations in molecular biology are not neatly confined to one ontological level, and hence that ideas of "reduction" derived from classical examples like the reduction of the phenomenological gas laws to molecular kinematics in nineteenth-century physics are simply inapplicable (Darden and Maull 1977; Schaffner 1993). Moreover, molecular biology does not have the kind of grand theory based around a set of laws or a set of mathematical models that is familiar from the physical sciences. Instead, highly specific mechanisms that have been uncovered in detail in one model organism seem to act as "exemplars" allowing the investigation of similar, although not necessarily identical, mechanisms in other organisms that employ the same, or related, molecular interactants. Darden and collaborators have argued that these "mechanisms"—specific collections of entities and their distinctive activities—are the fundamental unit of scientific discovery and scientific explanation, not only in molecular biology, but in a wide range of special sciences (Machamer, Darden, and Craver 2000) (see Mechanisms).

An important strand in the early debate over reduction concerned the different ways in which the gene itself is understood in Mendelian and molecular genetics. The gene of classical Mendelian genetics has been replaced by a variety of structural and functional units in contemporary molecular genetics (see Genetics). One response to this is pluralism about the gene (Falk 2000). Another is to identify a central tendency that unifies the various different ways in which the term 'gene' is used (Waters 1994, 2000). Identifying the different ways in which genes are conceived in different areas of molecular biology and their relations to one another is a major focus of current research (Beurton, Falk, and Rhineberger 2000; Moss 2002; Stotz, Griffiths, and Knight 2004). Another very active topic is the concept of genetic information, or developmental information more generally (Sarkar 1996a, 2004; Maynard Smith 2000; Griffiths 2001; Jablonka 2002) (see Biological Information; Molecular Biology).

The Philosophy of Developmental Biology

Developmental biology has received growing attention from philosophers in recent years. The debate over "adaptationism" introduced philosophers to the idea that explanations of traits in terms of natural selection have time and time again in the history of Darwinism found themselves in competition with explanations of the same traits from developmental biology.

Developmental biology throws light on the kinds of variation that are likely to be available for selection, posing the question of how far the results of evolution can be understood in terms of the options that were available ("developmental constraints") rather that the natural selection of those options (Maynard Smith et al. 1985). The question of when these explanations compete and when they complement one another is of obvious philosophical interest. The debate over developmental constraints looked solely at whether developmental biology could provide answers to evolutionary questions. However, as Ron Amundson pointed out, developmental biologists are addressing questions of their own, and, he argued, a different concept of 'constraint' is needed to address those questions (Amundson 1994). In the last decade several other debates in the philosophy of biology have taken on a novel aspect by being viewed from the standpoint of developmental biology. These include the analysis of biological teleology (Amundson and Lauder 1994), the units-of-selection debate (Griffiths and Gray 1994), and the nature of biological classification, which from the perspective of development is as much a debate about classifying the parts of organisms as about classifying the organisms themselves (Wagner 2001). The vibrant new field of evolutionary developmental biology is transforming many evolutionary questions within biology itself and hence causing philosophers to revisit existing positions in the philosophy of evolutionary biology (Brandon and Sansom 2005).

Increasing philosophical attention to developmental biology has also led philosophers of biology to become involved in debates over the concept of innateness, the long tradition of philosophical literature on this topic having previously treated innateness primarily as a psychological concept (Ariew 1996; Griffiths 2002).

The Philosophy of Ecology and Conservation Biology

Until recently this was a severely underdeveloped field in the philosophy of biology, which was surprising, because there is obvious potential for all three of the approaches to philosophy of biology discussed above. First, ecology is a demanding testing ground for more general ideas about science, for reasons explained below (see Ecology). Second, there is a substantial quantity of philosophical work in environmental ethics, and it seems reasonable to suppose that answering the questions that arise there would require a critical methodological examination of ecology and conservation biology. Finally, ecology contains a number of deep conceptual puzzles, which ecologists themselves have recognized and discussed extensively.

The most substantial contributions to the field to date include works by Kristin Shrader-Frechette and Earl McCoy (1993), Gregory Cooper (2003), and Lev Ginzburg and Mark Colyvan (2004). Cooper focuses on the particular methodological problems that confront ecology as a result of its subject matter—massively complex, and often unique, systems operating on scales that frequently make controlled experiment impractical—and on the consequent lack of connection between the sophisticated mathematical modeling tradition in ecology and ecological field work. Shrader-Frechette and McCoy's book is concerned primarily with how practical conservation activity can be informed by ecological theory despite the problems addressed by Cooper (for a related discussion, see Sarkar 1996b). Ginzburg and Colyvan, in contrast, argue forcefully that ecology may still produce simple, general theories that will account for the data generated by ecological field work in as satisfactory a manner as Newtonian dynamics accounted for the motion of the planets.

The concept of the niche stands in marked contrast to other ecological concepts in that 'niche' has been widely discussed by philosophers of biology (summarized in Sterelny and Griffiths 1999, 268–279). This, however, reflects the importance of the niche concept in *evolutionary* biology. Topics that merit much more attention than the little they have received to date include the concept of biodiversity and of stability (or, in its popular guise, the "balance of nature"). A recent extended philosophical discussion of these concepts, integrating themes from the philosophy of ecology and conservation biology with more traditional environmental philosophy, is found in Sarkar (2005) (see also Conservation Biology).

Conclusion

The philosophy of biology is a flourishing field, partly because it encompasses all three of the very different ways in which philosophy makes intellectual contact with the biological sciences, as discussed above. The scope of philosophical discussion has extended from its starting points in evolutionary biology to encompass systematics, molecular biology, developmental biology, and, increasingly, ecology and conservation biology. For those who wish to explore the field beyond this article and the related articles in this volume, recent textbooks include Sober (1993) and Sterelny and Griffiths (1999). Two valuable edited collections designed to supplement such a text are Sober (1994), which collects the classic papers on core debates, and Hull and Ruse (1998), which aims at a comprehensive survey using recent papers. Keller and Lloyd (1992) have edited an excellent collection on evolutionary biology aimed primarily at philosophers of biology.

PAUL E. GRIFFITHS

References

Amundson, Ron (1994), "Two Concepts of Constraint: Adaptationism and the Challenge from Developmental Biology," *Philosophy of Science* 61: 556–578.

Amundson, Ron, and George V. Lauder (1994), "Function Without Purpose: The Uses of Causal Role Function in Evolutionary Biology," *Biology and Philosophy* 9: 443–470.

Ariew, Andre (1996), "Innateness and Canalization," *Philosophy of Science* 63(suppl): S19–S27.

Ariew, Andre, Robert Cummins, and Mark Perlman (eds.) (2002), *Functions: New Essays in the Philosophy of Psychology and Biology*. New York and Oxford: Oxford University Press.

Beurton, Peter, Raphael Falk, and Hans-Joerg Rhineberger (2000), *The Concept of the Gene in Development and Evolution*. Cambridge: Cambridge University Press.

Brandon, Robert N. (1981), "Biological Teleology: Questions and Explanations," *Studies in the History and Philosophy of Science* 12: 91–105.

——— (1990), *Adaptation and Environment*. Princeton, NJ: Princeton University Press.

Brandon, Robert N., and Richard M. Burian (eds.) (1984), *Genes, Organisms, Populations: Controversies Over the Units of Selection*. Cambridge, MA: MIT Press.

Brandon, Robert N., and Scott Carson (1996), The Indeterministic Character of Evolutionary Theory: No "Hidden Variables Proof" but No Room for Determinism Either," *Philosophy of Science* 63: 315–337.

Brandon, Robert N., and Roger Sansom (eds.) (2005), *Integrating Evolution and Development*. Cambridge: Cambridge University Press.

Cooper, Gregory (2003), *The Science of the Struggle for Existence: On the Foundations of Ecology*. Edited by M. Ruse. Cambridge Studies in Biology and Philosophy. Cambridge: Cambridge University Press.

Darden, Lindley, and Nancy Maull (1977), "Interfield Theories," *Philosophy of Science* 44: 43–64.

Dawkins, Richard (1976), *The Selfish Gene*. Oxford: Oxford University Press.

Dupré, John (1993), *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

——— (ed.) (1987), *The Latest on the Best: Essays on Optimality and Evolution*. Cambridge, MA: MIT Press.

Ereshefsky, Marc (ed.) (1992), *The Units of Evolution*. Cambridge, MA: MIT Press.

Falk, Raphael (2000), "The Gene: A Concept in Tension," in P. Beurton, R. Falk, and H.-J. Rheinberger (eds.), *The Concept of the Gene in Development and Evolution*. Cambridge: Cambridge University Press.

Ghiselin, Michael T. (1974), "A Radical Solution to the Species Problem," *Systematic Zoology* 23: 536–544.

Ginzburg, Lev, and Mark Colyvan (2004), *Ecological Orbits: How Planets Move and Populations Grow*. Oxford and New York: Oxford University Press.

Griffiths, Paul E. (2001), "Genetic Information: A Metaphor in Search of a Theory," *Philosophy of Science* 68: 394–412.

——— (2002), "What Is Innateness?" *The Monist* 85: 70–85.

Griffiths, P. E, and R. D Gray (1994), "Developmental Systems and Evolutionary Explanation," *Journal of Philosophy* XCI: 277–304.

Hull, David L. (1974), *Philosophy of Biological Science*. Englewood Cliffs, NJ: Prentice-Hall Inc.

——— (1976), "Are Species Really Individuals?" *Systematic Zoology* 25: 174–191.

——— (1988), *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.

Hull, David L., and Michael Ruse (eds.) (1998), *The Philosophy of Biology*. Oxford: Oxford University Press.

Jablonka, Eva (2002), "Information Interpretation, Inheritance, and Sharing," *Philosophy of Science* 69: 578–605.

Keller, Evelyn Fox, and Elisabeth A. Lloyd (eds.) (1992), *Keywords in Evolutionary Biology*. Cambridge, MA: Harvard University Press.

Kitcher, Philip (1984), "1953 and All That: A Tale of Two Sciences," *Philosophical Review* 93: 335–373.

——— (1985), *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Cambridge, MA: MIT Press.

Lloyd, Elizabeth A. (1988), *The Structure and Confirmation of Evolutionary Theory*. Westport, CT: Greenwood Press.

Machamer, Peter, Lindley Darden, and Carl Craver (2000), "Thinking about Mechanisms," *Philosophy of Science* 67: 1–25.

Maynard Smith, John (2000), "The Concept of Information in Biology," *Philosophy of Science* 67: 177–194.

Maynard Smith, J., Richard M. Burian, Stuart Kauffman, Pere Alberch, J. Campbell, B. Goodwin, et al. (1985), "Developmental Constraints and Evolution," *Quarterly Review of Biology* 60: 265–287.

Millikan, Ruth G. (1984), *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

Mills, Susan, and John Beatty (1979), "The Propensity Interpretation of Fitness," *Philosophy of Science* 46: 263–286.

Moss, L. (2002), *What Genes Can't Do*. Cambridge, MA: MIT Press.

Neander, Karen (1991), "Functions as Selected Effects: The Conceptual Analyst's Defense," Philosophy of Science 58: 168–184.

——— (1995), "Misrepresenting and Malfunctioning," *Philosophical Studies* 79: 109–141.

Orzack, Steve, and Elliott Sober (eds.) (2001), *Optimality and Adaptation*. Cambridge: Cambridge University Press.

Pittendrigh, C. S. (1958), "Adaptation, Natural Selection and Behavior," in A. Roe and G. Simpson (eds.), *Behavior and Evolution*. New York: Academic Press.

Rosenberg, Alexander (1978), "The Supervenience of Biological Concepts," *Philosophy of Science* 45: 368–386.

——— (1980), *Sociobiology and the Preemption of Social Science*. Baltimore: Johns Hopkins University Press.

——— (1994), *Instrumental Biology or The Disunity of Science*. Chicago: Chicago University Press.

Rosenberg, Alexander, and Frederic Bouchard (2002), "Fitness," *Stanford Encyclopedia of Philosophy*. Available at http://plato.stanford.edu/archives/win2002/entries/fitness. Accessed August 2005.

Ruse, Michael (1979), *Sociobiology: Sense or Nonsense*. Dordrecht, Holland: Reidel.

Sarkar, Sahotra (1996a), "Biological Information: A Sceptical Look at Some Central Dogmas of Molecular Biology," in *The Philosophy and History of Molecular Biology: New Perspectives*. Dordrecht, Holland: Kluwer Academic Publishers.

——— (1996b), "Ecological Theory and Anuran Declines," *BioScience* 46: 199–207.

——— (1998), *Genetics and Reductionism*. Cambridge: Cambridge University Press.

——— (2004), "Molecular Models of Life: Philosophical Papers on Molecular Biology," Cambridge, MA: MIT Press.

——— (2005), *Biodiversity and Environmental Philosophy: An Introduction*. Cambridge: Cambridge University Press.

Schaffner, Kenneth F. (1967), "Approaches to Reduction," *Philosophy of Science* 34: 137–47.

——— (1969), "The Watson-Crick Model and Reductionism," *British Journal for the Philosophy of Science* 20: 325–348.

——— (1993), *Discovery and Explanation in Biology and Medicine*. Chicago and London: University of Chicago Press.

Shrader-Frechette, Kristin S., and Earl D. McCoy (1993), *Method in Ecology: Strategies for Conservation*. Cambridge and New York: Cambridge University Press.

Sober, Elliott (1984), *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, MA: MIT Press.

——— (1988), *Reconstructing the Past: Parsimony, Evolution and Inference*. Cambridge, MA: MIT Press.

——— (1993), *Philosophy of Biology*. Boulder, CO: Westview Press.

——— (ed.) (1994), *Conceptual Issues in Evolutionary Biology: An Anthology* (2nd ed.). Cambridge, MA: MIT Press.

Sober, Elliot, and David S. Wilson (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Sterelny, Kim, and Paul E. Griffiths (1999), *Sex and Death: An Introduction to the Philosophy of Biology*. Chicago: University of Chicago Press.

Stotz, Karola, Paul E. Griffiths, and Rob D. Knight (2004), "How Scientists Conceptualise Genes: An Empirical Study," *Studies in History and Philosophy of Biological and Biomedical Sciences* 35: 647–673.

Wagner, Günther P. (ed.) (2001), *The Character Concept in Evolutionary Biology*. San Diego: Academic Press.

Walsh, Dennis M. (2000), "Chasing Shadows: Natural Selection and Adaptation," *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 31C: 135–154.

Waters, C. Kenneth (1990), "Why the Antireductionist Consensus Won't Survive the Case of Classical Mendelian Genetics," in A. Fine, M. Forbes, and L. Wessells (eds.), *Proceedings of the Biennial Meeting of the Philosophy of Science Association*. East Lansing, MI: Philosophy of Science Association.

——— (1994), "Genes Made Molecular," *Philosophy of Science* 61: 163–185.

——— (2000), "Molecules Made Biological," *Revue Internationale de Philosophie* 4: 539–564.

Wilkerson, Timothy E. (1993), "Species, Essences and the Names of Natural Kinds," *Philosophical Quarterly* 43: 1–9.

Wilson, Robert A. (ed.) (1999), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press.

Wimsatt, William C. (1972), "Teleology and the Logical Structure of Function Statements," *Studies in the History and Philosophy of Science* 3: 1–80.

——— (1976), "Reductive Explanation: A Functional Account," in R. S. Cohen (ed.), *Proceedings of the Philosophy of Science Association, 1974*. East Lansing, MI: Philosophy of Science Association.

——— (1980), "Reductionistic Research Strategies and Their Biases in the Units of Selection Controversy," in

T. Nickles (ed.), *Scientific Discovery: Case Studies*. Dordrecht, Holland: D. Reidel Publishing Company.
Wright, Larry (1973), "Functions," *Philosophical Review* 82: 139–168.

See also **Altruism; Biological Individual; Biological Information; Conservation Biology; Ecology; Emergence; Evolution; Evolutionary Epistemology; Evolutionary Psychology; Firness; Function; Genetics; Immunology; Innate/Acquired Distinction; Laws of Nature; Mechanisms; Molecular Biology; Natural Selection; Neurobiology; Prediction; Probability; Reductionism; Scientific Models; Species; Supervenience**

PERCY WILLIAMS BRIDGMAN

(21 April 1882–20 August 1961)

Percy Williams Bridgman won a Nobel Prize for his experimental work in high-pressure physics in 1946. In addition, his constant interest in understanding and improving the scientific method led him to make significant and lasting contributions to the philosophy of science. Specifically, Bridgman is responsible for identifying and carefully explicating a method for defining scientific concepts called *operationalism.* As a philosopher of science, Bridgman recognized that the importance of Einstein's theory of special relativity was not limited to mechanics or even just to the physical sciences. He saw Einstein as looking for the meaning of simultaneity and finding that meaning by analyzing the physical operations necessary in order to use the concept in any concrete situation. Bridgman revealed his unwavering empiricist roots by claiming that scientific theories are not valuable for their so-called metaphysical consequences, but for what they actually do. Likewise, Bridgman thought that concepts in theories should not be abstract, metaphysical ideas, but rather concrete operations.

Life

Bridgman was born April 21, 1882, in Cambridge, Massachusetts, and attended public schools in Newton. He matriculated at Harvard College in 1900 and graduated *summa cum laude* in 1904. He immediately began his graduate studies in physics at Harvard, where he received his M.A. in 1905 and his Ph.D. in 1908. Bridgman remained at Harvard for his entire career, becoming an instructor of physics in 1910, assistant professor in 1913, and professor of physics in 1919. In 1926, he was named Hollis Professor of Mathematics and Natural Philosophy, and, in 1950, the Higgins University Professor. He retired in 1954 and became professor emeritus. Believing that people should not outlive their usefulness, Bridgman took his own life on August 20, 1961 (Walter 1990).

Physics

Bridgman was awarded the 1946 Nobel Prize for physics for his invention of an apparatus designed to obtain extremely high pressures and for the discoveries he made using it. Prior to Bridgman, the greatest pressures achieved in the laboratory were around 3,000 kg/cm^2 (Lindh 1946). At this time there were two limitations to reaching greater pressures. The first, which continues to be a constraint, is the strength of the containing material. However, this limitation diminishes as stronger materials are developed. The second limitation on attaining higher pressures was the problem of leakage that occurs even before the materials fail. This limitation was completely eliminated by Bridgman's apparatus. Bridgman's apparatus consists of a vessel containing the liquid to be compressed, surrounded by a soft packing material. It is designed so that the pressure in the packing material is automatically maintained at a fixed higher percentage than the pressure of the liquid, making leaks impossible (Bridgman 1946). Consequently, Bridgman was able to reach unprecedented pressures as high as 500,000 kg/cm^2. Lindh (1946) points out that the tremendous pressures made possible by his apparatus led Bridgman to discover many new polymorphous substances and new modifications of substances, and enabled him to amass a wealth of data about the properties of matter at high pressures. For example, Bridgman discovered two new modifications of both ordinary and heavy water in solid form, as well as two new modifications of phosphorous. He worked extensively on the effect of high pressures on electric resistance. This research led to the discovery of the existence of a resistance minimum for certain metals at high pressures. He also investigated the effect of high pressures on thermoelectric phenomena, heat conduction in gases, fluid viscosity, and the elastic properties of solid bodies. In addition, he made significant advances with his investigations of materials for containing substances under high pressures.

Philosophy

Long after his high-pressure results are made obsolete by new technologies and materials, Bridgman's contributions as a philosopher of science will remain influential. His most important contribution in this area is his treatment of operationalism. Bridgman should not be named the inventor of operationalism; he explicitly gives credit to Einstein for using it in developing the theory of special relativity (Bridgman 1927). In fact, Bridgman suggests that perhaps even Einstein was not the first to make progress by operationalizing concepts in scientific theories (Bridgman 1955). However, Bridgman deserves proper credit for explicitly identifying, analyzing, and explaining this important principle. The defining feature of operationalism lies in the idea that concepts are given meaning not by abstract, metaphysical musings, but by the processes used to measure them. This is contrasted with the former method exemplified by Newton's definition of absolute time as that which flows uniformly, independent of material happenings. For example, instead of an abstract definition of 'length' such as "extent in space," the operational definition of the 'length of *x*' would be the act of comparing a standard unit to *x*.

In his 1936 book *The Nature of Physical Theory,* Bridgman (1936) explains that by operationalizing concepts, their meanings are determined by physical processes—the great advantage of which is never having to retract theories:

> The more particular and important aspect of the operational significance of meaning is suggested by the fact that Einstein recognized that in dealing with physical situations the operations which give meaning to our physical concepts should properly be physical operations, actually carried out. For in so restricting the permissible operations, our theories reduce in the last analysis to descriptions of operations actually carried out in actual situations, and so cannot involve us in inconsistency or contradiction, since these do not occur in actual physical situations. Thus is solved at one stroke the problem of so constructing our fundamental physical concepts that we shall never have to revise them in the light of new experience. (9)

Bridgman did not believe that operationalism was restricted to scientific investigation. He was clear that, for instance, any question for which he could not conceive of a process by which to check the correctness of the answer must be regarded as a meaningless question. Bridgman thought that traditional metaphysical, philosophical questions, such as whether or not we have free will, should be considered meaningless because they cannot be operationalized. In this way Bridgman's philosophy is perfectly consistent with early logical empiricism.

Criticisms of Operationalism

In *The Logic of Modern Physics* Bridgman emphasizes that "the concept is synonymous with the corresponding set of operations" (1927, 5). This claim of synonymy raises problems for operationalism. For instance, it undermines the use of qualitative and dispositional properties such as hardness, which are very difficult to operationalize. Even in the realm of quantitative properties, Hempel (1966) argues that problems arise where more than one operation is possible. Consider the example of measuring temperature with a mercury thermometer and with an alcohol thermometer. According to operationalism, these have to be considered two *different* concepts—mercury-temperature and alcohol-temperature. The operations for each concept become more and more specific to particular conditions, to the point where theories adhering strictly to Bridgman's doctrine would become overburdened; and worse, they would lose generality. Hempel (1966) claims that "this would defeat one of the principal purposes of science; namely the attainment of a simple, systematically unified account of empirical phenomena" (94). Hempel agrees that operationalism is useful in certain contexts; however, he argues that it gives only "partial interpretations" of concepts. He claims that a scientific concept cannot be understood without knowing its systematic role. Concept formation and theory formation are interdependent; new theories are often generated out of discoveries about shared characteristics between concepts in different theories. This requires a certain flexibility of concept and theory formation not consistent with operationalism. Hempel gives the example of using the Sun to measure time. If a unit is marked as the Sun's return to a point in the sky each day, it cannot be questioned that the length of days are equal—it is true by definitional convention. However, when new operations are discovered for measuring the same phenomena, such as the invention of the pendulum clock, then it becomes possible to revise previous operations that turn out to be approximations. Hempel points out that this kind of concept revision can lead to scientific progress.

Ian Nyberg

References

Bridgman, Percy Williams (1927), *The Logic of Modern Physics*. New York: The Macmillan Company.

——— (1936), *The Nature of Physical Theory*. New York: Dover Publications.

——— (1946), "General Survey of Certain Results in the Field of High-Pressure Physics," in *Nobel Lectures, Physics 1942–1962*. Amsterdam: Elsevier Publishing.

——— (1955), *Reflections of A Physicist*. New York: Philosophical Library.

Hempel, Carl G. (1966), *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall.

Lindh, A. E. (1946), "Presentation Speech for the Nobel Prize in Physics 1946," in *Nobel Lectures, Physics 1942–1962*. Amsterdam: Elsevier Publishing.

Walter, Maila L. (1990), *Science and Cultural Crisis: An Intellectual Biography of Percy Williams Bridgman (1882–1961)*. Stanford, CA: Stanford University Press.

C

RUDOLF CARNAP

(18 May 1891–14 September 1970)

Rudolf Carnap (1891–1970), preeminent member of the Vienna Circle, was one of the most influential figures of twentieth-century philosophy of science and analytic philosophy (including the philosophies of language, logic, and mathematics). The Vienna Circle was responsible for promulgating a set of doctrines (initially in the 1920s) that came to be known as logical positivism or logical empiricism (see Logical Empiricism; Vienna Circle). This set of doctrines has provided the point of departure for most subsequent developments in the philosophy of science. Consequently Carnap must be regarded as one of the most important philosophers of science of the twentieth century. Nevertheless, his most lasting positive contributions were in the philosophy of logic and mathematics and the philosophy of language. Meanwhile, his systematic but ultimately unsuccessful attempt to construct an inductive logic has been equally influential, since its failure has convinced most philosophers that such a project must fail (see Inductive Logic).

Carnap was born in 1891 in Ronsdorf, near Barmen, now incorporated into the city of Wuppertal, in Germany (Carnap 1963a). In early childhood he was educated at home by his mother, Anna Carnap (neé Dörpfeld), who had been a schoolteacher. From 1898, he attended the Gymnasium at Barmen, where the family moved after his father's death that year. In school, Carnap's chief interests were in mathematics and Latin. From 1910 to 1914 Carnap studied at the Universities of Jena and Freiburg, concentrating first on philosophy and mathematics and, later, on philosophy and physics. Among his teachers in Jena were Bruno Bauch, a prominent neo-Kantian, and Gottlob Frege, a founder of the modern theory of quantification in logic. Bauch impressed upon him the power of Kant's conception that the geometrical structure of space was determined by the form of pure intuition. Though Carnap was impressed by Frege's ongoing philosophical projects, Frege's real (and lasting) influence came only later through a study of his writings. Carnap's formal intellectual work was interrupted between 1914 and 1918 while he did military service during World War I. His political views had already been of a mildly socialist/pacifist nature. The horrors of the war served to make them more explicit and more conscious, and to codify them somewhat more rigorously.

Space

After the war, Carnap returned to Jena to begin research. His contacts with Hans Reichenbach and others pursuing philosophy informed by current science began during this period (see Reichenbach, Hans). In 1919 he read Whitehead and Russell's *Principia Mathematica* and was deeply influenced by the clarity of thought that could apparently be achieved through symbolization (see Russell, Bertrand). He began the construction of a putative axiom system for a physical theory of space-time. The physicists—represented by Max Wien, head of the Institute of Physics at the University of Jena—were convinced that the project did not belong in physics. Meanwhile, Bauch was equally certain that it did not belong in philosophy. This incident was instrumental in convincing Carnap of the institutional difficulties faced in Germany of doing interdisciplinary work that bridged the chasm between philosophy and the natural sciences. It also probably helped generate the attitude that later led the logical empiricists to dismiss much of traditional philosophy, especially metaphysics. By this point in his intellectual development (the early 1920s) Carnap was already a committed empiricist who, nevertheless, accepted both the analyticity of logic and mathematics and the Frege-Russell thesis of logicism, which required that mathematics be formally constructed and derived from logic (see Analyticity).

Faced with this lack of enthusiasm for his original project in Jena, Carnap (1922) abandoned it to write a dissertation on the philosophical foundations of geometry, which was subsequently published as *Der Raum.* Most traditional commentators have regarded the dissertation as a fundamentally neo-Kantian work because it included a discussion of "intuitive space," determined by pure intuition, independent of all contingent experience, and distinct from both mathematical (or formal) space and physical space (see Friedman 1999). However, recent reinterpretations argue for a decisive influence of Husserl (Sarkar 2003). In contrast to Kant, Carnap restricted what could be grasped by pure intuition to some topological and metric properties of finite local regions of space. He identifies this intuitive space with an infinitesimal space and goes on to postulate that a global space may be constructed from it by iterative extension. In agreement with Helmholtz and Moritz Schlick (a physicist-turned-philosopher, and founder of the Vienna Circle) (see Schlick, Moritz), the geometry of physical space was regarded as an empirical matter. Carnap included a discussion of the role of non-Euclidean geometry in Einstein's theory of general relativity. By distinguishing among intuitive, mathematical, and physical spaces, Carnap attempted to resolve the apparent differences among philosophers, mathematicians, and physicists by assigning the disputing camps to different discursive domains. In retrospect, this move heralded what later became the most salient features of Carnap's philosophical work: tolerance for diverse points of view (so long as they met stringent criteria of clarity and rigor) and an assignment of these viewpoints to different realms, the choice between which is to be resolved not by philosophically substantive (e.g., epistemological) criteria but by pragmatic ones (see Conventionalism).

The Constructionist Phase

During the winter of 1921, Carnap read Russell's *Our Knowledge of the External World.* Between 1922 and 1925, this work led him (Carnap 1963a) to begin the analysis that culminated in *Der logische Aufbau der Welt* ([1928] 1967), which is usually regarded as Carnap's first major work. The purpose of the *Aufbau* was to construct the everyday world from a phenomenalist basis (see Phenomenalism). The phenomenalist basis is an epistemological choice (§§54, 58). Carnap distinguished between four domains of objects: autopsychological, physical, heteropsychological, and cultural (§58). The first of these consists of objects of an individual's own psychology; the second, of physical entities (Carnap does not distinguish between everyday material objects and the abstract entities of theoretical physics); the third consists of the objects of some other individual's psychology; and the fourth, of cultural entities (*geistige Gegenstände*), which include historical and sociological phenomena.

From Carnap's ([1928] 1967) point of view, "[a]n object ... is called *epistemically primary* relative to another one ... if the second one is recognized through the mediation of the first and thus presupposes, for its recognition, the recognition of the first" (§54). Autopsychological objects are epistemically primary relative to the others in this sense. Moreover, physical objects are epistemically primary to heteropsychological ones because the latter can be recognized only through the mediation of the former—an expression on a face, a reading in an instrument, etc. Finally, heteropsychological objects are epistemically primary relative to cultural ones for the same reason.

The main task of the *Aufbau* is construction, which Carnap ([1928] 1967) conceives of as the converse of what he regarded as reduction (which is far from what was then—or is now—conceived of as "reduction" in Anglophone philosophy) (see Reductionism):

> [*A*]*n object is 'reducible' to others . . . if all statements about it can be translated into statements which speak only about these other objects . . .* . By *constructing* a concept from other concepts, we shall mean the indication of its "constructional definition" on the basis of other concepts. By a *constructional definition* of the concept *a* on the basis of the concepts *b* and *c*, we mean a rule of translation which gives a general indication how any propositional function in which *a* occurs may be transformed into a coextensive propositional function in which *a* no longer occurs, but only *b* and *c*. If a concept is reducible to others, then it must indeed be possible to construct it from them (§35).

However, construction and reduction present different formal problems because, except in some degenerate cases (such as explicit definition), the transformations in the two directions may not have any simple explicit relation to each other. The question of reducibility/constructibility is distinct from that of epistemic primacy. In an important innovation in an empiricist context, Carnap argues that both the autopsychological and physical domains can be reduced to each other (in his sense). Thus, at the formal level, either could serve as the basis of the construction. It is epistemic primacy that dictates the choice of the former.

Carnap's task, ultimately, is to set up a constructional system that will allow the construction of the cultural domain from the autopsychological through the two intermediate domains. In the *Aufbau*, there are only informal discussions of how the last two stages of such a construction are to be executed; only the construction of the physical from the autopsychological is fully treated formally. As the basic units of the constructional system, Carnap chose what he calls "elementary experiences" (*Elementarerlebnisse,* or *elex*) (an extended discussion of Carnap's construction is to be found in Goodman 1951, Ch. 5). These are supposed to be instantaneous cross-sections of the stream of experience—or at least bits of that stream in the smallest perceivable unit of time—that are incapable of further analysis. The only primitive relation that Carnap introduces is "recollection of similarity" (*Rs*). (In the formal development of the system, *Rs* is introduced first and the *elex* are defined as the field of *Rs.*) The asymmetry of *Rs* is eventually exploited by Carnap to introduce temporal ordering.

Since the *elex* are elementary, they cannot be further analyzed to define what would be regarded as constituent qualities of them such as partial sensations or intensity components of a sensation. Had the *elex* not been elementary, Carnap could have used "proper analysis" to define such qualities by isolating the individuals into classes on the basis of having a certain (symmetric) relationship with each other. Carnap defines the process of "quasi-analysis" to be formally analogous to proper analysis but only defining "quasi-characteristics" or "quasi-constituents" because the *elex* are unanalyzable. Thus, if an *elex* is both *c* in color and *t* in temperature, *c* or *t* can be defined as classes of every *elex* having *c* or *t,* respectively. However, to say that *c* or *t* is a quality would imply that an *elex* is analyzable into simpler constituents. Quasi-analysis proceeds formally in this way (as if it is proper analysis) but defines only quasi-characteristics, thus allowing each *elex* to remain technically unanalyzable Quasi-analysis based on the relation "part similarity" (*Ps*), itself defined from *Rs*, is the central technique of the *Aufbau.* It is used eventually to define sense classes and, then, the visual sense, visual field places, the spatial order of the visual field, the order of colors and, eventually, sensations. Thus the physical domain is constructed out of the autopsychological. Carnap's accounts of the construction between the other two domains remain promissory sketches.

Carnap was aware that there were unresolved technical problems with his construction of the physical from the autopsychological, though he probably underestimated the seriousness of these problems. The systematic problems are that when a quality is defined as a class selected by quasi-analysis on the basis of a relation: (i) two (different) qualities that happen always to occur together (say, red and hot) will never be separated, and (ii) quality classes may emerge in which any two members bear some required relation to each other, but there may yet be no relation that holds between all members of the class. Carnap's response to these problems was extrasystematic: In the complicated construction of the world from the *elex,* he hoped that such examples would never or only very rarely arise. Nevertheless, because of these problems, and because the other constructions are not carried out, the attitude of the *Aufbau* is tentative and exploratory: The constructional system is presented as essentially unfinished. (Goodman 1951 also provides a lucid discussion of these problems.)

Some recent scholarship has questioned whether Carnap had any traditional epistemological concerns in the *Aufbau.* In particular, Friedman (e.g., 1992) has championed the view that Carnap's concerns in that work are purely ontological: The *Aufbau* is not concerned with the question of the source or status of knowledge of the external world (see Empiricism); rather, it investigates the bases on which such a world may be constructed (see Richardson 1998). Both Friedman and Richardson—as well as Sauer (1985) and Haack (1977) long before them—emphasize the Kantian roots of the *Aufbau.* If this reinterpretation is correct, then what exactly the *Aufbau* owes to Russell (and traditional empiricism) becomes uncertain. However, as Putnam (1994, 281) also points out, this reinterpretation goes too far: Though the project of the *Aufbau* is not identical to that of Russell's external world program, there is sufficient congruence between the two projects for Carnap to have correctly believed that he was carrying out Russell's program. In particular, the formal constructions of the *Aufbau* are a necessary prerequisite for the development of the epistemology that Russell had in mind: One must be able to construct the world formally from a phenomenalist basis before one can suggest that this construction shows that the phenomena are the source of knowledge of the world. Moreover, this reinterpretation ignores the epistemological remarks scattered throughout the *Aufbau* itself, including Carnap's concern for the epistemic primacy of the basis he begins with. Savage (2003) has recently pointed out that the salient difference between Russell's and Carnap's project is that whereas the former chose sense data as his point of departure, the latter chose elementary experiences. But this difference is simply a result of Carnap's having accepted the results of Gestalt psychology as having definitively shown what may be taken as individual experiential bases; other than that, that is, with respect to the issue of empiricism, it has no philosophical significance.

In any case, by this time of his intellectual development, Carnap had fully endorsed not only the logicism of the *Principia*, but also the form that Whitehead and Russell had given to logic (that is, the ramified theory of types including the axioms of infinity and reducibility) in that work. However, Henri Poincaré also emerges as a major influence during this period. Carnap did considerable work on the conceptual foundations of physics in the 1920s, and some of this work—in particular, his analysis of the relationship between causal determination and the structure of space—shows strong conventionalist attitudes (Carnap 1924; see also 1923 and 1926) (see Conventionalism; Poincaré, Henri).

Viennese Positivism

In 1926, at Schlick's invitation, Carnap moved to Vienna to become a *Privatdozent* (instructor) in philosophy at the University of Vienna for the next five years (see Vienna Circle). An early version of the *Aufbau* served as his *Habilitationsschrift.* He was welcomed into the Vienna Circle, a scientific philosophy discussion group organized by (and centered around) Schlick, who had occupied the chair for philosophy of the inductive sciences since 1922. In the meetings of the Vienna Circle, the typescript of the *Aufbau* was read and discussed. What Carnap seems to have found most congenial in the Circle—besides its members' concern for science and competence in modern logic—was their rejection of traditional metaphysics. Over the years, besides Carnap and Schlick, the Circle included Herbert Feigl, Kurt Gödel, Hans Hahn, Karl Menger, Otto Neurath, and Friedrich Waismann, though Gödel would later claim that he had little sympathy for the antimetaphysical position of the other members. The meetings of the Circle were characterized by open, intensely critical, discussion with no tolerance for ambiguity of formulation or lack of rigor in demonstration. The members of the Circle believed that philosophy was a collective enterprise in which progress could be made. These attitudes, even more than any canonical set of positions, characterized the philosophical movement—initially known as logical positivism and later as logical empiricism—that emerged from the work of the members of the Circle and a few others, especially Hans Reichenbach (see Reichenbach, Hans). However, besides rejecting traditional metaphysics, most members of the Circle accepted logicism and a sharp distinction between analytic and synthetic truths. The analytic was identified with the a priori; the synthetic with the *a posteriori* (see Analyticity). A. J. Ayer, who attended some meetings of the Circle in 1933 (after Carnap had left—see below), returned to Britain and published *Language, Truth and Logic* (Ayer 1936) (see Ayer, Alfred Jules). This short book did much to popularize the views of the Vienna Circle among Anglophone philosophers, though it lacks the sophistication that is found in the writings of the members of the Circle, particularly Carnap.

Under Neurath's influence, during his Vienna years, Carnap abandoned the phenomenalist language he had preferred in the *Aufbau* and came to accept physicalism (see Neurath, Otto; Physicalism). The epistemically privileged language is one in which sentences reporting empirical knowledge of the world ("protocol sentences") employ terms referring to material bodies and their observable properties (see Phenomenalism, Protocol Sentences). From Carnap's point of view, the chief advantage of a physicalist language is its intersubjectivity. Physicalism, moreover, came hand-in-hand with the thesis of the "unity of science," that is, that the different empirical sciences (including the social sciences) were merely different branches of a single unified science (see Unity of Science Movement). To defend this thesis, it had to be demonstrated that psychology could be based on a physicalist language In an important paper only published somewhat later, Carnap ([1932] 1934) attempted that demonstration (see Unity and Disunity of Science). Carnap's adoption of physicalism was final; he never went back to a phenomenalist language. However, what he meant by "physicalism" underwent radical transformations over the years. By the end of his life, it meant no more than the adoption of a nonsolipsistic language, that is, one in which intersubjective is possible (Carnap 1963b).

In the Vienna Circle, Wittgenstein's *Tractatus* was discussed in detail. Carnap found Wittgenstein's rejection of metaphysics concordant with the views he had developed independently. Partly because of Wittgenstein's influence on some members of the Circle (though not Carnap), the rejection of metaphysics took the form of an assertion that the sentences of metaphysics are meaningless in the sense of being devoid of cognitive content. Moreover, the decision whether a sentence is meaningful was to be made on the basis of the principle of verifiability, which claims that the meaning of a sentence is given by the conditions of its (potential) verification (see Verifiability). Observation terms are directly meaningful on this account (see Observation). Theoretical terms acquire meaning only through explicit definition from observation terms. Carnap's major innovation in these discussions within the Circle was to suggest that even the thesis of realism—asserting the "reality" of the external world—is meaningless, a position not shared by Schlick, Neurath, or Reichenbach. Problems generated by meaningless questions became the celebrated "pseudo-problems" of philosophy (Carnap [1928] 1967).

Wittgenstein's principle of verifiability posed fairly obvious problems in any scientific context. No universal generalization can ever be verified. Perhaps independently, Karl Popper perceived the same problem (see Popper, Karl). This led him to replace the requirement of verifiability with that of falsifiability, though only as a criterion to demarcate science from metaphysics, and not as one to be also used to demarcate meaningful from meaningless claims. It is also unclear what the status of the principle itself is, that is, whether it is meaningful by its own criterion of meaningfulness. Carnap, as well as other members of the Vienna Circle including Hahn and Neurath, realized that a weaker criterion of meaningfulness was necessary. Thus began the program of the "liberalization of empiricism." There was no unanimity within the Vienna Circle on this point. The differences between the members are sometimes described as those between a conservative "right" wing, led by Schlick and Waismann, which rejected both the liberalization of empiricism and the epistemological antifoundationalism of the move to physicalism, and a radical "left" wing, led by Neurath and Carnap, which endorsed the opposite views. The "left" wing also emphasized fallibilism and pragmatics; Carnap went far enough along this line to suggest that empiricism itself was a proposal to be accepted on pragmatic grounds. This difference also reflected political attitudes insofar as Neurath and, to a lesser extent, Carnap viewed science as a tool for social reform.

The precise formulation of what came to be called the criterion of cognitive significance took three decades (see Hempel 1950; Carnap 1956 and 1961) (see Cognitive Significance). In an important pair of papers, "Testability and Meaning," Carnap (1936–1937) replaced the requirement of verification with that of confirmation; at this stage, he made no attempt to quantify the latter. Individual terms replace sentences as the units of meaning. Universal generalizations are no longer problematic; though they cannot be conclusively verified, they can yet be confirmed. Moreover, in "Testability and Meaning," theoretical terms no longer require explicit definition from observational ones in order to acquire meaning; the connection between the two may be indirect through a system of implicit definitions. Carnap also provides an important pioneering discussion of disposition predicates.

The Syntactic Phase

Meanwhile, in 1931, Carnap had moved to Prague, where he held the chair for natural philosophy at the German University until 1935, when, under the shadow of Hitler, he emigrated to the United States. Toward the end of his Vienna years, a subtle but important shift in Carnap's philosophical interests had taken place. This shift was from a predominant concern for the foundations of physics to that for the foundations of mathematics and logic, even though he remained emphatic that the latter were important only insofar as they were used in the empirical sciences, especially physics.

In Vienna and before, following Frege and Russell, Carnap espoused logicism in its conventional sense, that is, as the doctrine that held that the concepts of mathematics were definable from those of logic, and the theorems of mathematics were derivable from the principles of logic. In the aftermath of Gödel's (1931) incompleteness theorems, however, Carnap abandoned this type of logicism and opted instead for the requirement that the concepts of mathematics and logic always have their customary (that is, everyday) interpretation in all contexts. He also began to advocate a radical conventionalism regarding what constituted "logic."

Besides the philosophical significance of Gödel's results, what impressed Carnap most about that work was Gödel's arithmetization of syntax. Downplaying the distinction between an object language and its metalanguage, Carnap interpreted this procedure as enabling the representation of the syntax of a language within the language itself. At this point Carnap had not yet accepted the possibility of semantics, even though he was aware of some of Tarski's work and had had some contact with the Polish school of logic. In this context, the representation of the syntax of a language within itself suggested to Carnap that all properties of a language could be studied within itself through a study of syntax.

These positions were codified in Carnap's major work from this period, *The Logical Syntax of Language* (Carnap 1934b and 1937). The English translation includes material that had to be omitted from the German original due to a shortage of paper; the omitted material was separately published in German as papers (Carnap 1934a and 1935). Conventionalism about logic was incorporated into the well-known principle of tolerance:

> *It is not our business to set up prohibitions but to arrive at conventions* [about what constitutes a logic] . . . . *In logic, there are no morals.* Every one is at liberty to build up his own logic, i.e., his own form of language, as he wishes. All that is required is that, if he wishes to discuss it, he must state his method clearly, and give syntactic rules instead of philosophical arguments. (Carnap 1937, 51–52; emphasis in the original)

Logic, therefore, is nothing but the syntax of language.

In *Syntax*, the principle of tolerance allows Carnap to navigate the ongoing disputes between logicism, formalism, and intuitionism/constructivism in the foundations of mathematics without abandoning any insight of interest from these schools. Carnap begins with a detailed study of the construction of two languages, I and II. The last few sections of *Syntax* also present a few results regarding the syntax of any language and also discuss the philosophical ramifications of the syntactic point of view. (Sarkar 1992 attempts a comprehensible reconstruction of the notoriously difficult formalism of *Syntax.*)

Language I, which Carnap calls "definite," is intended as a neutral core of all logically interesting languages, neutral enough to satisfy the strictures of almost any intuitionist or constructivist. It permits the definition of primitive recursive arithmetic and has bounded quantification (for all x up to some upper bound) but not much more. Its syntax is fully constructed formally. Language II, which is "indefinite" for Carnap, is richer. It includes Language I and has sufficient resources for the formulation of all of classical mathematics, and is therefore nonconstructive. Moreover, Carnap permits descriptive predicates in each language. Thus, the resources of Language II are strong enough to permit, in principle, the formulation of classical physics. The important point is that because of the principle of tolerance, the choice between Languages I and II or, for that matter, any other syntactically specified language, is not based on factual considerations. If one wants to use mathematics to study physics in the customary way, Language II is preferable, since as yet, nonconstructive mathematics remains necessary for physics. But the adoption of Language II, dictated by the pragmatic concern for doing physics, does not make Language I incorrect. This was Carnap's response to the foundational disputes of mathematics: By tolerance they are defined out of existence.

The price paid if one adopts the principle of tolerance is a radical conventionalism about what constitutes logic. Conventionalism, already apparent in Carnap's admission of both a phenomenalist and a physicalist possible basis for construction in the *Aufbau*, and strongly present in the works on the foundations of physics in the 1920s, had now been extended in *Syntax* to logic. As a consequence, what might be considered to be the most important question in any mathematical or empirical context—the choice of language—became pragmatic. This trend of relegating troublesome questions to the realm of pragmatics almost by fiat, thereby excusing them from systematic philosophical exploration, became increasingly prevalent in Carnap's views as the years went on.

Syntax contained four technical innovations in logic that are of significance: (i) a definition of analyticity that, as was later shown by S. C. Kleene, mimicked Tarski's definition of truth for a formalized language; (ii) a proof, constructed by Carnap independently of Tarski, that truth cannot be defined as a syntactic predicate in any consistent formalized language; (iii) a rule for infinite induction (in Language I) that later came to be called the omega rule; and (iv), most importantly, a generalization of Gödel's first incompleteness theorem that has come to be called the fixed-point lemma. With respect to (iv), what Carnap proved is that in a language strong enough to permit arithmetization, for any syntactic predicate, one can construct a sentence that would be interpreted as saying that it satisfies that predicate. If the chosen predicate is unprovability, one gets Gödel's result.

Besides the principle of tolerance, the main philosophical contribution of *Syntax* was the thesis that philosophy consisted of the study of logical syntax. Giving a new twist to the Vienna Circle's claim that metaphysical claims were meaningless, Carnap argues and tries to show by example that sentences making metaphysical claims are all syntactically ill-formed. Moreover, since the arithmetization procedure shows that all the syntactic rules of a language can be formulated within the language, even the rules that determine what sentences are meaningless can be constructed within the language. All that is left for philosophy is a study of the logic of science. But, as Carnap (1937) puts it: "The *logic of science* (logical methodology) is nothing else than the *syntax of the language of science*. . . . To share this view is to *substitute logical syntax for philosophy*" (7–8; emphasis in original). The claims of *Syntax* are far more grandiose—and more flamboyant—than anything in the *Aufbau.*

Semantics

In the late 1930s, Carnap abandoned the narrow syntacticism of *Syntax* and, under the influence of Tarski and the Polish school of logic, came to accept semantics. With this move, Carnap's work enters its final mature phase. For the first time, he accepted that the concept of truth can be given more than pragmatic content. Thereupon, he turned to the systematization of semantics with characteristic vigor, especially after his immigration to the United States, where he taught at the University of Chicago from 1936 to 1952. In his contribution to the *International Encyclopedia of Unified Science* (Carnap 1939), on the foundations of logic and mathematics, the distinctions among syntactic, semantic, and pragmatic considerations regarding any language are first presented in their mature form.

Introduction to Semantics, which followed in 1942, develops semantics systematically. In *Syntax* Carnap had distinguished between two types of transformations on sentences: those involving "the method of derivation" or "*d*-method," and those involving the "method of consequence" or "*c*-method." Both of these were supposed to be syntactic, but there is a critical distinction between them. The former allows only a finite number of elementary steps. The latter places no such restriction and is, therefore, more "indefinite." Terms defined using the *d*-method ("*d*-terms") include "derivable," "demonstrable," "refutable," "resoluble," and "irresoluble"; the corresponding *c*-terms are "consequence," "analytic," "contradictory," "L-determinate," and "synthetic." After the conversion to semantics, Carnap proposed that the *c*-method essentially captured what semantics allowed; the *c*-terms referred to semantic concepts.

Thus semantics involves a kind of formalization, though one that is dependent on stronger inference rules than the syntactical ones. In this sense, as Church (1956, 65) has perceptively pointed out, Carnap—and Tarski—reduce semantics to formal rules, that is, syntax. Thus emerges the interpretation of deductive logic that has since become the textbook version, so commonly accepted that is has become unnecessary to refer to Carnap when one uses it. For Carnap, the semantic move has an important philosophical consequence: Philosophy is no longer to be replaced just by the syntax of the language of science; rather, it is to be replaced by the syntax and the semantics of the language of science.

Carnap's (1947) most original—and influential—work in semantics is *Meaning and Necessity*, where the basis for an intensional semantics was laid down. Largely following Frege, intensional concepts are distinguished from extensional ones. Semantical rules are introduced and the analytic/synthetic distinction is clarified by requiring that any definition of analyticity must satisfy the (meta-) criterion that analytic sentences follow from the semantical rules alone. By now Carnap had fully accepted that semantic concepts and methods are more fundamental than syntactic ones: The retreat from the flamboyance of *Syntax* was complete. The most important contribution of *Meaning and Necessity* was the reintroduction into logic, in the new intensional framework, of modal concepts that had been ignored since the pioneering work of Lewis (1918). In the concluding chapter of his book, Carnap introduced an operator for necessity, gave semantic rules for its use, and showed how other modal concepts such as possibility, impossibility, necessary implication, and necessary equivalence can be defined from this basis.

By this point, Carnap had begun to restrict his analyses to exactly constructed languages, implicitly abandoning even a distant hope that they would have any direct bearing on natural languages. The problem with the latter is that their ambiguities made them unsuited for the analysis of science, which, ultimately, remained the motivation of all of Carnap's work. Nevertheless, Carnap's distinction between the analytic and the synthetic came under considerable criticism from many, including Quine (1951), primarily on the basis of considerations about natural languages (see Analyticity; Quine, Willard Van). Though philosophical fashion has largely followed Quine on this point, at least until recently, Carnap was never overly impressed by this criticism (Stein 1992). The analytic/synthetic distinction continued to be fundamental to his views, and, in a rejoinder to Quine, Carnap argued that nothing prevented empirical linguistics from exploring intensions and thereby discovering cases of synonymy and analyticity (Carnap 1955).

Carnap's (1950a) most systematic exposition of his final views on ontology is also from this period. A clear distinction is maintained between questions that are internal to a linguistic framework and questions that are external to it. The choice of a linguistic framework is to be based not on cognitive but on pragmatic considerations. The external question of "realism," which ostensibly refers to the "reality" of entities of a framework in some sense independent of it, rather than to their "reality" within it after the framework has been accepted, is rejected as noncognitive (see Scientific Realism). This appears to be an anti-"realist" position, but it is not in the sense that within a framework, Carnap is tolerant of the abstract entities that bother nominalists. The interesting question becomes the pragmatic one, that is, what frameworks are fruitful in which contexts, and Carnap's attitude toward the investigation of various alternative frameworks remains characteristically and consistently tolerant.

Carnap continued to explore questions about the nature of theoretical concepts and to search for a criterion of cognitive significance, preoccupations of the logical empiricists that date back to the Vienna Circle. Carnap (1956) published a detailed exposition of his final views regarding the relation between the theoretical and observational parts of a scientific language. This paper emphasizes the methodological and pragmatic aspects of theoretical concepts. It also contains his most subtle, though not his last, attempt to explicate the notion of the cognitive significance of a term and thus establish clearly the boundary between scientific and nonscientific discourse. However, the criterion he formulates makes theoretical terms significant only with respect to a class of terms, a theoretical language, an observation language, correspondence rules between them, and a theory. Relativization to a theory is critical to avoiding the problems that beset earlier attempts to find such a criterion. Carnap proves several theorems that are designed to show that the criterion does capture the distinction between scientific and nonscientific discourse. This criterion was criticized by Roozeboom (1960) and Kaplan (1975), but these criticisms depend on modifying Carnap's original proposal in important ways. According to Kaplan, Carnap accepted his criticism, though there is apparently no independent confirmation of that fact. However, Carnap (1961) did turn to a different formalism (Hilbert's ϵ-operator) in what has been interpreted as his last attempt to formulate such a criterion (Kaplan 1975), and this may indicate dissatisfaction with the 1956 attempt. If so, it remains unclear why: That attempt did manage to avoid the technical problems associated with the earlier attempts of the logical empiricists (see Cognitive Significance).

Probability and Inductive Logic

From 1941 onward, Carnap also began a systematic attempt to analyze the concepts of probability and to formulate an adequate inductive logic (a logic of confirmation), a project that would occupy him for the rest of his life. Carnap viewed this work as an extension of the semantical methods that he had been developing for the last decade. This underscores an interesting pattern in Carnap's intellectual development. Until the late 1930s Carnap viewed syntactic categories only as nonpragmatically specifiable; questions of truth and confirmation were viewed as pragmatic. His conversion to semantics saw the recovery of truth from the pragmatic to the semantic realm. Now, confirmation followed truth down the same pathway.

In *Logical Foundations of Probability* (1950b), his first systematic analysis of probability, Carnap distinguished between two concepts of probability: "statistical probability," which was the relevant concept to be used in empirical contexts and generally estimated from the relative frequencies of events, and "logical probability," which was to be used in contexts such as the confirmation of scientific hypotheses by empirical data. Though the latter concept, usually called the "logical interpretation" of probability, went back to Keynes (1921), Carnap provides its first systematic explication (see Probability).

Logical probability is explicated from three different points of view (1950b, 164–8): (i) as a conditional probability $c(h,e)$, which measures the degree of confirmation of a hypothesis h on the basis of evidence e (if $c(h,e) = r$, then r is determined by logical relations between h and e); (ii) as a rational degree of belief or fair betting quotient (if $c(h,e) = r$, then r represents a fair bet on h if e correctly describes the total knowledge available to a bettor); and (iii) as the limit of relative frequencies in some cases. According to Carnap, the first of these, which specifies a confirmation function ("*c*-function"), is the concept that is most relevant to the problem of induction. In the formal development of the theory, probabilities are associated with sentences of a formalized language.

In *Foundations*, Carnap (1950b) believed that a unique measure $c(h,e)$ of the degree of confirmation can be found, and he even proposed one (*viz*., Laplace's rule of succession), though he could not prove its uniqueness satisfactorily. His general strategy was to augment the standard axioms of the probability calculus by a set of "conventions on adequacy" (285), which turned out to be equivalent to assumptions about the rationality of degrees of belief that had independently been proposed by both Ramsey and de Finetti (Shimony 1992). In a later work, *The Continuum of Inductive Methods*, using the conventions on adequacy and some plausible symmetry principles, Carnap (1952) managed to show that all acceptable *c*-functions could be parameterized by a single parameter, a real number, $\lambda \in [0,\infty]$. The trouble remained that there is no intuitively appealing a priori strategy to restrict λ to some preferably very small subset of $[0,\infty]$. At one point, Carnap even speculated that it would have to be fixed empirically. Unfortunately, some higher-order induction would then be required to justify the procedure for its estimation, and potentially, this leads to infinite regress (see Confirmation Theory; Inductive Logic).

Carnap spent 1952–1954 at the Institute for Advanced Study at Princeton, New Jersey, where he continued to work on inductive logic, often in collaboration with John Kemeny. He also returned to the foundations of physics, apparently motivated by a desire to trace and explicate the relations between the physical concept of entropy and an abstract concept of entropy appropriate for inductive logic. His discussion with physicists proved to be disappointing and he did not publish his results. (These were edited and published by Abner Shimony [Carnap 1977] after Carnap's death.)

In 1954 Carnap moved to the University of California at Los Angeles to assume the chair that had become vacant with Reichenbach's death in 1953. There he continued to work primarily on inductive logic, often with several collaborators, over the next decade. There were significant modifications of his earlier attempts to formulate a systematic inductive logic (see Carnap and Jeffrey 1971 and Jeffrey 1980. An excellent introduction to this part of Carnap's work on inductive logic is Hilpinen 1975). Obviously impressed by the earlier work of Ramsey and de Finetti, Carnap (1971a) returned to the second of his three 1950 explications of logical probability and emphasized the use of inductive logic in decision problems.

More importantly, Carnap, in "A Basic System of Inductive Logic" (1971b and 1980), finally recognized that attributing probabilities to sentences was too restrictive. If a conceptual system uses real numbers and real-valued functions, no language can express all possible cases using only sentences or classes of sentences. Because of this, he now began to attribute probabilities to events or propositions (which are taken to be synonymous). This finally brought some concordance between his formal methods and those of mathematical statisticians interested in epistemological questions. Propositions are identified with sets of models; however, the fields of the sets are defined using the atomic propositions of a formalized language. Thus, though probabilities are defined as measures of sets, they still remain relativized to a particular formalized language. Because of this, and because the languages considered remain relatively simple (mostly monadic predicate languages), much of this work remains similar to the earlier attempts.

By this point Carnap had abandoned the hope of finding a unique *c*-function. Instead, he distinguished between subjective and objective approaches in inductive logic. The former emphasizes individual freedom in the choice of necessary conventions; the latter emphasizes the existence of limitations. Though Carnap characteristically claimed to keep an open mind about these two approaches, his emphasis was on finding rational a priori principles that would systematically limit the choice of *c*-functions. Carnap was still working on this project when he died on September 14, 1970. He had not finished revising the last sections of the second part of the "Basic System," both parts of which were published only posthumously.

Toward the end of his life, Carnap's concern for political and social justice had led him to become an active supporter of an African-American civil rights organization in Los Angeles. According to Stegmüller (1972, lxvi), the "last photograph we have of Carnap shows him in the office of this organization, in conversation with various members. He was the only white in the discussion group."

The Legacy

Thirty-five years after Carnap's death it is easier to assess Carnap's legacy, and that of logical empiricism, than it was in the 1960s and 1970s, when a new generation of analytic philosophers and philosophers of science apparently felt that they had to reject that work altogether in order to be able to define their own philosophical agendas. This reaction can itself be taken as evidence of Carnap's seminal influence, but, nevertheless, it is fair to say that Carnap and logical empiricism fell into a period of neglect in the 1970s from which it began to emerge only in the late 1980s and early 1990s. Meanwhile it became commonplace among philosophers to assume that Carnap's projects had failed.

Diagnoses of this failure have varied. For some it was a result of the logical empiricists' alleged inability to produce a technically acceptable criterion for cognitive significance (see Cognitive Significance). For others, it was because of Quine's dicta against the concept of analyticity and the analytic/synthetic distinction (see Analyticity; Quine, Willard Van). Some took Popper's work to have superseded that of Carnap and the logical empiricists (see Popper, Karl Raimund). Many viewed Thomas Kuhn's seminal work on scientific change to have shown that the project of inductive logic was misplaced; they, and others, generally regarded Carnap's attempt to explicate inductive logic to have been a failure (see Kuhn, Thomas; Scientific Change). Finally, a new school of "scientific realists" attempted to escape Carnap's arguments against external realism (see Realism).

There can be little doubt that Carnap's project of founding inductive logic has faltered. He never claimed that he had gone beyond preliminary explorations of possibilities, and, though there has been some work since, by and large, epistemologists of science have abandoned that project in favor of less restrictive formalisms, for instance, those associated with Bayesian or Neyman-Pearson statistics (see Bayesianism; Statistics, Philosophy of). But, with respect to every other case mentioned in the last paragraph, the situation is far less clear. It has already been noted that Carnap's final criterion for cognitive significance does not suffer from any technical difficulty no matter what its other demerits may be. Quine's dicta against analyticity no longer appear as persuasive as they once did (Stein 1992); Quine's preference for using natural—rather than formalized—language in the analysis of science has proved to be counterproductive; and his program of naturalizing epistemology has yet to live up to its initial promise. Putnam's "internal realism" is based on and revives Carnap's views on ontology, and Kuhn is perhaps now better regarded as having contributed significantly to the sociology rather than to the epistemology of science.

However, to note that some of the traditionally fashionable objections to Carnap and logical empiricism cannot be sustained does not show that that work deserves a positive assessment on its own. There still remains the question: What, exactly, did Carnap contribute? The answer turns out to be straightforward: The textbook picture of deductive logic that is in use today is the one that Carnap produced in the early 1940s after he came to acknowledge the possibility of semantics. The fixed-point lemma has turned out to be an important minor contribution to logic. The reintroduction of modal logic into philosophy opened up new vistas for Kripke and others in the 1950s and 1960s. Carnap's views on ontology continue to influence philosophers today. Moreover, even though the project of inductive logic seems unsalvageable to most philosophers, it is hard to deny that Carnap managed to clarify significantly the ways in which concepts of probability must be deployed in the empirical sciences and why the problem of inductive logic is so difficult. But, most of all, Carnap took philosophy to a new level of rigor and clarity, accompanied by an open-mindedness (codified in the principle of tolerance) that, unfortunately, is not widely shared in contemporary analytic philosophy.

Sahotra Sarkar

References

Ayer, A. J. (1936), *Language, Truth and Logic*. London: Gollancz.

Carnap, R. (1922), *Der Raum*. Berlin: von Reuther and Reichard.

——— (1923), "Über die Aufgabe der Physik und die Anwendung des Grundsatzes der Einfachsteit," *Kant-Studien* 28: 90–107.

——— (1924), "Dreidimensionalität des Raumes und Kausalität: Eine Untersuchung über den logischen Zusammenhang zweier Fiktionen," *Annalen der Philosophie und philosophischen Kritik* 4: 105–130.

——— (1926), *Physikalische Begriffsbildung*. Karlsruhe, Germany: Braun.

——— ([1928] 1967), *The Logical Structure of the World and Pseudoproblems in Philosophy*. Berkeley and Los Angeles: University of California Press.

——— ([1932] 1934. *The Unity of Science*. London: Kegan Paul, Trench, Trubner & Co.

——— (1934a), “Die Antinomien und die Unvollständigkeit der Mathematik,” *Monatshefte für Mathematik und Physik* 41: 42–48.

——— (1934b), *Logische Syntax der Sprache*. Vienna: Springer.

——— (1935), “Ein Gültigskriterium für die Sätze der klassischen Mathematik,” *Monatshefte für Mathematik und Physik* 42: 163–190.

——— (1936–1937), “Testability and Meaning,” *Philosophy of Science* 3 and 4: 419–471 and 1–40.

——— (1937), *The Logical Syntax of Language*. London: Kegan Paul, Trench, Trubner & Co.

——— (1939), *Foundations of Logic and Mathematics*. Chicago: University of Chicago Press.

——— (1947), *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

——— (1950a), “Empiricism, Semantics, and Ontology,” *Revue Internationale de Philosophie* 4: 20–40.

——— (1950b), *Logical Foundations of Probability*. Chicago: University of Chicago Press.

——— (1952), *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

——— (1955), “Meaning and Synonymy in Natural Languages,” *Philosophical Studies* 7: 33–47.

——— (1956), “The Methodological Character of Theoretical Concepts,” in H. Feigl and M. Scriven (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minneapolis: University of Minnesota Press, 38–76.

——— (1961), “On the Use of Hilbert's ϵ-Operator in Scientific Theories,” in Y. Bar-Hillel, E. I. J. Poznanski, M. O. Rabin, and A. Robinson (eds.), *Essays on the Foundations of Mathematics*. Jerusalem: Magnes Press, 156–164.

——— (1963a), “Intellectual Autobiography,” in P. A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*. La Salle, IL: Open Court, 3–84.

——— (1963b), “Replies and Systematic Expositions,” in P. A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*. La Salle, IL: Open Court, 859–1013.

——— (1971a), “Inductive Logic and Rational Decisions,” in R. Carnap and R. C. Jeffrey (eds.), *Studies in Inductive Logic and Probability* (Vol. 1). Berkeley and Los Angeles: University of California Press, 5–31.

——— (1971b), “A Basic System of Inductive Logic, Part I,” in R. Carnap and R. C. Jeffrey (eds.), *Studies in Inductive Logic and Probability* (Vol. 1). Berkeley and Los Angeles: University of California Press, 33–165.

——— (1977), *Two Essays on Entropy*. Berkeley and Los Angeles: University of California Press.

——— (1980), “A Basic System of Inductive Logic, Part II,” in R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability* (Vol. 2). Berkeley and Los Angeles: University of California Press, 8–155.

Carnap, R., and R. C. Jeffrey (eds.) (1971), *Studies in Inductive Logic and Probability* (Vol. 1). Berkeley and Los Angeles: University of California Press.

Church, A. (1956), *Introduction to Mathematical Logic*. Princeton: Princeton University Press.

Friedman, M. (1992), “Epistemology in the *Aufbau*,” *Synthese* 93: 191–237.

——— (1999), *Reconsidering Logical Empiricism*. New York: Cambridge University Press.

Gödel, K. (1931), “Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme 1,” *Monatshefte für Mathematik und Physik* 38: 173–198.

Goodman, N. (1951), *The Structure of Experience*. Cambridge, MA: Harvard University Press.

Haack, S. (1977), “Carnap's *Aufbau*: Some Kantian Reflections,” *Ratio* 19: 170–175.

Hempel, C. G. (1950), “Problems and Changes in the Empiricist Criterion of Meaning,” *Revue Internationale de Philosophie* 11: 41–63.

Hilpinen, R. (1975), “Carnap's New System of Inductive Logic,” in J. Hintikka (ed.), *Rudolf Carnap, Logical Empiricist: Materials and Perspectives*. Dordrecht, Netherlands: Reidel, 333–359.

Jeffrey, R. C. (ed.) (1980), *Studies in Inductive Logic and Probability* (Vol. 2). Berkeley and Los Angeles: University of California Press.

Kaplan, D. (1975), “Significance and Analyticity: A Comment on Some Recent Proposals of Carnap,” in J. Hintikka (ed.), *Rudolf Carnap, Logical Empiricist: Materials and Perspectives*. Dordrecht, Netherlands: Reidel, 87–94.

Keynes, J. M. (1921), *A Treatise on Probability*. London: Macmillan & Co.

Lewis, C. I. (1918), *A Survey of Symbolic Logic*. Berkeley and Los Angeles: University of California Press.

Putnam, H. (1994), “Comments and Replies,” in P. Clark and B. Hale (eds.), *Reading Putnam*. Oxford: Blackwell, 242–295.

Quine, Willard Van (1951), “Two Dogmas of Empiricism,” *Philosophical Review* 60: 20–43.

Richardson, A. (1998), *Carnap's Construction of the World*. Cambridge: Cambridge University Press.

Roozeboom, W. (1960), “A Note on Carnap's Meaning Criterion,” *Philosophical Studies* 11: 33–38.

Sarkar, S. (1992), “‘The Boundless Ocean of Unlimited Possibilities’: Logic in Carnap's *Logical Syntax of Language*,” *Synthese* 93: 191–237.

——— (2003), “Husserl's Role in Carnap's *Der Raum*,” in T. Bonk (ed.), *Language, Truth and Knowledge: Contributions to the Philosophy of Rudolf Carnap*. Dordrecht, Netherlands: Kluwer, 179–190.

Sauer, W. (1985), “Carnap's ’*Aufbau*’ in Kantianishcer Sicht,” *Grazer Philosophische Studien* 23: 19–35.

Savage, C. W. (2003), “Carnap's *Aufbau* Rehabilitated,” in T. Bonk (ed.), *Language, Truth and Knowledge: Contributions to the Philosophy of Rudolf Carnap*. Dordrecht, Netherlands: Kluwer, 79–86.

Shimony, A. (1992), “On Carnap: Reflections of a Metaphysical Student,” *Synthese* 93: 261–274.

Stegmüller, W. (1972), “Homage to Rudolf Carnap,” in R. C. Buck and R. S. Cohen (eds.), *PSA 1970*. Dordrecht, Netherlands: Reidel, lii–lxvi.

Stein, H. (1992), “Was Carnap Entirely Wrong, After All?” *Synthese* 93: 275–295.

See also **Analyticity; Confirmation Theory; Cognitive Significance; Conventionalism; Hahn, Hans; Hempel, Carl Gustav; Induction, Problem of; Inductive Logic; Instrumentalism; Logical Empiricism; Neurath, Otto; Phenomenalism; Protocol Sentences; Quine, Willard Van; Reichenbach, Hans; Scientific Realism; Schlick, Moritz; Space-Time; Verifiability; Vienna Circle**

CAUSALITY

Arguably no concept is more fundamental to science than that of causality, for investigations into cases of existence, persistence, and change in the natural world are largely investigations into the causes of these phenomena. Yet the metaphysics and epistemology of causality remain unclear. For example, the ontological categories of the causal relata have been taken to be objects (Hume [1739] 1978), events (Davidson 1967), properties (Armstrong 1978), processes (Salmon 1984), variables (Hitchcock 1993), and facts (Mellor 1995). (For convenience, causes and effects will usually be understood as events in what follows.) Complicating matters, causal relations may be singular (*Socrates' drinking hemlock* caused *Socrates' death*) or general (*Drinking hemlock* causes *death*); hence the relata might be tokens (e.g., instances of properties) or types (e.g., types of events) of the category in question. Other questions up for grabs are: Are singular causes metaphysically and/or epistemologically prior to general causes or vice versa (or neither)? What grounds the intuitive asymmetry of the causal relation? Are macrocausal relations reducible to microcausal relations? And perhaps most importantly: Are causal facts (e.g., the holding of causal relations) reducible to noncausal facts (e.g., the holding of certain spatiotemporal relations)?

Some Issues in Philosophy of Causality

The Varieties of Causation

Causes can apparently contribute to effects in a variety of ways: by being background or standing conditions, "triggering events," omissions, factors that enhance or inhibit effects, factors that remove a common preventative of an effect, etc. Traditionally accounts of causation have focussed on triggering events, but contemporary accounts are increasingly expected to address a greater range of this diversity.

There may also be different notions of cause characteristic of the domains of different sciences (see Humphreys 1986; Suppes 1986): The seemingly indeterministic phenomena of quantum physics may require treatment different from either the seemingly deterministic processes of certain natural sciences or the "quasi-deterministic" processes characteristic of the social sciences, which are often presumed to be objectively deterministic but subjectively uncertain. Also relevant here is the distinction between teleological (intentional, goal-oriented) and nonteleological causality: While the broadly physical sciences tend not to cite motives and purposes, the plant, animal, human, and social sciences often explicitly do so. Contemporary treatments of teleological causality generally aim to avoid the positing of anything like entelechies or "vital forces" (of the sort associated with nineteenth-century accounts of biology), and also to avoid taking teleological goals to be causes that occur after their effects (see Salmon 1989, Sec. 3.8, for a discussion). On Wright's (1976) account, consequence etiology teleological behaviors (e.g., stalking a prey) are not caused by future catchings (which, after all, might not occur), but rather by the fact that the behavior in question has been often enough successful in the past that it has been evolutionarily selected for and for creatures capable of intentional representation, alternative explanations may be available. While teleological causes raise interesting questions for the causal underpinnings of behavior (especially concerning whether a naturalistically acceptable account of intentionality can be given), the focus in what follows will be on nonteleological causality, reflecting the primary concern of most contemporary philosophers of causation.

Singular vs. General Causation

Is all singular causation ultimately general? Different answers reflect different understandings of the notion of 'production' at issue in the platitude "Causes produce their effects." On generalist (or covering-law) accounts (see the section on "Hume and Pearson: Correlation, Not Causation" below), causal production is a matter of law: Roughly, event *c* causes event *e* just in case *c* and *e* are instances of terms in a law connecting events of *c*'s type with events of *e*'s type. The generalist interpretation is in part motivated by the need to ground inductive reasoning: Unless causal relations are subsumed by causal laws, one will be unjustified in inferring that events of *c*'s type will, in the future, cause events of *e*'s type. Another motivation stems from thinking that identifying a sequence of events as causal requires identifying the sequence as falling under a (possibly unknown) law.

Alternatively, singularists (see "Singularist Accounts" below) interpret causal production as involving a singular causal process (variously construed) that is metaphysically prior to laws. Singularists also argue for the epistemological priority of singular causes, maintaining that one can identify a sequence as causal without assuming that the sequence falls under a law, even when the sequence violates modal presuppositions (as in Fair's [1979] case: Intuitively, one could recognize a glass's breaking as causal, even if one antecedently thought glasses of that type were unbreakable).

Counterfactual accounts (see "Counterfactual Accounts" below) analyze singular causes in terms of counterfactual conditionals (as a first pass, event *c* causes event *e* just in case if *c* had not occurred, then *e* would not have occurred). Whether a counterfactual account should be considered singularist, however, depends on whether the truth of the counterfactuals is grounded in laws connecting types of events or in, for example, propensities (objective single-case chances) understood as irreducible to laws. Yet another approach to the issue of singular versus general causes is to deny that either is reducible to the other, and rather to give independent treatments of each type (as in Sober 1984).

Reduction vs. Nonreduction

There are at least three questions of reducibility at issue in philosophical accounts of causation, which largely cut across the generalist/singularist distinction. The first concerns whether causal facts (e.g., the holding of causal relations) are reducible to noncausal facts (e.g., the holding of certain spatiotemporal relations). Hume's generalist reduction of causality (see "Hume and Pearson: Correlation, Not Causation" below) has a projectivist or antirealist flavor: According to Hume, the seeming "necessary connexion" between cause and effect is a projection of a psychological habit of association between ideas, which habit is formed by regular experience of events of the cause type being spatially contiguous and temporally prior to events of the effect type. Contemporary neo-Humeans (see "Hempel: Explanation, Not Causation" and "Probabilistic Relevance Accounts" below) dispense with Hume's psychologism, focusing instead on the possibility of reducing causal relations and laws to objectively and noncausally characterized associations between events. (Whether such accounts are appropriately deemed antirealist is a matter of dispute, one philosopher's reductive elimination being another's reductive introduction.) By way of contrast, nonreductive generalists (often called realists—see "Causal Powers, Capacities, Universals, Forces" below) take the modally robust causal connection between event types to be an irreducible feature of reality (see Realism). Singularists also come in reductive or realist varieties (see "Singularist Accounts" below).

A second question of reducibility concerns whether a given account of causation aims to provide a conceptual analysis of the concept (hence to account for causation in bizarre worlds, containing magic, causal action at a distance, etc.) or instead to account for the causal relation in the actual world, in terms of physically or metaphysically more fundamental entities or processes. These different aims make a difference in what sort of cases and counterexamples philosophers of causation take to heart when developing or assessing theories. A common intermediate methodology focuses on central cases, leaving the verdict on far-fetched cases as "spoils for the victor."

A third question of reducibility concerns whether macro-causal relations (holding between entities, or expressed by laws, in the special sciences) are reducible to micro-causal relations (holding between entities, or expressed by laws, in fundamental physics). This question arises from a general desire to understand the ontological and causal underpinnings of the structured hierarchy of the sciences, and from a need to address, as a special case, the "problem of mental causation," of whether and how mental events (e.g., a feeling of pain) can be causally efficacious vis-à-vis certain effects (e.g., grimacing) that appear also to be caused by the brain events (and ultimately, fundamental physical events) upon which the mental events depend.

Causal reductionists (Davidson 1970; Kim 1984) suggest that mental events (more generally, macro-level events) are efficacious in virtue of supervening on (or being identical with) efficacious physical events. Many worry, however, that these approaches render macro-level events causally irrelevant (or "epiphenomenal"). Nonreductive approaches to macro-level causation come in both physicalist and nonphysicalist varieties (Wilson 1999 provides an overview; see Physicalism). Some physicalists posit a relation (e.g., the determinable/determinate relation or proper parthood) between macro- and micro-level events that entails that the set of causal powers of a given macro-level event *m* (roughly, the set of causal interactions that the event, in appropriate circumstances, could enter into) is a proper subset of those of the micro-level event *p* upon which *m* depends. On this "proper subset" strategy, the fact that the sets of causal powers are different provides some grounds for claiming that *m* is efficacious in its own right, but since each individual causal power of *m* is identical with a causal power of *p*, the two events are not in causal competition. On another nonreductive strategy—emergentism—the causal efficacy of at least some macro-level events (notably, mental events) is due to their having genuinely new causal powers not possessed by the physical events on which the mental events depend (see Emergence). When the effect in question is physical, such powers violate the causal closure of the physical (the claim that every physical effect has a fully sufficient physical cause); but such a violation arguably is not at odds with any cherished scientific principles, such as conservation laws (see McLaughlin 1992).

Features of Causality: Asymmetry, Temporal Direction, Transitivity

Intuitively, causality is asymmetric: if event *c* causes event *e*, then *e* does not cause *c*. Causality also generally proceeds from the past to the future. How to account for these data remains unclear. The problem of accounting for asymmetry is particularly pressing for accounts that reductively analyze causality in terms of associations, for it is easy to construct cases in which the associations are reversible, but the causation is not (as in Sylvain Bromberger's case in which the height *h* of a flagpole is correlated with the length *l* of the shadow it casts, and *vice versa*, and intuitively *h* causes *l*, but *l* does not cause *h*). Both the asymmetry and temporal direction of causality can be accommodated (as in Hume) by stipulatively identifying causal with temporal asymmetry: causes differ from their effects in being prior to their effects. This approach correctly rules out *l*'s causing *h* in the case above. But it also rules out simultaneous and backwards causation, which are generally taken to be live (or at least not too distant) possibilities.

Accounts on which the general temporal direction of causation is determined by physical or psychological processes may avoid the latter difficulties. On Reichenbach's (1956) account, the temporal direction of causation reflects the direction of "conjunctive forks": processes where a common cause produces joint effects, and where, in accordance with what Reichenbach called "the principle of the common cause", the probabilistic dependence of the effects on each other is "screened off"—goes away—when the common cause is taken into account. Such forks are, he claimed, always (or nearly always) open to the future and closed to the past. Some have suggested that the direction of causation is fixed by the direction of increasing entropy, or (more speculatively) by the direction of collapse of the quantum wave packet. Alternatively, Price (1991) suggests that human experience of manipulating causes provides a basis for the (projected) belief that causality is forward-directed in time (see "Counterfactuals and Manipulability" below). These accounts explain the usual temporal direction of causal processes, while allowing the occasional exception.

It remains the case, however, that accommodating the asymmetry of causation by appeal to the direction of causation rules out reducing the direction of time to the (general) direction of causation, which some (e.g., Reichenbach) have wanted to do. More importantly, neither stipulative nor nonstipulative appeals to temporal direction seem to adequately explain the asymmetry of causation, which intuitively has more to do with causes producing their effects (in some robust sense of 'production') than with causes being prior to their effects. Nonreductive accounts on which causality involves manifestations of powers or transfers of energy (or other conserved quantities) may be better situated to provide the required explanation, if such manifestations or transfers can be understood as directed (which remains controversial).

Another feature commonly associated with causality is transitivity: if *c* causes *d*, and *d* causes *e*, then *c* causes *e*. This assumption has come in for question of late, largely due to the following sort of case (see Kvart 1991): A man's finger is severed in a factory accident; a surgeon reattaches the finger, which afterwards becomes perfectly functional. The accident caused the surgery, and surgery caused the finger's functionality; but it seems odd to say that the accident caused the finger's functionality (see Hall 2000 for further discussion).

Challenges to Causality

Galileo, Newton, and Maxwell—How, Not Why

From the ancient through modern periods, accounts of natural phenomena proceeded by citing the powers and capacities of agents, bodies, and mechanisms to bring about effects (see Hankinson 1998; Clatterbaugh 1999). Galileo's account of the physics of falling bodies initiated a different approach to scientific understanding, on which this was a matter of determining how certain measurable quantities were functionally correlated (the "how" of things, or the kinematics), as opposed to determining the causal mechanisms responsible for these correlations (the "why" of things, or the dynamics). This descriptive approach enabled scientific theories to be formulated with comparatively high precision, which in turn facilitated predictive and retrodictive success; by way of contrast, explanations in terms of (often unobservable) causal mechanisms came to be seen as explanatorily otiose at best and unscientific at worst.

Newton's famous claim in the *Principia* ([1687] 1999) that "*hypotheses non fingo*" ("I frame no hypotheses"), regarding gravitation's "physical causes and seats" is often taken as evidence that he advocated a descriptivist approach (though he speculated at length on the causes of gravitational forces in the *Optics*). And while Maxwell drew heavily upon Faraday's qualitative account of causally efficacious electromagnetic fields (and associated lines of force) in the course of developing his theories of electricity and magnetism, he later saw such appeals to underlying causes as heuristic aids that could be dropped from the final quantitative theory.

Such descriptivist tendencies have been encouraged by perennial worries about the metaphysical and epistemological presuppositions of explicitly causal explanations (see "Causal Powers, Capacities, Universals, Forces" below) and the concomitant seeming availability of eliminativist or reductivist treatments of causal notions in scientific laws. For example, Russell (1912) influentially argued that since the equations of physics do not contain any terms explicitly referring to causes or causal relations and moreover (in conflict with the presumed asymmetry of causality) appear to be functionally symmetric (one can write $a = F/m$ as well as $F = ma$), causality should be eliminated as "a relic of a bygone age." Jammer (1957) endorsed a view in which force-based dynamics is a sophisticated form of kinematics, with force terms being mere "methodological intermediaries" enabling the convenient calculation of quantities (e.g., accelerations) entering into descriptions. And more recently, van Fraassen (1980) has suggested that while explanations going beyond descriptions may serve various pragmatic purposes, these have no ontological or causal weight beyond their ability to "save the phenomena."

Whether science really does, or should, focus on the (noncausally) descriptive is, however, deeply controversial. Galileo himself sought for explanatory principles going beyond description (see Jammer 1957 for a discussion), and, notwithstanding Newton's professed neutrality about their physical seats, he took forces to be the "causal principle[s] of motion and rest." More generally, notwithstanding the availability of interpretations of scientific theories as purely descriptive, there are compelling reasons (say, the need to avoid a suspect action at a distance) for taking the causally explanatory posits of scientific theories (e.g., fields and forces) ontologically seriously. The deeper questions here, of course, concern how to assess the ontological and causal commitments of scientific theories; and at present there is no philosophical consensus on these important matters. In any case it is not enough, in assessing whether causes are implicated by physical theories, to note that terms like 'cause' do not explicitly appear in the equations of the theory, insofar as the commitments of a given theory may transcend the referents of the terms appearing in the theory, and given that many terms—force, charge, valence—that do appear are most naturally defined in causal terms ('force,' for example, is usually defined as that which causes acceleration). It is also worth noting that the apparent symmetry of many equations, as well as the fact that cause terms do not explicitly appear in scientific equations, may be artifacts of scientists' using the identity symbol as an all-purpose connective between functional quantities, which enables the quantities to be manipulated using mathematical techniques but is nonetheless implicitly understood as causally directed, as in $F = ma$.

Nor does scientific practice offer decisive illumination of whether scientific theorizing is or is not committed to causal notions: As with Maxwell, it remains common for scientists to draw upon apparently robustly causal notions when formulating or explaining a theory, even while maintaining that the theory expresses nothing beyond descriptive functional correlations of measurable quantities. Perhaps it is better to attend to what scientists do rather than what they say. That they rarely leave matters at the level of descriptive laws linking observables is some indication that they are not concerned with just the "how" question—though, to be sure, the tension between descriptive and causal/explanatory questions may recur at levels below the surface of observation.

Hume and Pearson: Correlation, Not Causation

The Galilean view of scientific understanding as involving correlations among measurable quantities was philosophically mirrored in the empiricist view that all knowledge (and meaning) is ultimately grounded in sensory experience (see Empiricism). The greatest philosophical challenge to causality came from Hume, who argued that there is no experience of causes being efficacious, productive, or powerful vis-à-vis their effects; hence "we only learn by experience the frequent *conjunction* of objects, without being ever able to comprehend any thing like *connexion* between them" ([1748] 1993, 46). In place of realistically interpreted "producing" theories of causation (see Strawson 1987 for a taxonomy), Hume offered the first regularity theory of causation, according to which event *c* causes event *e* just in case *c* and *e* occur, and events of *c*'s type have (in one's experience) been universally followed by, and spatially contiguous to, events of *e*'s type. (As discussed, such constant conjunctions were the source of the psychological imprinting that was, for Hume, the ultimate locus of causal connection.) Hume's requirement of contiguity may be straightforwardly extended to allow for causes to produce distant effects, via chains of spatially contiguous causes and effects.

Hume's requirement of universal association is not sufficient for causation (night always follows day but day does not cause night). Nor is Hume's requirement necessary, for even putting aside the requirement that one experience the association in question, there are many causal events that happen only once (e.g., the big bang). The immediate move of neo-Humeans (e.g., Hempel and Oppenheim 1948; Mackie 1965) is to understand the generalist component of the account in terms of laws of nature, which express the lawful sufficiency of the cause for the effect, and where (reflecting Hume's reductive approach) the sufficiency is not to be understood as grounded in robust causal production. One wonders, though, what is grounding the laws in question if associations are neither necessary nor sufficient for their holding and robust production is not allowed to play a role (see Laws of Nature). If the laws are grounded in brute fact, it is not clear that the reductive aim has been served (but see the discussion of Lewis's account of laws, below).

In any case, neo-Humean accounts face several problems concerning events that are inappropriately deemed causes ("spurious causes"). One is the problem of joint effects, as when a virus causes first a fever, and then independently causes a rash: Here the fever is lawfully sufficient for, hence incorrectly deemed a cause of, the rash. Another involves violations of causal asymmetry: Where events of the cause's type are lawfully necessary for events of the effect's type, the effect will be lawfully sufficient for (hence inappropriately deemed a cause of) the cause. Cases of preemption also give rise to spurious causes: Suzy's and Billy's rockthrowings are each lawfully sufficient for breaking the bottle; but given that Suzy's rock broke the bottle (thereby preempting Billy's rock from doing so), how is one to rule out Billy's rockthrowing as a cause?

The above cases indicate that lawful sufficiency alone is not sufficient for causality. One response is to adopt an account of events in which these are finely individuated, so that, for example, the bottle-breaking resulting from Suzy's rock-throwing turns out to be of a different event type than a bottle-breaking resulting from Billy's rock-throwing (in which case Billy's rock-throwing does not instantiate a rock-throwing–bottle-breaking law, and so does not count as a cause). Another response incorporates a proviso that the lawful sufficiency at issue is sufficiency in the circumstances (as in Mackie's "INUS" condition account, in which a cause is an *i*nsufficient but *n*ecessary part of a condition that is, in the circumstances, *u*nnecessary but *s*ufficient for the effect).

Nor is lawful sufficiency alone necessary for causality, as the live possibility of irreducibly probabilistic causality indicates. This worry is usually sidestepped by a reconception of laws according to which these need express only some lawlike pattern of dependence; but this reconception makes it yet more difficult for regularity theorists to distinguish spurious from genuine causes and laws (see "Probabilistic Relevance Accounts" below for developments). One neo-Humean response is to allow certain a priori constraints to enter into determining what laws there are in a world (as in the "best system" theory of laws of Lewis 1994) in which the laws are those that systematize the phenomena with the best combination of predictive strength and formal simplicity, so as to accommodate probabilistic (and even uninstantiated) laws (see also "Hempel: Explanation, Not Causation" below).

The view that causation is nothing above (appropriately complex) correlations was widespread following the emergence of social statistics in the nineteenth century and was advanced by Karl Pearson, one of the founders of modern statistics, in 1890 in *The Grammar of Science*. Pearson's endorsement of this view was, like Hume's, inspired by a rejection of causality as involving mysterious productive powers, and contributed to causes (as opposed to associations) being to a large extent expunged from statistics and from the many sciences relying upon statistics. An intermediate position between these extremes, according to which causes are understood to go beyond correlations but are not given any particular metaphysical interpretation (in particular, as involving productive powers), was advanced by the evolutionary biologist Sewall Wright, the inventor of path analysis. Wright (1921) claimed that path analysis not only enabled previously known causal relations to be appropriately weighted, but moreover enabled the testing of causal hypotheses in cases where the causal relations were (as yet) unknown. Developed descendants and variants of Wright's approach have found increasing favor of late (see "Bayesian Networks and Causal Models" below), contributing to some rehabilitation of the notion of causality in the statistical sciences.

Hempel: Explanation, Not Causation

Logical empiricists were suspicious of causation understood as a metaphysical connection in nature; instead they located causality in language, interpreting causal talk as talk of explanation (see Salmon 1989 for a discussion). On Hempel and Oppenheim's (1948) influential D-N (deductive-nomological) model of scientific explanation, event *c* explains event *e* just in case a statement expressing the occurrence of *e* is the conclusion of an argument with premises, one of which expresses the holding of a universal generalization to the effect that events of *c*'s type are associated with events of *e*'s type and another of which expresses the fact that *c* occurred (see Explanation). Imposing certain requirements on universal generalizations (e.g., projectibility) enabled the D-N account to avoid cases of spurious causation due to accidental regularities (*s*'s being a screw in Smith's car does not explain why *s* is rusty, even if all the screws in Smith's car rusty). And requiring that explanations track temporal dependency relations (as a variation on identifying causal with temporal asymmetry) can prevent, for example, the length of the shadow from explaining the height of the flagpole. It is less clear, however, how to deal with explanatory preemption, as when Jones, immediately after ingesting a pound of arsenic, is run over by a bus and dies. Hempel's account allows us to explain Jones' death by citing the law that anyone who ingests a pound of arsenic dies within 24 hours, along with the fact that Jones ate a pound of arsenic; but such an explanatory arguments only cite laws and facts that are causally relevant to the event being explained; but this is in obvious tension with the empiricist's goal of characterizing causation in terms of explanation.

To accommodate the possibility of irreducibly probabilistic association, as well as explanations (characteristic of the social sciences) proceeding under conditions of partial uncertainty, Hempel (1965) proposed an inductive-statistical (I-S) model, in which event *c* explains event *e* if *c* occurs and it is an inductively grounded law that the probability of an event of type *e* given an event of type *c* is high (see Explanation; Inductive Logic). This account is subject of counterexamples in which a cause produces an effect but with a low probability, as in Scriven's case (discussed by him prior to Hempel's extension, and developed in Scriven 1975), where the probability of paresis given syphilis is low, but when paresis occurs, syphilis is the reason. A similar point applies to many quantum processes. Such cases gave rise to two different approaches to handling probabilistic explanation (or causation, by those inclined to accept this notion). One approach (see Railton 1978) locates probabilistic causality in propensities; the other (see "Probabilistic Relevance Accounts" below) in more sophisticated probabilistic relations.

At this point the line between accounts that are reductive (in the sense of reducing causal to noncausal goings-on) and nonreductive, as well as the line between singularist and generalist accounts, begins to blur. For while a propensity-based account of causality initially looks nonreductive and singularist, some think that propensities can be accommodated on a sophisticated associationist account of laws; and while an account based on relations of probabilistic relevance initially looks reductive and generalist, whether it is so depends on how the probabilities are interpreted (as given by frequencies, irreducible propensities, etc.).

Contemporary Generalist Accounts

Probabilistic Relevance Accounts

A natural response to Scriven-type cases is to understand positive causal relevance in terms of probability raising (Suppes 1970): Event *c* causes event *e* just in case the probability of events of *e*'s type is higher given events of *c*'s type than without. (Other relevance relations, such as being a negative causal factor, can be defined accordingly.) A common objection to such accounts (see Rosen 1978) proceeds by constructing "doing the hard way" cases in which it seems that causes lower the probability of their effects (e.g., where a mishit golf ball ricochets off a tree, resulting in a hole-in-one; or where a box contains a radioactive substance *s* that produces decay particles but the presence of *s* excludes the more effective radioactive substance *s*′). Such cases can often be handled, however, by locating a neutral context (where the golf ball is not hit at all, or where no radioactive substance is in the box) relative to which events of the given type do raise the probability of events of the effect type.

A more serious problem for probability-raising accounts is indicated by Simpson's paradox, according to which any statistical relationship between two variables may be reversed by including additional factors in the analysis. The "paradox" reflects the possibility that a variable C can be positively correlated with a variable E in a population and yet C be negatively correlated with E in every partition of the population induced by a third variable X. Just this occurred in the Berkeley sex discrimination case. Relative to the population of all men and women applying to graduate school at the University of California, Berkeley, being male (C) was positively correlated with being admitted (E). But relative to every partition of the population containing the men and women applying to a particular department (X), this correlation was reversed. In this case the difference between the general and specific population statistics reflected the fact that while in every department it was easier for women to be admitted than men, women were more likely to apply to departments that were harder (for *everyone*) to get into. The general population statistic was thus confounded: Being a male, *simpliciter*, was not in fact causally relevant to getting into graduate school at UC Berkeley; rather (assuming no other confounding was at issue), applying to certain departments rather than others was what was relevant.

To use probabilistic accounts as a basis for testing hypotheses and making predictions—and especially in order to identify effective strategies (courses of action) in the social sciences and, indeed, in everyday life—statistical confounding needs to be avoided. Nancy Cartwright (1979) suggested that avoiding confounding requires that the relevant probabilities be assessed relative to background contexts within which all other causal factors (besides the variable C, whose causal relevance is at issue) are held fixed. Opinions differ regarding whether events of type C must raise the probability of events of type E in at least one such context, in a majority of contexts, or in every such context. So one might take C to be a positive causal factor for E just in case $P(E \mid C \wedge X_i) \geq P(E \mid \neg C \wedge X_i)$ for all background contexts X_i, with strict inequality for at least one X_i. On this approach, smoking would be a positive causal factor for having lung cancer just in case smoking increases the chance of lung cancer in at least one background context and does not lower it in any background context.

Practically, Cartwright's suggestion has the disadvantage that one is frequently not in a position to control for all alternative causal factors (though in some circumstances one can avoid having to do this; see "Bayesian Networks and Causal Models" below). Philosophically, the requirement threatens reductive versions of probabilistic accounts with circularity. Attempts have been made to provide a noncircular means of specifying the relevant background contexts (e.g., Salmon 1984), but it is questionable whether these attempts succeed, and many are presently prepared to agree with Cartwright: "No causes in, no causes out."

As mentioned, probabilistic accounts may or may not be reductive, depending on whether the probabilities at issue are understood as grounded in associations (as in Suppes 1970) or else in powers, capacities, or propensities (as in Humphreys 1989 and Cartwright 1989). In the latter interpretation, further divisions are introduced: If the propensities are taken to be irreducible to laws (as in Cartwright's account), then the associated probabilistic relevance account is more appropriately deemed singularist. Complicating the taxonomy here is the fact that most proponents of probabilistic accounts are not explicit as regards what analysis should be given of the probabilities at issue.

Bayesian Networks and Causal Models

Philosophical worries concerning whether statistical information adequately tracks causal influence are echoed in current debates over the interpretation of the statistical techniques used in the social sciences. As noted above ("Hume and Pearson: Correlation, Not Causation"), these techniques have frequently been interpreted as relating exclusively to correlations, but recently researchers in computer science, artificial intelligence, and statistics have developed interpretations of these approaches as encoding explicitly causal information (see Spirtes, Glymour, and Scheines 1993; Pearl 2000).

In the causal modeling approach, one starts with a set of variables (representing properties) and a probability distribution over the variables. This probability distribution, partially interpreted with prior causal knowledge, is assumed to reflect a causal structure (a set laws expressing the causal relations between the variables, which laws may be expressed either graphically or as a set of structurally related functional equations). Given that certain conditions (to be discussed shortly) hold between the probabilities and the causal structure, algorithmic techniques are used to generate the set of all causal structures consistent with the probabilities and the prior causal knowledge. Techniques also exist for extracting information regarding the results of interventions (corresponding to manipulations of variables). Such strategies appear to lead to improved hypothesis testing and prediction of effects under observation and intervention.

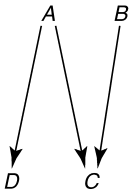


Fig. 1. *A* is a common causal factor of *D* and *C*.

Another advantage claimed for such accounts is that they provide a means of avoiding confounding without imposing the often impracticable requirement that the relevant probabilities be assessed against background contexts taking into account all causal factors, it rather being sufficient to take into account all common causal factors. As a simple illustration, suppose that *A* and *B* are known to causally influence *C*, as in Figure 1.

To judge whether *D* influence *C*, Cartwright (1979) generally recommends holding fixed both *A* and *B*, while Spirtes *et al.* (1993) instead recommend holding fixed only *A*. Cartwright allows, however, that attention to just common causal factors is possible when the causal Markov condition holds. Since this is one of the conditions that must hold in order to implement the causal modeling approach, the restriction to common causal factors is not really an advantage over Cartwright's account.

While causal modeling approaches may lead to improved causal inference concerning complex systems (in which case they are of some epistemological interest), it is unclear what bearing they have on the metaphysics of causality. Spirtes *et al.* (1993) present their account not so much as an analysis of causality as a guide to causal inference. Pearl (2000), however, takes the appeal to prior causal intuitions to indicate that causal modeling approaches are nonreductive (and moreover based on facts about humans' cognitive capacities to make effective causal inferences in simple cases). In any case, the potential of causal models to provide a basis for a general theory of causality is limited by the need for certain strong conditions to be in place in order for the algorithms to be correctly applied.

One of these is the aforementioned causal Markov condition (of which Reichenbach's [1956] "principle of the common cause" was a special case), which says that once one conditions on the complete set P_v of "causal parents" (direct causes) of a variable *V, V* will be probabilistically independent of all other variables except *V*'s descendants; that is for all variables *X*, where *X* is not one of *V*'s descendants, $P(V|P_v \wedge X) = P(V|P_v)$. (In particular, where *V* is a joint effect, conditioning on the causal parents of *V* screens off the probabilistic influence of the other joint effects on *V*.) Here again there is the practical problem that in the social sciences, where the approaches are supposed to be applicable, one is often not in a position to specify states with sufficient precision to guarantee that the condition is met. A metaphysical problem is that (contrary to Reichenbach's apparent assumption that the condition holds in all cases involving a common cause of joint effects) the causal Markov condition need not hold in cases of probabilistic causation: When a particle may probabilistically decay either by emitting a high-energy electron and falling into a low-energy state or by emitting a low-energy electron and falling into a different energy state, the joint effects in either case will not be probabilistically independent of each other, even conditioning on the cause; and certain cases of macrocausation appear also to violate the condition.

A second assumption of the causal modeling technique is what Spirtes *et al.* (1993) call "faithfulness" (also known as "stability" in Pearl 2000), according to which probabilistic dependencies faithfully reveal causal connections. In particular, if *Y* is probabilistically independent of *X*, given *X*'s parents, then *X* is assumed not to cause *Y*. Again, this condition cannot be assumed to hold in all cases, since some variables (properties) may sometimes prevent and sometimes produce and effect (as when birth control pills are a cause of thrombosis yet also prevent thrombosis, insofar as pregnancy causes thrombosis and the pills prevent pregnancy). In circumstances where the positive and negative contributions of *X* to *Y* are equally effective, the probabilistic dependence of effect on cause may cancel out, and thus *X* may inappropriately be taken not to be causally relevant to *Y*.

Causal Powers, Capacities, Universals, Forces

As mentioned, some proponents of probabilistic relevance accounts endorse metaphysical interpretations of the probabilities at issue. Such positions fall under the broader category of nonreductive ("realist") covering-law theories, in which laws express (or are grounded in) more than mere associations. The job such accounts face is to provide an alternative basis for causal laws. Among other possibilities, these bases are taken to be relations of necessitation or "probabilification" among universals (Dretske 1977; Tooley 1977; Armstrong 1978), (law-based) capacities or powers associated with objects or properties (Shoemaker 1980; Martin

1993), or fundamental forces or interactions (Bohm 1957; Strawson 1987).

Such accounts sidestep many of the problems associated with reductive covering law accounts. Since laws are not just a matter of association, a realist has the means to deny, in the virus–fever–rash case, that there is a law connecting fevers with rashes; similarly, in cases of preemption a realist may claim that, for example, Billy's rock throwing and the bottle's breaking did not instance the law in question (even without endorsing a fine-grained account of event individuation). Of course, much depends here on the details of the proposed account of laws. In the Dretske-Tooley-Armstrong account, causal laws are contingent, brute relations between universals. Some find this less than satisfying from a realist point of view, insofar as it is compatible with, for example, the property of having spin 1 bestowing completely different causal powers (say, all those actually bestowed by having spin $\frac{1}{2}$) on its possessing particulars. Other realists are more inclined to see the nature of properties and particulars as essentially dependent on the causal laws that actually govern them, a view that is not implausible for scientific entities: "[C]ausal laws are not like externally imposed legal restrictions that, so to speak, merely limit the course of events to certain prescribed paths .... [T]he causal laws satisfied by a thing ... are inextricably bound up with the basic properties of the thing which helps to define what it is" (Bohm 1957, 14).

The primary problem facing realist accounts is that they require accepting entities and relations (universals, causal powers, forces) that many philosophers and scientists find metaphysically obscure and/or epistemologically inaccessible. How one evaluates these assessments often depends on one's other commitments. For example, many traditional arguments against realist accounts (e.g., Hume's arguments) are aimed at showing that these do not satisfy a strict epistemological standard, according to which the warranted posit of a contingent entity requires that the entity be directly accessible to experience (or a construction from entities that are so accessible). But if inference to the existence of an unexperienced entity (as the best explanation of some phenomenon) is at least sometimes an acceptable mode of inference, such a strict epistemological standard (and associated arguments) will be rejected; and indeed, positive arguments for contemporary realist accounts of causality generally proceed via such inferences to the best explanation—often of the patterns of association appealed to by reductivist accounts.

Contemporary Singularist Accounts

Singularists reject the claim that causes follow laws in the order of explanation, but beyond this there is considerable variety in their accounts. *Contra* Hume, Anscombe (1971) takes causation to be a (primitive) relation that may be observed in cuttings, pushings, fallings, etc. It is worth noting that a primitivist approach to causality is compatible with even a strict empiricism (compare Hume's primitivist account of the resemblance relation). The empiricist Ducasse (1926) also locates causation in singular observation, but nonprimitively: A cause is the change event observed to be immediately prior and spatiotemporally contiguous to an effect event. While interesting in allowing for a non-associative, non-primitivist, empiricist causality, Ducasse's account is unsatisfactory in allowing only the coarse-grained identification of causes (as some backward temporal segment of the entire observed change); hence it fails to account for most ordinary causal judgments. Note that singularists basing causation on observation need not assert that one's knowledge of causality proceeds only via observations of the preferred sort; they rather generally maintain that such experiences are sufficient to account for one's acquiring the concept of causation, then both allow that causation need not be observed and that confirming singular causal claims may require attention to associations.

Another singularist approach takes causation to be theoretically inferred, as that relation satisfying (something like) the Ramsey sentence consisting of the platitudes about causality involving asymmetry, transitivity, and so on (see Tooley 1987). One problem here is that, as may be clear by now, such platitudes do not seem to uniformly apply to all cases. Relatedly, one may wonder whether the platitudes are consistent; given the competing causal intuitions driving various accounts of causality, it would be surprising if they were.

Finally, a wide variety of singularist accounts analyze causality in terms of singular processes. Such accounts are strongly motivated by the intuition that in a case of preemption such as that of Suzy and Billy, what distinguishes Suzy's throw as a cause is that it initiates a process ending in the bottle breaking, while the process initiated by Billy's throw never reaches completion (see Menzies 1996 for a discussion). Commonly, process for singularists attempt (like Ducasse) to provide a non-primitivist causality that is both broadly empiricist, in not appealing to any properly metaphysical elements, and non-associationist, in recognition of the diffculties that associationist accounts have (both with preemption and with distinguishing genuine causality from accidental regularity). Hence they typically fill in the "process" intuition by identifying causality with fundamental physical processes, including transfers or interactions, as in Fair's (1979) account of causation as identical with the transfer of energy momentum, Salmon's (1984) "mark transmission" account, and Dowe's (1992) account in which the transfer of any conserved quantity will suffice.

An objection to the claim that physical processes are sufficient for causality is illustrated by Cartwright's (1979) case of a plant sprayed with herbicide that improbably survives and goes on to flourish (compare also Kvart's 1991 finger-severing case, discussed previously). While transfers and interactions of the requisite sort can be traced from spraying to flourishing, intuitively the former did not cause the latter; however, accepting that the spraying did cause the flourishing may not be an overly high price to pay. A deeper worry concerns the epistemological question of how accounts of physical process link causation, understood as involving theoretical relations or processes of fundamental physics, with causation as ordinarily experienced. Fair suggests that ordinary experience involves macroprocesses, which are in turn reducible to the relevant physical processes; but even supposing that such reductions are in place, ordinary causal judgments do not seem to presuppose them.

Contemporary Counterfactual Accounts

Counterfactual accounts of causality, which may also be traced back to Hume, take as their starting point the intuition that a singular cause makes an important difference in what happens. As a first pass, *c* causes *e* (where *c* and *e* are actually occurring events) only if, were *c* not to occur, then *e* would not occur. As a second pass, *c* causes *e* only if *c* and *e* are connected by a chain of such dependencies (see Lewis 1973), so as to ensure that causation is transitive (causation is thus the "ancestral," or transitive closure, of counterfactual dependence). In addition to the requirement of counterfactual necessity of causes for effects, counterfactual accounts also commonly impose a requirement of counterfactual sufficiency of causes for effects: If *c* were to occur, then *e* would occur. Insofar as counterfactual accounts are standardly aimed at reducing causal to noncausal relations, and given plausible assumptions concerning evaluation of counterfactuals, the latter requirement is satisfied just by *c* and *e*'s actually occurring (which occurrences, as above, are assumed); hence standard counterfactual accounts do not have a nontrivial notion of counterfactual sufficiency. A nontrivial notion of counterfactual sufficiency can be obtained by appeal to nested counterfactuals (see Vihvelin 1995): *c* causes *e* only if, if neither *c* nor *e* had occurred, then if *c* had occurred, *e* would have occurred.

Problems, Events, and Backtrackers

While counterfactual accounts are often motivated by a desire to give a reductive account of causality that avoids problems with reductive covering-law accounts (especially those of joint effects and of preemption), it is unclear whether counterfactual accounts do any better by these problems. First, consider the problem of joint effects. Suppose a virus causes first a fever, then a rash, and that the fever and rash could only have been caused by the virus. It seems correct to reason in the following "backtracking" fashion: If the fever had not occurred, then the viral infection would not have occurred, in which case the rash would not have occurred. But then the counterfactual "If the fever had not occurred, the rash would not have occurred" turns out true, which here means that the fever causes the rash, which is incorrect. Proponents of counterfactual accounts have responses to these objections, which require accepting controversial accounts of the truth conditions for counterfactuals (see Lewis 1979). Even so, the responses appear not to succeed (see Bennett 1984 for a discussion).

Second, consider the problem of preemption. In the Suzy-Billy case, it seems correct to reason that if Suzy had not thrown her rock, then Billy's rock would have gotten through and broken the bottle. Hence the counterfactual "If Suzy's throw had not occurred, the bottlebreaking would not have occurred" turns out false; so Suzy's rockthrowing turns out not to be a cause, which is incorrect. In cases (as here) of so-called "early preemption," where it makes sense to suppose that there was an intermediate event *d* between the effect and the cause on which the effect depended, this result can be avoided: Although the breaking does not counterfactually depend on Suzy's rockthrowing, there is a chain of counterfactual dependence linking the breaking to Suzy's rockthrowing (and no such chain linking the breaking to Billy's), and so her throw does end up being a cause (and Billy's does not). But the appeal to an intermediate event seems *ad hoc*, and in any case cannot resolve cases of "late preemption." Lewis (developing an idea broached in Paul 1998) eventually responded to such cases by allowing that an event may be counted as a cause if it counterfactually influences the mode of occurrence of the effect (e.g., how or when it occurs), as well as if it counterfactually influences the occurrence of the effect, *simpliciter*.

Counterfactuals and Manipulability

Where counterfactual accounts may be most useful is in providing a basis for understanding or formalizing the role that manipulability plays in the concept of causation. One such approach sees counterfactuals as providing the basis for an epistemological, rather than a metaphysical, account of causation (see Pearl 2000 for discussion). The idea here is that counterfactuals nicely model the role manipulability (actual or imagined) plays in causal inference, for a natural way to determine whether a counterfactual is true is to manipulate conditions so as to actualize the antecedent. Another approach takes counterfactuals to provide a basis for a generalist account of causal explanation (see Woodward 1997), according to which such explanations track stable or invariant connections and the notion of invariance is understood nonepistemologically in terms of a connection's continuing to hold through certain counterfactual (not necessarily human) "interventions." Whether the notion of manipulability is itself a causal notion, and so bars the reduction of causal to noncausal facts, is still an open question.

JESSICA WILSON

References

Anscombe, G. E. M. (1971), *Causality and Determination*. Cambridge: Cambridge University Press.

Armstrong, David M. (1978), *Universals and Scientific Realism*, vol. 2: *A Theory of Universals*. Cambridge: Cambridge University Press.

Bennett, Jonathan (1984), "Counterfactuals and Temporal Direction," *Philosophical Review* 93: 57–91.

Bohm, David (1957), *Causality and Chance in Modern Physics*. London: Kegan Paul.

Cartwright, Nancy (1979), "Causal Laws and Effective Strategies," *Noûs* 13: 419–438.

——— (1989), *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.

Clatterbaugh, Kenneth (1999), *The Causation Debate in Modern Philosophy, 1637–1739*. New York: Routledge.

Davidson, Donald (1967), "Causal Relations," *Journal of Philosophy* 64: 691–703. Reprinted in Davidson (2001), *Essays on Actions and Events*. Oxford: Oxford University Press, 149–162.

——— (1970), "Mental Events," in L. Foster and J. Swanson (eds.), *Experience and Theory*. Amherst: Massachusetts University Press. Reprinted in Davidson (2001), *Essays on Actions and Events*. Oxford: Oxford University Press, 207–224.

Dowe, Phil (1992), "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory," *Philosophy of Science* 59: 195–216.

Dretske, Fred (1977), "Laws of Nature," *Philosophy of Science* 44: 248–268.

Ducasse, C. J. (1926), "On the Nature and Observability of the Causal Relation," *Journal of Philosophy* 23: 57–68.

Fair, David (1979), "Causation and the Flow of Energy," *Erkenntnis* 14: 219–250.

Hall, Ned (2000), "Causation and the Price of Transitivity," *Journal of Philosophy* 97: 198–222.

Hankinson, R. J. (1998), *Cause and Explanation in Ancient Greek Thought*. Oxford: Oxford University Press.

Hempel, Carl (1965), *In Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hempel, Carl, and Paul Oppenheim (1948), "Studies in the Logic of Explanation," *Philosophy of Science* 15: 135–175.

Hitchcock, Christopher (1993), "A Generalized Probabilistic Theory of Causal Relevance," *Synthese* 97: 335–364.

Hume, David ([1739] 1978), *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge. Oxford: Oxford University Press.

——— ([1748] 1993), *An Enquiry Concerning Human Understanding*. Edited by Eric Steinberg. Indianapolis: Hackett Publishing.

Humphreys, Paul (1986), "Causation in the Social Sciences: An Overview," *Synthese* 68: 1–12.

——— (1989), *The Chances of Explanation*. Princeton, NJ: Princeton University Press.

Jammer, Max (1957), *Concepts of Force*. Cambridge, MA: Harvard University Press.

Kim, Jaegwon (1984), "Epiphenomenal and Supervenient Causation," *Midwest Studies in Philosophy IX: Causation and Causal Theories*, 257–270.

——— (1993), *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

Kvart, Igal (1991), "Transitivity and Preemption of Causal Relevance," *Philosophical Studies* LXIV: 125–160.

Lewis, David (1973), "Causation," *Journal of Philosophy* 70: 556–567.

——— (1979), "Counterfactual Dependence and Time's Arrow," *Noûs* 13:455–476.

——— (1983), *Philosophical Papers*, vol. 1. Oxford: Oxford University Press.

——— (1994), "Humean Supervenience Debugged," *Mind* 1: 473–490.

Mackie, John L. (1965), "Causes and Conditions," *American Philosophical Quarterly* 2: 245–264.

Martin, C. B. (1993), "Power for Realists," in Keith Cambell, John Bacon, and Lloyd Reinhardt (eds.), *Ontology, Causality, and Mind: Essays on the Philosophy of D. M. Armstrong*. Cambridge: Cambridge University Press.

McLaughlin, Brian (1992), "The Rise and Fall of British Emergentism," in Ansgar Beckerman, Hans Flohr, and Jaegwon Kim (eds.), *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. Berlin: De Gruyter, 49–93.

Mellor, D. H. (1995), *The Facts of Causation*. Oxford: Oxford University Press.

Menzies, Peter (1996), "Probabilistic Causation and the Pre-Emption Problem," *Mind* 105: 85–117.
Newton, Isaac ([1687] 1999), *The Principia: Mathematical Principles of Natural Philosophy*. Translated by I. Bernard Cohen and Anne Whitman. Berkeley and Los Angeles: University of California Press.
Paul, Laurie (1998), "Keeping Track of the Time: Emending the Counterfactual Analysis of Causation," *Analysis* LVIII: 191–198.
Pearl, Judea (2000), *Causality*. Cambridge: Cambridge University Press.
Price, Huw (1992), "Agency and Causal Asymmetry," *Mind* 101: 501–520.
Railton, Peter (1978), "A Deductive-Nomological Model of Probabilistic Explanation," *Philosophy of Science* 45: 206–226.
Reichenbach, Hans (1956), *The Direction of Time*. Berkeley and Los Angeles: University of California Press.
Rosen, Deborah (1978), "In Defense of a Probabilistic Theory of Causality," *Philosophy of Science* 45: 604–613.
Russell, Bertrand (1912), "On the Notion of Cause," *Proceedings of the Aristotelian Society* 13: 1–26.
Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
——— (1989), "Four Decades of Scientific Explanation," in Philip Kitcher and Wesley Salmon (eds.), *Scientific Explanation*, 3–219.
Scriven, Michael (1975), "Causation as Explanation," *Noûs* 9: 238–264.
Shoemaker, Sydney (1980), "Causality and Properties," in Peter van Inwagen (ed.), *Time and Cause*. Dordrecht, Netherlands: D. Reidel, 109–135.
Sober, Elliott (1984), "Two Concepts of Cause," in Peter D. Asquith and Philip Kitcher (eds.), *PSA 1984*. East Lansing, MI: Philosophy of Science Association, 405–424.
Sosa, Ernest, and Michael Tooley (eds.) (1993), *Causation*. Oxford: Oxford University Press..
Spirtes, P., C. Glymour, and R. Scheines (1993), *Causation, Prediction, and Search*. New York: Springer-Verlag.
Strawson, Galen (1987), "Realism and Causation," *Philosophical Quarterly* 37: 253–277.
Suppes, Patrick (1970), *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.
——— (1986), "Non-Markovian Causality in the Social Sciences with Some Theorems on Transitivity," *Synthese* 68: 129–140.
Tooley, Michael (1977), "The Nature of Laws," *Canadian Journal of Philosophy* 7: 667–698.
——— (1987), *Causation: A Realist Approach*. Oxford: Oxford University Press.
van Fraassen, Bas (1980), *The Scientific Image*. Oxford: Oxford University Press.
Vihvelin, Kadhri (1995), "Causes, Effects, and Counterfactual Dependence," *Australasian Journal of Philosophy* 73: 560–583.
Wilson, Jessica (1999), "How Superduper Does a Physicalist Supervenience Need to Be?" *Philosophical Quarterly* 49: 33–52.
Woodward, James (1997), "Explanation, Invariance, and Intervention," *Philosophy of Science* 64 (Proceedings): S26–S41.
Wright, Larry (1976), *Teleological Explanation*. Berkeley and Los Angeles: University of California Press.
Wright, Sewall (1921), "Correlation and Causation," *Journal of Agricultural Research* 20: 557–585.

CHAOS THEORY

See **Prediction**

PHILOSOPHY OF CHEMISTRY

Although many influential late-nineteenth- and early-twentieth-century philosophers of science were educated wholly or in part as chemists (Gaston Bachelard, Pierre Duhem, Emile Myerson, Wilhelm Ostwald, Michael Polanyi), they seldom reflected directly on the epistemological, methodological, or metaphysical commitments of their science. Subsequent philosophers of science followed suit, directing little attention to chemistry in comparison with physics and biology despite the industrial, economic, and academic success of the chemical sciences. Scattered examples of philosophical reflection on chemistry by chemists do exist; however, philosophically sensitive historical analysis and sustained conceptual analysis are relatively recent phenomena. Taken together, these two developments demonstrate that chemistry addresses general issues in the philosophy of science and, in addition, raises important questions in the interpretation of chemical theories, concepts, and experiments.

Historians of chemistry have also raised a number of general philosophical questions about chemistry. These include issues of explanation, ontology, reduction, and the relative roles of theories, experiments, and instruments in the advancement of the science.

Worries about the explanatory nature of the alchemical, corpuscularian, and phlogiston theories are well documented (cf. Bensaude-Vincent and Stengers 1996; Brock 1993). Lavoisier and—to a lesser extent historically—Dalton initiated shifts in the explanatory tasks and presuppositions of the science. For instance, prior to Lavoisier many chemists "explained" a chemical by assigning it to a type associated with its experimental dispositions (e.g., flammability, acidity, etc.). After Lavoisier and Dalton, "explanation" most often meant the isolation and identification of a chemical's constituents. Eventually, the goal of explanation changed to the identification of the transformation processes in the reactants that gave rise to the observed properties of the intermediaries and the products. It is at that time that chemists began to write the now familiar reaction equations, which encapsulate this change. These transformations cannot be described simply as the coming of new theories; they also involved changes in explanatory presuppositions and languages, as well as new experimental techniques (Bensaude-Vincent and Stengers 1996; Nye 1993).

Thinking in terms of transformation processes led chemists to postulate atoms as the agents of the transformation process. But atoms were not observable in the nineteenth century. How is the explanatory power of these unobservable entities accounted for? Also, do chemical elements retain their identity in compounds (Paneth 1962)? Something remains the same, yet the properties that identify elements (e.g., the green color of chlorine gas) do not exist in compounds (e.g., sodium chloride or common table salt).

The history of chemistry also raises a number of interesting questions about the character of knowledge and understanding in the science. Whereas philosophers have historically identified theoretical knowledge with laws or sets of propositions, the history of chemistry shows that there are different kinds of knowledge that function as a base for understanding. Much of chemistry is experimental, and much of what is known to be true arises in experimental practice independently of or only indirectly informed by theoretical knowledge. When chemists have theorized, they have done so freely, using and combining phenomenological, constructive (in which the values of certain variables are given by experiment or other theory), and deductive methods. Chemists have rarely been able to achieve anything like a strict set of axioms or first principles that order the phenomena and serve as their explanatory base (Bensaude-Vincent and Stengers 1996; Gavroglu 1997; Nye 1993).

Historical research has also raised the issue of whether theory has contributed most to the progress of chemistry. While philosophers often point to the conceptual "revolution" wrought by Lavoisier as an example of progress, historians more often point to the ways in which laboratory techniques have been an important motor of change in chemistry, by themselves or in tandem with theoretical shifts (Bensaude-Vincent and Stengers 1996; Nye 1993). For example, during the nineteenth century, substitution studies, in which one element is replaced by another in a compound, were driven largely by experimental practices. Theoretical concepts did not, in the first instance, organize the investigation (Klein 1999). Similar remarks can be made regarding the coming of modern experimental techniques such as various types of chromatography and spectroscopy (Baird 2000; Slater 2002). Chemists often characterize a molecule using such techniques, and while the techniques are grounded in physical theory, the results must often be interpreted in chemical language. These examples lead back to questions about the nature of chemical knowledge. Arguably, the knowledge appears to be a mix of "knowing how" and "knowing that" which is not based solely in the theory (chemical or physical) available at the time.

A number of the philosophical themes raised in the history of chemistry continue to reverberate in current chemistry. For instance, what is the proper ontological base for chemical theory, explanation, and practice? While many chemists would unfailingly resort to molecular structure as the explanation of what is seen while a reaction is taking place, this conception can be challenged from two directions. From one side, echoing the eighteenth-century conception of the science, chemistry begins in the first instance with conceptions and analyses of the qualitative properties of material stuff (Schummer 1996; van Brakel 2001). One might call the ontology associated with this conception a metaphysically nonreductive dispositional realism, since it focuses on properties and how they appear under certain conditions and does not attempt to interpret them in any simpler terms. In this conception, reference to the underlying molecular structure is subsidiary or even otiose, since the focus is on the observable properties of the materials. Given that molecular structure is difficult to justify within quantum mechanics (see below), one can argue that it is justifiable to remain with the observable properties. If this view is adopted, however, the justification of the ontology of material stuff becomes a pressing matter. Quantum mechanics will not supply the justification, since it does not deliver the qualitative properties of materials. Further, it still seems necessary to account for the phenomenal success that molecular explanations afford in planning and interpreting chemical structures and reactions.

From the other side, pure quantum mechanics makes it difficult to speak of the traditional atoms-within-a-molecule approach referred to in the reaction equations and structural diagrams (Primas 1983; Weininger 1984). Quantum mechanics tells us that the interior parts of molecules should not be distinguishable; there exists only a distribution of nuclear and electronic charges. Yet chemists rely on the existence and persistence of atoms and molecules in a number of ways. To justify these practices, some have argued that one can forgo the notion of an atom based on the orbital model and instead identify spatial regions within a molecule bounded by surfaces that have a zero flux of energy across the surfaces (Bader 1990). Currently, it is an open question whether this representation falls naturally out of quantum mechanics, and so allows one to recover atoms as naturally occurring substituents of molecules, or whether the notion of 'atom' must be presupposed in order for the identification to be made. Here again there is a question of the character of the theory that will give the desired explanation.

A number of examples supporting the claim that quantum mechanics and chemistry are uneasy bedfellows will be discussed below as they relate to the issue of reductionism, but their relationship also raises forcefully the long-standing issue of how theory guides chemical practice. Even in this era of supercomputers, only the energy states of systems with relatively few electrons or with high degrees of symmetry can be calculated with a high degree of faithfulness to the complete theoretical description. For most chemical systems, various semi-empirical methods must be used to get theoretically guided results. More often than not, it is the experimental practice independent of any theoretical calculation that gets the result. Strictly theoretical predictions of novel properties are rather rare, and whether they are strictly theoretical is a matter that can be disputed. A case in point is the structure of the CH_2 molecule. Theoretical chemists claimed to have predicted novel properties of the molecule, *viz.*, its nonlinear geometry, prior to any spectroscopic evidence (Foster and Boys 1960). While it is true that the spectroscopic evidence was not yet available, it may have been the case that reference to analogous molecules allowed the researchers to set the values for some of the parameters in the equations. So the derivation may not have been as a priori as it seemed.

Like the other special sciences, chemistry raises the issue of reductionism quite forcefully. However, perhaps because most philosophers have accepted at face value Dirac's famous dictum that chemistry has become nothing more than the application of quantum mechanics to chemical problems (cf. Nye 1993, 248), few seem to be aware of the difficulties of making good on that claim using the tools and concepts available in traditional philosophical analyses of the sciences. No one doubts that chemical forces are physical in nature, but connecting the chemical and physical mathematical structures and/or concepts proves to be quite a challenge. Although problems involving the relation between the physical and the chemical surround a wide variety of chemical concepts, such as aromaticity, acidity (and basicity), functional groups, and substituent effects (Hoffmann 1995), three examples will be discussed here to illustrate the difficulties: the periodic table, the use of orbitals to explain bonding, and the concept of molecular shape. Each also raises issues of explanation, representation, and realism.

Philosophers and scientists commonly believe that the periodic table has been explained by—and thus reduced to—quantum mechanics. This is taken to be an explanation of the configuration of the electrons in the atom, and, as a result of this, an explanation of the periodicity of the table. In the first case, however, configurations of electrons in atoms and molecules are the result of a particular approximation, in which the many-electron quantum wavefunction is rewritten as a series of one-electron functions. In practice, these one-electron functions are derived from the hydrogen wavefunction and, upon integration, lead to the familiar spherical *s* orbital, the dumbbell-shaped *p* orbitals, and the more complicated *d* and *f* orbitals. Via the Pauli exclusion principle, which states that the spins are to be paired if two electrons are to occupy one orbital and no more than two electrons may occupy an orbital, electrons are assigned to these orbitals. If the approximation is not made (and quantum mechanics tells us it should not be, since the approximation relies on the distinguishability of electrons), the notion of individual quantum numbers—and thus configurations—is no longer meaningful. In addition, configurations themselves are not observable; absorption and emission spectra are observed and interpreted as energy transitions between orbitals of different energies.

In the second case, quantum mechanics explains only part of the periodic table, and often it does not explain the features of the table that are of most interest to chemists (Scerri 1998). Pauli's introduction of the fourth quantum number, "spin" or "spin angular momentum," leads directly to the *Aufbau* principle, which states that the periodic table is constructed by placing electrons in lower energy levels first and then demonstrating that atoms with similar configurations have similar chemical properties. In this way, one can say that because chlorine and fluorine need one more electron to achieve a closed shell, they will behave similarly. However, this simple, unqualified explanation suffers from a number of anomalies. First, the filling sequence is not always strictly obeyed. Cobalt, nickel, and copper fill their shells in the sequence $3d^7 4s^2$, $3d^8 4s^2$, $3d^{10} 4s^1$. The superscripts denote the number of electrons in the subshell; the *s* shell can hold a maximum of two electrons and the five *d* orbitals ten. The observed order of filling is curious from the perspective of the unmodified *Aufbau* principle for a number of reasons. The 4*s* shell, which is supposed to be higher in energy than the 3*d* shell, has been occupied and closed first. Then, there is the "demotion" of one 4*s* electron in nickel to a 3*d* electron in copper. Second, configurations are supposed to explain why elements falling into the same group behave similarly, as in the example of chlorine and fluorine. Yet nickel, palladium, and platinum are grouped together because of their marked chemical similarities despite the fact that their outer shells have different configurations ($4s^2$, $5s^0$ and $6s^1$, respectively). These and other anomalies can be resolved using alternative derivations more closely tied to fundamental quantum mechanics, but the derivations require that the orbital approximation be dropped, and it was that approximation that was the basis for the assignment into the *s*, *p*, and *d* orbitals in the first place. It thus becomes an open question whether quantum mechanics, via the *Aufbau* principle, has explained the chemical periodicities encapsulated in the table.

Similar questions about the tenuous relation between physics and chemistry surround the concept of bonding. At a broad level, chemists employ two seemingly inconsistent representations, the valence bond (VB) and molecular orbital (MO) theories, to explain why atoms and molecules react. Both are calculational approximations inherited from atomic physics. The relations between the two theories and respective relations to the underlying quantum mechanics raise many issues of theory interpretation and realism about the chemical concepts (see below). Subsidiary concepts such as resonance are also invoked to explain the finer points of bonding. Chemists have offered competing realist interpretations of this concept, and philosophers have offered various realist and instrumentalist interpretations of it as well (Mosini 2000).

More specifically, a host of philosophical issues are raised within the molecular orbital theory. Here, bonding is pictured as due to the interaction of electrons in various orbitals. As noted earlier, the familiar spherical and dumbbell shapes arise only because of the orbital approximation. However, there is no reason to expect that the hydrogenic wavefunctions will look anything like the molecular ones (Bader 1990; Woody 2000). After the hydrogenic wavefunctions have been chosen as the basis for the calculation, they must be processed mathematically to arrive at a value for the energy of the orbital that is at all close to the experimentally observed value. The molecular wave equation is solved by taking linear combinations of the hydrogenic wave functions, forming the product of these combinations (the "configuration interaction" approach), and using the variational method to produce a minimal energy solution to the equation. The familiar orbitals appear only when these three steps in the complete solution have been omitted (Woody 2000). Thus, the idea that the familiar orbitals are responsible for the bonding is thrown into question. Yet the orbitals classify and explain how atoms and molecules bond extremely well. That they do provide deep, unified, and fertile representations and explanations seems curious from the perspective of fundamental quantum mechanics. Clearly, more analysis is required to understand the relation clearly. If one insists on a philosophical account of reduction that requires the mathematical or logical derivability of one theory from another, how orbitals achieve their power remains obscure. Even when one abandons that philosophical account, it is not clear how the representations have the organizing and explanatory power they do (see Reductionism).

As a final illustration of the difficulty of connecting physics and chemistry in any sort of strict fashion, consider the concept of molecular shape. Partly through the tradition of orbitals described above and partly through a historical tradition of oriented bonding that arose well before the concept of orbitals was introduced, chemists commonly explain many behaviors of molecules as due to their three-dimensional orientation in space. Molecules clearly react as if they are oriented in three-dimensional space. For example, the reaction $I^- + CH_3Br \rightarrow ICH_3 + Br^-$ is readily explained by invoking the notion that the iodine ion (I^-) attacks the carbon (C) on the side away from the bromine (Br) atom. (This can be detected by substituting deuterium atoms for one of the hydrogens [H] and measuring subsequent changes in spectroscopic properties.) As noted before, however, such explanations are suspect within quantum mechanics, since talk of oriented bonds and quasi-independent substituents in the reaction is questionable. Orientations must be "built into" the theory by parameterizing some of the theoretical variables. Unfortunately, there is no strict quantum mechanical justification for the method by which orientation in space is derived. Orientation relies on a notion of a nuclear frame surrounded by electrons. This notion is constructed via the Born-Oppenheimer approximation, which provides a physical rationale for why the nuclear positions should be slowly varying with respect to the electronic motions. In some measurement regimes, the approximation is invalid, and correct predictions are achieved only by resorting to a more general molecular Hamiltonian. In more common measurement regimes, the approximation is clearly valid. But, as with the issues involved in the case of the periodic table and orbitals, the physics alone do not tell us why it is valid. There is a physical justification for the procedure, but this justification has no natural representation with the available physical theory (Weininger 1984). Should justification based on past experience be trusted, or should the theory correct the interpretive practice? In any case, the chemistry is consistent with, but not yet derivable from, the physics.

All three examples are connected with a methodological issue mentioned earlier, *viz.*, the type of theory that chemists find useful. As previously noted, chemists often must parameterize the physical theories at their disposal to make them useful. All three of the cases described above involve such parameterization, albeit in different ways. How is it that such parameterizations uncover useful patterns in the data? Are they explanatory? When are they acceptable and when not (Ramsey 1997)? These and a host of similar questions remain to be answered.

Other epistemological and ontological issues raised in the practice of chemistry remain virtually unexplored. For instance, the question of whether one molecule is identical to another is answered by referring to some set of properties shared by the two samples. Yet the classification of two molecules as of the "same" type will vary, since different theoretical representations and experimental techniques detect quite different properties (Hoffmann 1995). For instance, reference can be made to the space-filling property of molecules, their three-dimensional structure, or the way they respond to an electric field. Additionally, the determination of sameness must be made in light of the question, For what function or purpose? For instance, two molecules of hemoglobin, which are large biological molecules, might have different isotopes of oxygen at one position. While this difference might be useful in order to discover the detailed structure of the hemoglobin molecule, it is usually irrelevant when talking about the molecule's biological function.

How the explanatory practices of chemistry stand in relation to the available philosophical accounts and to the practices of other sciences remains an important question. Chemical explanations are very specific, often lacking the generality invoked in philosophical accounts of explanation. Moreover, chemists invoke a wide variety of models, laws, theories, and mechanisms to explain the behavior and structure of molecules. Finally, most explanations require analogically based and/or experimentally derived adjustments to the theoretical laws and regularities in order for the account to be explanatory.

As mentioned earlier, chemistry is an extremely experimental science. In addition to unifying and fragmenting research programs in chemistry, new laboratory techniques have dramatically changed the epistemology of detection and observation in chemistry (e.g., from tapping manometers to reading NMR [nuclear magnetic resonance] outputs). As yet, however, there is no overarching, complete study of the changes in the epistemology of experimention in chemistry: for example, what chemists count as observable (and how this is connected to what they consider to be real), what they assume counts as a complete explanation, what they assume counts as a successful end to an experiment, etc.

Additionally, what are the relations between academic and industrial chemistry? What are the relative roles of skill, theory, and experiment in these two arenas of inquiry? Last but not least, there are pressing ethical questions. The world has been transformed by chemical products. Chemistry has blurred the distinction between the natural and the artificial in confusing ways (Hoffmann 1995). For instance, catalytically produced ethanol is chemically identical to the ethanol produced in fermentation. So is the carbon dioxide produced in a forest fire and in a car's exhaust. Why are there worries about the exhaust fumes but not the industrially produced ethanol? Is this the appropriate attitude? Last but not least, chemicals have often replaced earlier dangerous substances and practices; witness the great number of herbicides and insecticides available at the local garden center and the prescription medicines available at the pharmacy. Yet these replacements are often associated with a cost. One need think only of DDT or thalidomide to be flung headlong into ethical questions regarding the harmfulness and use of human-made products.

Many of the above topics have not been analyzed in any great depth. Much remains to be done to explore the methodology and philosophy of the chemical sciences.

Jeffry L. Ramsey

References

Bader, R. (1990), Atoms in Molecules: a Quantum Theory. New York: Oxford University Press.

Baird, D. (2000), "Encapsulating Knowledge: The Direct Reading Spectrometer," *Foundations of Chemistry* 2: 5–46.

Bensaude-Vincent, B., and I. Stengers (1996), *A History of Chemistry*. Translated by D. van Dam. Cambridge, MA: Harvard University Press.

Brock, W. (1993), *The Norton History of Chemistry*. New York: W. W. Norton.

Foster, J. M., and S. F. Boys (1960), "Quantum Variational Calculations for a Range of CH_2 Configurations," *Reviews of Modern Physics* 32: 305–307.

Gavroglu, K. (1997), "Philosophical Issues in the History of Chemistry," *Synthese* 111: 283–304.

Hoffmann, R. (1995), *The Same and Not the Same*. New York: Columbia University Press.

Klein, U. (1999), "Techniques of Modeling and Paper-Tools in Classical Chemistry," in M. Morgan and M. Morrison (eds.), *Models as Mediators*. New York: Cambridge University Press, 146–167.

Mosini, V. (2000), "A Brief History of the Theory of Resonance and of Its Interpretation," *Studies in History and Philosophy of Modern Physics* 31B: 569–581.

Nye, M. J. (1993), *From Chemical Philosophy to Theoretical Chemistry*. Berkeley and Los Angeles: University of California Press.

Paneth, F. (1962), "The Epistemological Status of the Concept of Element," *British Journal for the Philosophy of Science* 13: 1–14, 144–160.

Primas, H. (1983), *Chemistry, Quantum Mechanics and Reductionism*. New York: Springer Verlag.

Ramsey, J. (1997), "Between the Fundamental and the Phenomenological: The Challenge of 'Semi-Empirical' Methods," *Philosophy of Science* 64: 627–653.

Scerri, E. (1998), "How Good Is the Quantum Mechanical Explanation of the Periodic System?" *Journal of Chemical Education* 75: 1384–1385.

Schummer, J. (1996), *Realismus und Chemie*. Wuerzburg: Koenigshausen and Neumann.

Slater, L. (2002), "Instruments and Rules: R. B. Woodward and the Tools of Twentieth-century Organic Chemistry," *Studies in History and Philosophy of Science* 33: 1–32.

van Brakel, J. (2001), *Philosophy of Chemistry: Between the Manifest and the Scientific Image*. Leuven, Belgium: Leuven University Press.

Weininger, S. (1984), "The Molecular Structure Conundrum: Can Classical Chemistry Be Reduced to Quantum Chemistry?" *Journal of Chemical Education* 61: 939–944.

Woody, A. (2000), "Putting Quantum Mechanics to Work in Chemistry: The Power of Diagrammatic Representation," *Philosophy of Science* 67(suppl): S612–S627.

See also **Explanation; Laws of Nature; Quantum Mechanics; Reductionism**

NOAM CHOMSKY

(7 December 1928–)

Avram Noam Chomsky received his Ph.D in linguistics from the University of Pennsylvania and has been teaching at Massachusetts Institute of Technology since 1955, where he is currently Institute Professor. Philosophers are often familiar with the early work of Chomsky (1956, 1957, 1959a, and 1965), which applied the methods of formal language theory to empirical linguistics, but his work has also incorporated a number of philosophical assumptions about the nature of scientific practice—many of which are defended in his writings.

This entry will first describe the development and evolution of Chomsky's theory of generative linguistics, highlighting some of the philosophical assumptions that have been in play. It will then turn to some of the methodological debates in generative linguistics (and scientific practice more generally), focusing on Chomsky's role in these debates.

The Development and Evolution of Generative Grammar

A number of commentators have suggested that Chomsky's early work in generative linguistics initiated a kind of Kuhnian paradigm shift in linguistic theory. While Chomsky himself would reject this characterization (at least for his initial work in generative grammar), it is instructive to examine the development of generative linguistics, for it provides an excellent laboratory for the study of the development of a young science, and in particular it illuminates some of the philosophical prejudice that a young science is bound to encounter.

Chomsky's role in the development of linguistic theory and cognitive science generally can best be appreciated if his work is placed in the context of the prevailing intellectual climate in the 1950s—one in which behaviorism held sway in psychology departments and a doctrine known as American Structuralism was prevalent in linguistics departments.

American Structuralism, in particular as articulated by Bloomfield (1933 and 1939), adopted a number of key assumptions that were in turn adopted from logical empiricism (see Logical Empiricism). Newmeyer (1986, Ch. 1) notes that the following assumptions were in play:

- All useful generalizations are inductive generalizations.
- Meanings are to be eschewed because they are occult entities—that is, because they are not directly empirically observable.
- Discovery procedures like those advocated in logical empiricism should be developed for the proper conduct of linguistic inquiry.
- There should be no unobserved processes.

One of the ways in which these assumptions translated into theory was in the order that various levels of linguistic description were to be tackled. The American Structuralists identified four levels: phonemics (intuitively the study of sound patterns), morphemics (the study of words, their prefixes and suffixes), syntax (the study of sentence-level structure), and discourse (the study of cross-sentential phenomena). The idea was that proper methodology would dictate that one begin at the level of phonemics, presumably because it is closer to the data; then proceed to construct a theory of morphemics on the foundations of phonemics; and then proceed to construct a theory of syntax, etc.

Notice the role that the concepts of logical empiricism played in this proposed methodology. One finds radical reductionism in the idea that every level must be reducible to the more basic phonemic level; verificationism in the contention that the phonemic level is closely tied to sense experience; and discovery procedures in the suggestion that this overall order of inquiry should be adopted (see Reductionism; Verifiability).

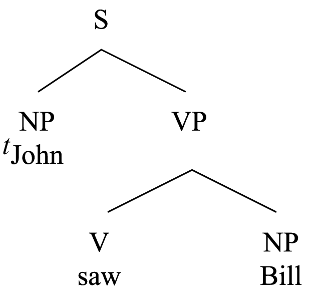
Chomsky rejected most if not all of these assumptions early on (see Chomsky [1955] 1975, introduction, for a detailed discussion). As regards discovery procedures, for example, he rejected them while still a matriculating graduate student, then holding a position in the Harvard Society of Fellows:

> By 1953, I came to the same conclusion [as Morris Halle] if the discovery procedures did not work, it was not because I had failed to formulate them correctly, but because the entire approach was wrong. ... [S]everal years of intense effort devoted to improving discovery procedures had come to naught, while work I had been doing during the same period on generative grammars and explanatory theory, in almost complete isolation, seemed to be consistently yielding interesting results. (1979: 131)

Chomsky also rejected the assumption that all processes should be "observable"—early theories of transformational grammar offered key examples of unobservable processes. For example, in his "aspects theory" of generative grammar (Chomsky 1965), the grammar is divided into two different "levels of representation," termed initially *deep structure* and *surface structure*. The deep-structure representations were generated by a context-free phase structure grammar—that is, by rules (of decomposition, "→") of the following form, where S stands for sentence, NP for noun phrase, VP for verb phrase, etc.

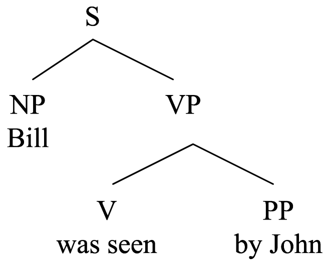
S → NP VP
VP → V NP
NP → John
NP → Bill
V → saw

These rewriting rules then generated linguistic representations of the following form:



(1)

Crucially for Chomsky, the objects of analysis in linguistic theory were not the terminal strings of words, but rather *phrase markers*—structured objects like (1). Transformational rules then operated on these deep-structure representations to yield surface-structure representations. So, for example, the operation of passivization would take a deep-structure representation like (1) and yield the surface-structure representation (abstracting from detail) in (2):



(2)

The sentence in (3) is therefore a complex object consisting of (at a minimum) an ordered pair of the two representations corresponding to (1) and (2):

Bill was seen by John. (3)

Clearly, Chomsky was committed not only to "unobserved" processes in the guise of transformations, but also to unobserved levels of representation. No less significant was the nature of the data that Chomsky admitted—not utterances or written strings, but rather speakers' judgments of acceptability and meaning. Thus, (3) is not a datum because it has been written or spoken, but rather because speakers have intuitions that it is (would be) an acceptable utterance. Here again, Chomsky broke with prevailing methodology in structuralist linguistics and, indeed, behaviorist psychology, by allowing intuitions rather than publicly available behaviors as data.

Generative grammar subsequently evolved in response to a number of internal pressures. Crucially, the number of transformations began to proliferate in a way that Chomsky found unacceptable. Why was this proliferation unacceptable? Early on in the development of generative grammar, Chomsky had made a distinction between the *descriptive adequacy* and the *explanatory adaquacy* of an empirical linguistic theory (Chomsky 1965 and 1986b). In particular, if a linguistic theory is to be explanatorially adequate, it must not merely describe the facts, but must do so in a way that explains how humans are able to learn languages. Thus, linguistics was supposed to be embeddable into cognitive science more broadly. But if this is the case, then there is a concern about the unchecked proliferation of rules—such rule systems might be descriptively adequate, but they would fail to account for how we learn languages (perhaps due to the burden of having to learn all those language-specific rules).

Chomsky's initial (1964 and 1965) solution to this problem involved the introduction of conditions on transformations (or constraints on movement), with the goal of reducing the complexity of the descriptive grammar. In Chomsky (1965), for example, the recursive power of the grammar is shifted from the transformations to the phrase structure rules alone. In the "extended standard theory" of the 1970s, there was a reduction of the phrase structure component with the introduction of "X-bar theory," and a simplification of the constraints on movement. This was followed by a number of proposals to reduce the number and types of movement rules themselves. This came to a head (Chomsky 1977 and 1981a) with the abandonment of specific transformations altogether for a single rule ("move-α"), which stated, in effect, that one could move anything anywhere. This one rule was then supplemented with a number of constraints on movement. As Chomsky and a number of other generative linguists were able to show, it was possible to reduce a great number of transformations to a single-rule move-α and to a handful of constraints on movement.

Chomsky (1981b, 1982, and 1986a) synthesized subsequent work undertaken by linguists working in a number of languages, ranging from the romance languages to Chinese and Japanese, showing that other natural languages had similar but not identical constraints, and it was hypothesized that the variation was due to some limited parametric variation among human languages. This work established the "principles and parameters" framework of generative grammar. To get an idea of this framework, consider the following analogy: Think of the language faculty as a prewired box containing a number of switches. When exposed to environmental data, new switch settings are established. Applying this metaphor, the task of the linguist is to study the initial state of the language faculty, determine the possible parametric variations (switch settings), and account for language variation in terms of a limited range of variation in parameter settings. In Chomsky's view (2000, 8), the principles-and-parameters framework "gives at least an outline of a genuine theory of language, really for the first time." Commentators (e.g., Smith 2000, p. xi) have gone so far as to say that it is "the first really novel approach to language of the last two and a half thousand years." In what sense is it a radical departure? For the first time it allowed linguists to get away from simply constructing rule systems for individual languages and to begin exploring in a deep way the underlying similarities of human languages (even across different language families), to illuminate the principles that account for those similarities, and ultimately to show how those principles are grounded in the mind/brain. In Chomsky's view, the principles-and-parameters framework has yielded a number of promising results, ranging from the discovery of important similarities between *prima facie* radically different languages like Chinese and English to insights into the related studies of language acquisition, language processing, and acquired linguistic deficits (e.g., aphasia). Perhaps most importantly, the principles-and-parameters framework offered a way to resolve the tension between the two goals of descriptive adequacy and explanatory adequacy.

Still working within the general principles-and-parameters framework, Chomsky (1995 and 2000, Ch. 1) has recently articulated a research program that has come to be known as the "minimalist program," the main idea behind which is the working hypothesis that the language faculty is not the product of messy evolutionary tinkering—for example, there is no redundancy, and the only resources at work are those that are driven by "conceptual necessity." Chomsky (1995, ch. 1) initially seemed to hold that in this respect the language faculty would be unlike other biological functions, but more recently (2001) he seems to be drawn to D'Arcy Thompson's theory that the core of evolutionary theory consists of physical/mathematical/chemical principles that sharply constrain the possible range of organisms. In this case, the idea would be not only that those principles constrain low-level biological processes (like sphere packing in cell division) but also that such factors might be involved across the board—even including the human brain and its language faculty.

In broadest outline, the minimalist program works as follows: There are two levels of linguistic representation, phonetic form (PF) and logical form (LF), and a well-formed sentence (or linguistic *structure*) must be an ordered pair $<\pi, \lambda>$ of these representations (where π is a phonetic form and λ is the logical form). PF is taken to be the level of representation that is the input to the performance system (e.g., speech generation), and LF is, in Chomsky's terminology, the input to the conceptual/intensional system. Since language is, if nothing else, involved with the pairing of sounds and meanings, these two levels of representation are conceptually necessary. A minimal theory would posit no other levels of representation.

It is assumed that each sentence (or better, *structure*, Σ) is constructed out of an *array* or *numeration*, N, of lexical items. Some of the items in the numeration will be part of the pronounced (written) sentence, and others will be part of a universal inventory of lexical items freely inserted into all numerations. Given the numeration N, the computational system (C_{HL}) attempts to derive (compute) well-formed PF and LF representations, converging on the pair $<\pi, \lambda>$. The derivation is said to *converge* at a certain level if it yields a representation that is interpretable at that level. If it fails to yield an interpretable representation, the derivation *crashes*. Not all converging derivations yield structures that belong to a given language L. Derivations must also meet certain economy conditions.

Chomsky (2000, 9) notes that the import of the minimalist program is not yet clear. As matters currently stand, it is a subresearch program within the principles-and-parameters framework that is showing some signs of progress—at least enough to encourage those working within the program. As always, the concerns are to keep the number of principles constrained, not just to satisfy economy constraints, but to better facilitate the embedding of linguistics into theories of language acquisition, cognitive psychology, and, perhaps most importantly, general biology.

Some Conceptual Issues in Generative Grammar

While Chomsky would argue that he does not have a philosophy of science per se and that his philosophical observations largely amount to common sense, a number of interesting debates have arisen in the wake of his work. The remainder of this entry will review some of those debates.

On the Object of Study

Chomsky (1986b) draws the distinction between the notions of *I*-language and *E*-language, where *I*-language is the language faculty discussed above, construed as a chapter in cognitive psychology and ultimately human biology. *E*-language, on the other hand, comprises a loose collection of theories that take language to be a shared social object, established by convention and developed for purposes of communication, or an abstract mathematical object of some sort.

In Chomsky's view (widely shared by linguists), the notion of a 'language' as it is ordinarily construed by philosophers of language is fundamentally incoherent. One may talk about "the English language" or "the French language" but these are loose ways of talking. Typically, the question of who counts as speaking a particular language is determined more by political boundaries than actual linguistic variation. For example, there are dialects of German that, from a linguistic point of view, are closer to Dutch than to standard German. Likewise, in the Italian linguistic situation, there are a number of so-called dialects only some of which are recognized as "official" languages by the Italian government. Are the official languages intrinsically different from the "mere" dialects? Not in any linguistic sense. The decision to recognize the former as official is entirely a political decision. In the words attributed to Max Weinreich: A language is a dialect with an army and a navy. In this case, a language is a dialect with substantial political clout and maybe a threat of separatism.

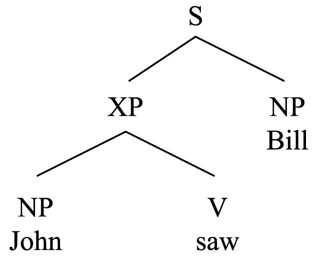
Chomsky (1994 and 2000, Chap. 2) compares talk of languages (i.e., *E*-languages) to saying that two cities are "near" each other; whether two cities are near depends on one's interests and one's mode of transportation and very little on brute facts of geography. In the study of language, the notion of 'sameness' is no more respectable than that of 'nearness' in geography. Informally we might group together ways of speaking that seem to be similar (relative to one's interests), but such groupings have no real scientific merit. As a subject of natural inquiry, the key object of study has to be the language faculty and its set of possible parametric variations.

Not only is the notion of an *E*-language problematic, but it will not help to retreat to talk about *E-dialects*. The problem is that what counts as a separate *E*-dialect is also incoherent from a scientific point of view. For example, Chomsky (2000, 27) reports that in his idiolect, the word "ladder" rhymes with "matter" but not with "madder." For others, the facts do not cut in this way. Do they speak the same dialect as Chomsky or not? There is no empirical fact of the matter here; it all depends on individuals' desires to identify linguistically with each other. Even appeals to mutual intelligibility will not do, since what one counts as intelligible will depend much more on one's patience, one's ambition, and one's familiarity with the practices of one's interlocutors than it will on brute linguistic facts.

If the notion of *E*-language and *E*-dialect are incoherent, is it possible to construct a notion of *E-idiolect*?—that is, to identify idiolects by external criteria like an individual's spoken or written language? Apparently it is not. Included in what a person says or writes are numerous slips of the tongue, performance errors, etc. How is one we to rule those out of the individual's *E*-idiolect? Appeal to the agent's linguistic community will not do, since that would in turn require appeal to an *E*-language, and, for the reasons outlined above, there is no meaningful way to individuate *E*-languages. In the *I*-language approach, however, the problem of individuating *I*-idiolects takes the form of a coherent empirical research project. The idiolect (*I*-idiolect) is determined by the parametric state of *A*'s language faculty, and the language faculty thus determines *A*'s linguistic *competence*. Speech production that diverges from this competence can be attributed to *performance* errors. Thus, the competence/performance distinction is introduced to illuminate the distinction between sounds and interpretations that are part of *A*'s grammar and those that are simply mistakes. The *E*-language perspective has no similar recourse.

The Underdetermination of Theory by Evidence

One of the first philosophical issues to fall out from the development of generative grammar has been the dispute between Quine (1970) and Chomsky (1969 and 2000, Chap. 3) on the indeterminacy of grammar. Similar to his argument for the indeterminacy of meaning, Quine held that there is no way to adjudicate between two descriptively adequate sets of grammatical rules. So, for example, imagine two rule sets, one envisioning structures like (1) above, and another positing the following:



(4)

Chomsky maintained that Quine's argument is simply a recapitulation of the standard scientific problem of the underdetermination of theory by evidence. And in any case it is not clear that there are no linguistic tests that would allow us to choose between (1) and (4)—there are, after all, "constituency tests" (e.g., involving movement possibilities) that would allow us to determine whether the XP or the VP is the more plausible constituent.

Even if there are several grammars that are consistent with the available linguistic facts (the facts are not linguistic *behavior*, for Chomsky, but intuitions about acceptability and possible interpretation), one still has the additional constraint of which theory best accounts for the problem of language acquisition, acquired linguistic deficits (e.g., from brain damage), linguistic processing, etc. In other words, since grammatical theory is embedded within cognitive psychology, the choice between candidate theories can, in principle, be radically constrained. But further, even if there were two descriptively adequate grammars, each of which could be naturally embedded within cognitive psychology, there remain standard best-theory criteria (simplicity, etc.) that can help us to adjudicate between the theories.

Intrinsic Versus Relational Properties in Linguistics

In the philosophy of science, it is routine to make the distinction between "intrinsic" and "relational" properties. So, for example, the rest mass of an object would be an intrinsic property, and its weight would be a relational property (since it depends upon the mass of the body that the object is standing on). Similarly, in the philosophy of psychology there is a common distinction between "individualistic" and "externalist" properties. So, for example, there is the question of whether psychological states supervene on individualistic properties (intrinsic properties that hold of the individual in isolation) or whether they supervene on "externalist" properties (in effect, on relations between the agent and its environment) (see Supervenience). Chomsky has argued that all scientifically interesting psychological and linguistic properties supervene upon individualist (intrinsic) properties. In particular, there is a brute fact about the state of an individual's language faculty, and that fact is determined in turn by facts about the individual in isolation—not by the environment in which the individual is embedded. Because the language faculty is part of one's biological endowment, the nature of the representations utilized by the language faculty are fixed by biology and are not sensitive to environmental issues such as whether one is moving about on Earth or a phenomenologically identical planet with a different microstructure (e.g., Putnam's "Twin Earth").

Thus Chomsky takes issue with philosophers like Burge (1986), who argues in *Individualism and Psychology* that the content of the representations posited in psychology are determined at least in part by environmental factors. Chomsky holds that if the notion of content involves externalist or environmental notions, then it is not clear that it can play an interesting role in naturalistic inquiry in cognitive psychology (see Psychology, Philosophy of).

If environmentalism is to be rejected in psychology, then it naturally must be rejected in the semantics of natural language as well. That is: if the task of the linguist is to investigate the nature of *I*-language; if the nature of *I*-language is a chapter of cognitive psychology; and if cognitive psychology is an individualistic rather than a relational science, semantics will want to eschew relational properties like reference (where 'reference' is construed as a relation between a linguistic form and some object in the external environment). Thus Chomsky (1975 and 2000) rejects the notion of reference that has been central to the philosophy of language for the past three decades, characterizing it as an ill-defined technical notion (certainly one with no empirical applications), and, following Strawson, suggests that in the informal usage individuals 'refer' but linguistic objects do not.

This conclusion has immediate results for the notion of theory change in science. If the notion of reference is suspect or incoherent, then it can hardly be employed in an account of theory change (as in Putnam 1988). How then can one make sense of theory change? Chomsky (2000) suggests the following:

> Some of the motivation for externalist approaches derives from the concern to make sense of the history

> of science. Thus, Putnam argues that we should take the early Niels Bohr to have been referring to electrons in the quantum-theoretic sense, or we would have to "dismiss all of his 1900 beliefs as totally wrong" (Putnam 1988), perhaps on a par with someone's beliefs about angels, a conclusion that is plainly absurd . . . .
>
> Agreeing . . . that an interest in intelligibility in scientific discourse across time is a fair enough concern, still it cannot serve as the basis for a general theory of meaning; it is, after all, only one concern among many, and not a central one for the study of human psychology. Furthermore, there are internalist paraphrases. Thus we might say that in Bohr's earlier usage, he expressed beliefs that were literally false, because there was nothing of the sort he had in mind in referring to electrons; but his picture of the world and articulation of it was structurally similar enough to later conceptions so that we can distinguish his beliefs about electrons from beliefs about angels. (43)

Against Teleological Explanation in Linguistics

A number of philosophers and linguists have thought that some progress can be made in the understanding of language by thinking of it as principally being a medium of communication. Chomsky rejects this conception of the nature of language, arguing that the language faculty is not *for* communication in any interesting sense. Of course, by rejecting the contention that language is a social object established for purposes of communication, Chomsky has not left much room for thinking of language in this way. But he also rejects standard claims that *I*-language must have evolved for the selectional value of communication; he regards such claims as without basis in fact. On this score, Chomsky sides with Gould (1980), Lewontin (1990), and many evolutionary biologists in supposing that many of our features (cognitive or anatomical) did not necessarily evolve for selectional reasons, but may have been the result of arbitrary hereditary changes that have perhaps been co-opted (see Evolution). Thus the language faculty may not have evolved for purposes of communication but may have been co-opted for that purpose, despite its nonoptimal design for communicative purposes.

In any case, Chomsky cautions that even if there was selectional pressure for the language faculty to serve as a means of communication, selection is but one of many factors in the emerging system. Crucially (2000, 163), "physical law provides narrow channels within which complex organisms may vary," and natural selection is only one factor that determines how creatures may vary within these constraints. Other factors (as Darwin himself noted) will include nonadaptive modifications and unselected functions that are determined from structure (see Function).

Inductive Versus Abductive Learning

A number of debates have turned on whether language acquisition requires a dedicated language faculty or whether "general intelligence" is enough to account for linguistic competence. Chomsky considers the general-intelligence thesis hopelessly vague, and argues that generalized inductive learning mechanisms make the wrong predictions about which hypotheses children would select in a number of cases. Consider the following two examples from Chomsky (1975 and [1980] 1992a):

The man is tall (5)

Is the man tall? (6)

Chomsky observes that confronted with evidence of question formation like that in (5)–(6) and given a choice between hypothesis (H1) and (H2), the generalized inductive learning mechanism will select (H1):

Move the first "is" to the front of the sentence. (H1)

Move the first "is" following the first NP to the front of the sentence. (H2)

But children apparently select (H2), since in forming a question from (7), they never make the error of producing (8), but always opt for (9):

The man who is here is tall. (7)

*Is the man who here is tall? (8)

Is the man who is here tall? (9)

Note that this is true despite the fact that the only data they have been confronted with previous to encountering (7) is simple data like (5)–(6). Chomsky's conclusion is that whatever accounts for children's acquisition of language, it cannot be generalized inductive learning mechanisms, but rather must be a system with structure-dependent principles/rules. Chomsky (1975, ch.1; 2000, 80) compares such learning to Peircian abduction. In other contexts this has been cast as a thesis about the "modularity" of language—that is, that there is a dedicated language acquisition device, and it, rather than some vague notion of general intelligence, accounts for the acquisition of language (see Evolutionary Psychology).

Putnam (1992) once characterized Chomsky's notion of a mental organ like the language faculty as being "biologically inexplicable," but Chomsky ([1980] 1992b) has held that it is merely "unexplained" and not inexplicable; in his view the language faculty is thus on the same footing as many other features of our biology (see Abduction; Inductive Logic).

The Science-Forming Faculty

On a more general and perhaps more abstract level, Chomsky has often spoken of a "science-forming faculty," parallel to our language faculty. The idea is that science could not have formed in response to mere inductive generalizations, but that human beings have an innate capacity to develop scientific theories (Chomsky credits C. S. Peirce with this basic idea). Despite Star Trekkian assumptions to the contrary, extraterrestrials presumably do not have a science-forming faculty like humans do, and may go about theorizing in entirely different ways, which would not be recognized as "scientific" by humans.

While the science-forming faculty may be limited, Chomsky would not concede that its limits are necessarily imposed by the selectional pressures of the prehistoric past. Just as the language faculty may not have evolved principally in response to selectional pressures, so too the science-forming faculty may have emerged quite independently of selectional considerations. As Chomsky (2000, Chap. 6) notes (citing Lewontin 1990), insects may seem marvelously well adapted to flowering plants, but in fact insects evolved to almost their current diversity and structure millions of years before flowering plants existed. Perhaps it is a similar situation with the science-forming faculty. That is, perhaps it is simply a matter of good fortune for humans that the science-forming faculty is reliable and useful, since it could not have evolved to help us with quantum physics, for example.

The Limits of Science

The notion of a science-forming faculty also raises some interesting questions about the limits of human ability to understand the world. Since what one can know through naturalistic endeavors is bounded by the human science-forming faculty, which in turn is part of the human biological endowment, it stands to reason that there are questions that will remain mysteries—or at least outside the scope of naturalistic inquiry:

> Like other biological systems, SFF [the science-forming faculty] has its potential scope and limits; we may distinguish between *problems* that in principle fall within its range, and *mysteries* that do not. The distinction is relative to humans; rats and Martians have different problems and mysteries and, in the case of rats, we even know a fair amount about them. The distinction also need not be sharp, though we certainly expect it to exist, for any organism and any cognitive faculty. The successful natural sciences, then, fall within the intersection of the scope of SFF and the nature of the world; they treat the (scattered and limited) aspects of the world that we can grasp and comprehend by naturalistic inquiry, in principle. The intersection is a chance product of human nature. Contrary to speculations since Peirce, there is nothing in the theory of evolution, or any other intelligible source, that suggests that it should include answers to serious questions we raise, or even that we should be able to formulate questions properly in areas of puzzlement. (2000, Ch. 4)

The question of what the "natural" sciences are, then, might be answered, narrowly, by asking what they have achieved; or more generally, by inquiry into a particular faculty of (the human) mind, with its specific properties.

The Mind/Body Problem and the Question of Physicalism

Chomsky has consistently defended a form of methodological monism (he is certainly no dualist); but, for all that, he is likewise no materialist. In Chomsky's view the entire mind/body question is ill-formed, since there is no coherent notion of physical body. This latter claim is not in itself unique; Crane and Mellor (1990) have made a similar point. There is a difference, however; for Crane and Mellor, developments in twentieth-century science have undermined physicalism, but for Chomsky the notion of physical body was already undermined by the time of Newton:

> Just as the mechanical philosophy appeared to be triumphant, it was demolished by Newton, who reintroduced a kind of "occult" cause and quality, much to the dismay of leading scientists of the day, and of Newton himself. The Cartesian theory of mind (such as it was) was unaffected by his discoveries, but the theory of body was demonstrated to be untenable. To put it differently, Newton eliminated the problem of "the ghost in the machine" by exorcising the machine; the ghost was unaffected. (2000, 84)

In Chomsky's view, then, investigations into the mind (in the guise of cognitive science generally or linguistics in particular) can currently proceed without worrying about whether they hook up with what is known about the brain, or even fundamental particles. The unification of science remains a goal, but in Chomsky's view it is not the study of mind that must be revised so as to conform to physical theory, but rather physical theory may eventually have to incorporate what is learnt in the study of the mind. According to Chomsky this is parallel to the situation that held prior to the unification of chemistry and physics; it was not chemistry that needed to be modified to account for what was known about physics, but in fact just the opposite:

> Large-scale reduction is not the usual pattern; one should not be misled by such dramatic examples as the reduction of much of biology to biochemistry in the middle of the twentieth century. Repeatedly, the more ''fundamental'' science has had to be revised, sometimes radically, for unification to proceed. (2000, 82)

Conclusion

The influence of Chomsky's work has been felt in a number of sciences, but perhaps the greatest influence has been within the various branches of cognitive science. Indeed, Gardner (1987) has remarked that Chomsky has been the single most important figure in the development of cognitive science. Some of Chomsky's impact is due to his role in arguing against behaviorist philosophers such as Quine; some of it is due to work that led to the integration of linguistic theory with other sciences; some of it is due to the development of formal tools that were later employed in disciplines ranging from formal language theory (cf. the "Chomsky Hierarchy") to natural language processing; and some of it is due to his directly engaging psychologists on their own turf. (One classic example of this was Chomsky's [1959b] devastating review of Skinner's [1957] *Verbal Behavior*. See also his contributions to Piatelli-Palmerini 1980) (See Behaviorism).

With respect to his debates with various philosophers, Chomsky has sought to expose what he has taken to be double standards in the philosophical literature. In particular he has held that while other sciences are allowed to proceed where inquiry takes them without criticism by armchair philosophers, matters change when the domain of inquiry shifts to mind and language:

> The idea is by now a commonplace with regard to physics; it is a rare philosopher who would scoff at its weird and counterintuitive principles as contrary to right thinking and therefore untenable. But this standpoint is commonly regarded as inapplicable to cognitive science, linguistics in particular. Somewhere in-between, there is a boundary. Within that boundary, science is self-justifying; the critical analyst seeks to learn about the criteria for rationality and justification from the study of scientific success. Beyond that boundary, everything changes; the critic applies independent criteria to sit in judgment over the theories advanced and the entitities they postulate. This seems to be nothing more than a kind of ''methodological dualism,'' far more pernicious than the tradtional metaphysical dualism, which was a scientific hypothesis, naturalistic in spirit. Abandoning this dualist stance, we pursue inquiry where it leads. (2000, 112)

Peter Ludlow

References

Bloomfield, L. (1933), *Language*. New York: Holt, Rinehart and Winston.

——— (1939), *Linguistic Aspects of Science: International Encyclopedia of Unified Science* (Vol. 1, No. 4). Chicago: University of Chicago Press.

Burge, T. (1986), "Individualism and Psychology," *Philosophical Review* 95: 3–45.

Chomsky, N. ([1955] 1975), *The Logical Structure of Linguistic Theory*. Chicago: University of Chicago Press.

——— (1956), "Three Models for the Description of Language," *I.R.E. Transactions of Information Theory* IT-2: 113–124.

——— (1957), *Syntactic Structures*. The Hague: Mouton.

——— (1959a), "On Certain Formal Properties of Grammars," *Information and Control* 2: 137–167.

——— (1959b), "Review of B. F. Skinner, *Verbal Behavior*," *Language* 35, 26–57.

——— (1964), *Current Issues in Linguistic Theory*. The Hague: Mouton & Co.

——— (1965), *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

——— (1969), "Quine's Empirical Assumptions," in D. Davidson and J. Hintikka (eds.), *Words and Objections: Essays on the Work of Willard Van Quine*. Dordrecht, Netherlands: D. Reidel.

——— (1975), *Reflections on Language*. New York: Pantheon.

——— (1977), "Conditions on Rules of Grammar," in *Essays on Form and Interpretation*. Amsterdam: Elsevier North-Holland, 163–210.

——— (1979), *Language and Responsibility*. New York: Pantheon.

——— (1981a), *Lectures on Government and Binding*. Dordrecht, Netherlands: Foris Publications.

——— (1981b), "Principles and Parameters in Syntactic Theory," in N. Horstein and D. Lightfoot (eds.), *Explanation in Linguistics: The Logical Problem of Language Acquisition*. London: Longman.

——— (1982), *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.

——— (1986a), *Barriers*. Cambridge, MA: MIT Press.

——— (1986b), *Knowledge of Language*. New York: Praeger.

——— ([1980] 1992a), "On Cognitive Structures and Their Development," in B. Beakley and P. Ludlow (eds.), *The*

Philosophy of Mind: Classical Problems/Contemporary Issues. Cambridge, MA: MIT Press, 393–396.

——— ([1980] 1992b), "Discussion of Putnam's Comments," in B. Beakley and P. Ludlow (eds.), *The Philosophy of Mind: Classical Problems/Contemporary Issues*. Cambridge, MA: MIT Press, 411–422.

——— ([1988] 1992c), "From *Language and Problems of Knowledge*," in B. Beakley and P. Ludlow (eds.), *The Philosophy of Mind: Classical Problems/Contemporary Issues*. Cambridge, MA: MIT Press, 47–50.

——— (1994), "Noam Chomsky," in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 153–167.

——— (1995), *The Minimalist Program*. Cambridge, MA: MIT Press.

——— (2000), *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

——— (2001), "Beyond Explanatory Adequacy," *MIT Occasional Papers in Linguistics*, no. 20. MIT Department of Linguistics.

Crane, T., and D. H. Mellor (1990), "There Is No Question of Physicalism," *Mind* 99: 185–206.

Gardner, H. (1987), *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books.

Gould, S. J. (1980), *The Panda's Thumb: More Reflections in Natural History*. New York: W. W. Norton.

Lewontin, R. (1990), "The Evolution of Cognition," in D. N. Osherson and E. E. Smith (eds.), *An Invitation to Cognitive Science* (Vol. 3). Cambridge, MA: MIT Press, 229–246.

Newmeyer, Frederick (1986), *Linguistic Theory in America* (2nd ed.). San Diego: Academic Press.

Piatelli-Palmerini, M. (ed.) (1980), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.

Putnam, H. (1988), *Representation and Reality*. Cambridge, MA: MIT Press.

Quine, Willard Van (1970), "Methodological Reflections on Current Linguistic Theory," *Synthese*, 21: 386–398.

Skinner, B. F. (1957), *Verbal Behavior*. New York: Appleton-Century-Crofts.

Smith, N. (2000), Foreword, in Chomsky, *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

See also **Behaviorism; Cognitive Science; Innate/Acquired Distinction; Linguistics, Philosophy of; Psychology, Philosophy of; Physicalism**

CLASSICAL MECHANICS

Over the centuries classical mechanics has been a steady companion of the philosophy of science. It has played different parts, ranging (i) from positing its principles as a priori truths to the insight—pivotal for the formation of a modern philosophy of science—that modern physics requires a farewell to the explanatory ideal erected upon mechanics; (ii) from the physiological analyses of mechanical experiences to axiomatizations according to the strictest logical standards; and (iii) from the mechanistic philosophy to conventionalism (see Conventionalism; Determinism; Space-Time). Core structures of modern science, among them differential equations and conservation laws, as well as core themes of philosophy, among them determinism and the ontological status of theoretical terms, have emerged from this context. Two of the founders of classical mechanics, Galileo and Sir Isaac Newton, have often been identified with the idea of modern science as a whole. Owing to its increasing conceptual and mathematical refinement during the nineteenth century, classical mechanics gave birth to the combination of formal analysis and philosophical interpretation that distinguished the modern philosophy of science from the earlier *Naturphilosophie*. The works of Helmholtz, Mach, and Poincaré molded the historical character of the scientist-philosopher that would fully bloom during the emergence of relativity theory and quantum mechanics (see Mach, Ernest; Quantum Mechanics; Poincaré, Henri; Space-Time).

Mechanics became "classical" at the latest with the advent of quantum mechanics. Already relativistic field theory had challenged mechanics as the leading scientific paradigm. Consequently, present-day philosophers of science usually treat classical mechanics within historical case studies or as the first touchstone for new proposals of a general kind. Nonetheless, there are at least two lines on which classical mechanics in itself remains a worthy topic for philosophers. On the one hand, ensuing from substantial mathematical progress during the twentieth century, classical mechanics has developed into the theory of classical dynamical systems. Among the problems of interest to formally minded philosophers of science are the relationships between the different conceptualizations, issues of stability and chaotic behavior, and whether classical mechanics is a special case of the conceptual structures characteristic of other physical theories. On the other hand, classical mechanics or semiclassical approaches continue to be applied widely by working scientists and engineers. Those real-world applications involve a variety of features that substantially differ from the highly idealized textbook models to which physicists and philosophers are typically accustomed, and they require solution strategies whose epistemological status is far from obvious.

Often classical mechanics is used synonymously with Newtonian mechanics, intending that its content is circumscribed by Newton's famous three laws. The terms 'analytical mechanics' and 'rational mechanics' stress the mathematical basis and theoretical side of mechanics as opposed to 'practical mechanics,' which originally centered on the traditional simple machines: lever, wedge, wheel and axle, tackle block, and screw. But the domain of mechanics was being constantly enlarged by the invention of new mechanical machines and technologies. Most expositions of mechanics also include the theory of elasticity and the mechanics of continua. The traditional distinction between mechanics and physics was surprisingly long-lived. One reason was that well into the nineteenth century, no part of physics could live up to the level of formal sophistication that mechanics had achieved. Although of little importance from a theoretical point of view, it is still common to distinguish statics (mechanical systems in equilibrium) and dynamics that are subdivided into a merely geometrical part, kinematics, and kinetics.

Ancient Greek and Early Modern Mechanics

In Greek antiquity, mechanics originally denoted the art of voluntarily causing motions against the nature of the objects moved. Other than physics, it was tractable by the methods of Euclidean geometry. Archimedes used mechanical methods to determine the center of mass of complex geometrical shapes but did not recognize them as valid geometrical proofs. His Euclidean derivation of the law of the lever, on the other hand, sparked severe criticism from Mach ([1883] 1960), who held that no mathematical derivation whatsoever could replace experience.

Pivotal for bringing about modern experimental science was Galileo's successful criticism of Aristotelian mechanics. Aristotle had divided all natural motions into celestial motions, which were circular and eternal, and terrestrial motions, which were rectilinear and finite. Each of the four elements (earth, water, fire, air) moved toward its natural place. Aristotle held that heavy bodies fell faster than light ones because velocity was determined by the relation of motive force (weight) and resistance. All forces were contact forces, such that a body moving with constant velocity was continuously acted upon by a force from the medium. Resistance guaranteed that motion remained finite in extent. Thus there was no void, nor motion in the void. Galileo's main achievement was the idealization of a constantly accelerated motion in the void that is slowed down by the resistance of the medium. This made free fall amenable to geometry. For today's reader, the geometrical derivation of the law is clumsy; the ratio of different physical quantities $v = s/t$ (where s stands for a distance, t for a time interval, and v for a velocity) was not yet meaningful. Galileo expressly put aside what caused bodies to fall and referred to experimentation. Historians and philosophers have broadly discussed what and how Galileo reasoned from empirical evidence. Interpreters wondered, in particular, whether the thought experiment establishing the absurdity of the Aristotelian position was conclusive by itself or whether Galileo had simply repackaged empirical induction in the deductive fashion of geometry.

Huygens' most important contribution to mechanics was the derivation of the laws of impact by invoking the principle of energy conservation. His solution stood at the crossroads of two traditions. On the one hand, it solved the foundational problem of Cartesian physics, the program of which was to reduce all mechanical phenomena to contact forces exchanged in collision processes. On the other hand, it gave birth to an approach based upon the concept of energy, which became an alternative to the Newtonian framework narrowly understood.

The work of Kepler has repeatedly intrigued philosophers of diverging orientations. Utilizing the mass of observational data collected by Brahe, Kepler showed that the planetary orbits were ellipses. Kepler's second law states that the line joining the sun to the planet sweeps through equal areas in equal times, and the third law states that the square of the periods of revolution of any two planets are in the same proportion as the cubes of their semi-major axes. All three laws were merely kinematical. In his *Mysterium cosmographicum*, Kepler

([1596] 1981) identified the spacings of the then known six planets with the five platonic solids. It will be shown below that the peculiar shape of the solar system, given Newton's laws of universal gravitation, can be explained without reference to Kepler's metaphysical belief in numerical harmony.

On the Status of Newton's Laws

The most important personality in the history of classical mechanics was Newton. Owing to the activities of his popularizers, he became regarded as the model scientist; this admiration included a devotion to the methodology of his *Philosophiae naturalis principia mathematica* (Newton [1687, 1713] 1969), outlined in a set of rules preceding book III. But interpreters disagree whether Newton really pursued the Baconian ideal of science and licensed induction from phenomena or must be subsumed under the later descriptivist tradition that emerged with Mach ([1883] 1960) and Kirchhoff (1874). At any rate, the famous declaration not to feign hypothesis targeted the Cartesian and Leibnizian quest for a metaphysical basis of the principles of mechanics.

After the model of Euclid, the *Principia* began with eight definitions and three axioms or laws. Given Newton's empiricist methodology, interpreters have wondered about their epistemological status and the logical relations among them. Certainly, the axioms were neither self-evident truths nor mutually independent:

1. Every body continues in its state of rest or uniform motion in a right (i.e., straight) line, unless it is compelled to change that state by forces impressed upon it.
2. The change of motion is proportional to the motive force impressed and is made in the direction of the right line in which that force is impressed.
3. To every action there is always opposed an equal reaction; or the mutual actions of two bodies upon each other are always equal, and directed to contrary parts.

The proper philosophical interpretation of these three laws remained contentious until the end of the nineteenth century. Kant claimed to have deduced the first and third laws from the synthetic a priori categories of causality and reciprocity, respectively. The absolute distinction between rest and motion in the first law was based on absolute space and time. Within Kant's transcendental philosophy, Euclidean space-time emerged from pure intuition a priori. This was the stand against which twentieth-century philosophy of science rebelled, having relativity theory in its support (see Space-Time).

Admitting that the laws were suggested by previous experiences, some interpreters, including Poincaré, considered them as mere conventions (see Conventionalism). Positing absolute Euclidean space, the first law states how inertial matter moves in it. But it is impossible to obtain knowledge about space independent of everything else. Thus, the first law represents merely a criterion for choosing a suitable geometry. Mach ([1883] 1960) rejected absolute space and time as metaphysical. All observable motion was relative, and thus, at least in principle, all material objects in the universe were mutually linked. Mach's principle, as it became called, influenced the early development of general relativity but is still controversial among philosophers (see Barbour and Pfister 1995; Pooley and Brown 2002). According to Mach, the three laws were highly redundant and followed from a proper empiricist explication of Newton's definitions of mass and force. Defining force, with Newton, as an action exerted upon a body to change its state, that is, as an acceleration, the first and second laws can be straightforwardly derived. Mach rejected Newton's definition of mass as quantity of matter as a pseudo-definition and replaced it with the empirical insight that mass is the property of bodies determining acceleration. This was equivalent to the third law. Mach's criticism became an important motivation for Albert Einstein's special theory of relativity. Yet, even the members of the Vienna Circle disagreed whether this influence was formative or Mach merely revealed the internal contradictions of the Newtonian framework (see Vienna Circle).

As to the second law, one may wonder whether the forces are inferred from the observed phenomena of motion or the motions are calculated from given specific forces. Textbooks often interpret the second law as providing a connection from forces to motion, but this is nontrivial and requires a proper superposition of the different forces and a due account of the constraints. The opposite interpretation has the advantage of not facing the notorious problem of characterizing force as an entity in its own right, which requires a distinction between fundamental forces from fictitious or inertial forces.

Book III of Newton's *Principia* introduced the law of universal gravitation, which finally unified the dynamics of the celestial and terrestrial spheres. But, as it stood, it involved an action at a distance through the vacuum, which Newton regarded as the greatest absurdity. To Bentley he wrote: "Gravity must be caused by an Agent acting according to certain Laws; but whether this Agent be material or immaterial, I have left to the Consideration of my readers" (Cohen 1958, 303). Given the law of universal gravity, the peculiar shape of the solar system was a matter of initial conditions, a fact that Newton ascribed to the contrivance of a voluntary agent.

The invention of the calculus was no less important for the development of mechanics than was the law of gravitation. But its use in the *Principia* was not consistent and intertwined with geometrical arguments, and it quickly turned out that Leibniz's version of the calculus was far more elegant. That British scientists remained loyal to Newton's fluxions until the days of Hamilton proved to be a substantial impediment to the use of calculus in science.

Celestial Mechanics and the Apparent Triumph of Determinism

No other field of mechanics witnessed greater triumphs of prediction than celestial mechanics: the return of Halley's comet in 1758; the oblateness of the Earth in the 1740s; and finally the discovery of Neptune in 1846 (cf. Grosser 1962). However, while Neptune was found at the location that the anomalies in the motion of Uranus had suggested, Le Verrier's prediction of a planet Vulcan to account for the anomalous perihelion motion of Mercury failed. Only general relativity would provide a satisfactory explanation of this anomaly (see Space-Time).

But in actual fact little follows from Newton's axioms and the inverse square law of gravitation alone. Only the two-body problem can be solved analytically by reducing it to a one-body problem for relative distance. The three-body problem requires approximation techniques, even if the mass of one body can be neglected. To ensure the convergence of the respective perturbation series became a major mathematical task. In his lunar theory, Clairaut could derive most of the motion of the lunar apsides from the inverse-square law, provided the approximation was carried far enough. But d'Alembert warned that further iterations might fail to converge; hence subsequent analysts calculated to higher and higher orders. D'Alembert's derivation of the precession and nutation of the Earth completed a series of breakthroughs around 1750 that won Newton's law a wide acceptance. In 1785, Laplace explained the remaining chief anomalies in the solar system and provided a (flawed) indirect proof for the stability of the solar system from the conservation of angular momentum (cf. Wilson 1995).

One might draw an inductivist lesson from this history and conceive the increased precision in the core parameters of the solar system, above all the planetary masses, as a measure of explanatory success for the theory containing them (cf. Harper and Smith 1995). Yet, the historically more influential lesson for philosophers consisted in the ideal of Laplace's Demon, the intellect that became the executive officer of strict determinism:

> Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain, and the future, as the past, would be present to its eyes. (Laplace [1795] 1952, 4)

But the Demon is threatened by idleness in many ways. If the force laws are too complex, determinism becomes tautologous; if Newton's equations cannot be integrated, perturbative and statistical strategies are mandatory; calculations could still be too complex; exact knowledge of the initial state of a system presupposes that the precision of measurement could be increased at will.

A further idealization necessary to make the Laplacian ideal thrive was the point mass approach of Boscovich ([1763] 1966). But some problems cannot be solved in this way—for instance, whether a point particle moving in a head-on orbit directed at the origin of a central force is reflected by the singularity or goes right through it. In many cases, celestial mechanics treats planets as extended bodies or gyroscopes rather than point masses. Nonrigid bodies or motion in resistant media require even further departures from the Laplacian ideal. As Wilson put it, "applied mathematicians are often forced to pursue roundabout and shakily rationalized expedients if any progress is to be made" (2000, 296).

Only some of these expedients can be mathematically justified. This poses problems for the relationship of mathematical and physical ontology. Often unsolvable equations are divided into tractable satellite equations. Continuum mechanics is a case in point. The Navier-Stokes equations are virtually intractable by analytic methods; Prandtl's boundary layer theory splits a flow in a pipe into one equation for the fluid's boundary and one for the middle of the flow. Prandtl's theory failed in the case of turbulence, while statistical investigations brought useful results. Accordingly, in this case predictability required abandoning determinism altogether years before the advent of quantum mechanics (cf. von Mises 1922).

From Conserved Quantities to Invariances: Force and Energy

The framework of Newton's laws was not the only conception of mechanics used by 1800 (cf. Grattan-Guinness 1990). There were purely algebraic versions of the calculus based on variational problems developed by Euler and perfected in Lagrange's *Analytical Mechanics* ([1788] 1997). Engineers developed a kind of energy mechanics that emerged from Coulomb's friction studies. Lazare Carnot developed it into an alternative to Lagrange's reduction of dynamics to equilibrium, that is, to statics. Dynamics should come first, and engineers had to deal with many forces that did not admit a potential function.

The nature of energy—primary entity or just inferred quantity—was no less in dispute than that of force. Both were intermingled in content and terminology. The eighteenth century was strongly influenced by the *vis viva* controversy launched by Leibniz. What was the proper quantity in the description of mechanical processes: mv or mv^2 (where $m =$ mass and $v =$ velocity)? After the Leibniz-Clarke correspondence, the issue became associated with atomism and the metaphysical character of conserved quantities. In the nineteenth century, 'energy,' or work, became a universal principle after the discovery of the mechanical equivalent of heat. Helmholtz gave the principle a more general form based upon the mechanical conception of nature. For many scientists this suggested a reduction of the other domains of physical science to mechanics. But this program failed in electrodynamics.

Through the works of Ostwald and Helm, the concept of energy became the center of a movement that spellbound German-speaking academia from the 1880s until the 1900s. But, as Boltzmann untiringly stressed, energeticists obtained the equations of motion only by assuming energy conservation in each spatial direction. But why should Cartesian coordinates have a special meaning?

Two historico-critical studies of the energy concept set the stage for the influential controversy between Planck and Mach (1908–1910). While Mach ([1872] 1911) considered the conservation of work as an empirical specification of the instinctive experience that perpetual motion is impossible, Planck (1887) stressed the independence and universality of the principle of energy conservation. Planck later lauded Mach's positivism as a useful antidote against exaggerated mechanical reductionism but called for a stable world view erected upon unifying variational principles and invariant quantities.

Group theory permitted mathematical physicists to ultimately drop any substantivalist connotation of conserved quantities, such as energy. The main achievement was a theorem of Emmy Noether that identified conserved quantities, or constants of motion, with invariances under one-parameter group transformations (cf. Arnold 1989, 88–90).

Variational Principles and Hamiltonian Mechanics

In 1696–1697 Johann Bernoulli posed the problem of finding the curve of quickest descent between two points in a homogeneous gravitational field. While his own solution used an analogy between geometrical optics and mechanics, the solutions of Leibniz and Jacob Bernoulli assumed that the property of minimality was present in the small as in the large, arriving thus at a differential equation. The title of Euler's classic treatise describes the scope of the variational calculus: *The Method of Finding Curved Lines That Show Some Property of Maximum or Minimum* (1744). The issue of minimality sparked philosophical confusion. In 1746, Pierre Moreau de Maupertuis announced that in all natural processes the quantity $\int v ds = \int v^2 dt$ attains its minimum; he interpreted this as a formal teleological principle that avoided the outgrowths of the earlier physicotheology of Boyle and Bentley. Due to his defective examples and a priority struggle, the principle lost much of its credit. Lagrange conceived of it simply as an effective problem-solving machinery. There was considerable mathematical progress during the nineteenth century, most notably by Hamilton, Gauss, and Jacobi. Only Weierstrass found the first sufficient condition for the variational integral to actually attain its minimum value. In 1900, Hilbert urged mathematicians to systematically develop the variational calculus (see Hilbert, David). He made action principles a core element of his axiomatizations of physics. Hilbert and Planck cherished the principles' applicability, after appropriate specification, to all fields of reversible physics; thus they promised a formal unification instead of the discredited mechanical reductionism.

There exist differential principles, among them d'Alembert's principle and Lagrange's principle of virtual work, that reduce dynamics to statics, and integral action principles that characterize the actual dynamical evolution over a finite time by the stationarity of an integral as compared with other possible evolutions. Taking M_q as the space of all possible motions $q(t)$ between two points in configuration space, the action principle states that the actual motion q extremizes the value of the integral $W[q] = \int_a^b F(t, q(t), \dot{q}(t))dt$ in comparison with all possible motions, the varied curves $(q + \delta q) \in M_q$ (the endpoints of the interval [*a*, *b*] remain fixed).

If one views the philosophical core of action principles in a temporal teleology, insofar as a particle's motion between *a* and *b* is also determined by the fixed state, this contradicts the fact that almost all motions that can be treated by way of action principles are reversible. Yet, already the mathematical conditions for an action principle to be well-defined suggest that the gist of the matter lies in its global and modal features. Among the necessary conditions is the absence of conjugate points between *a* and *b* through which all varied curves have to pass. Sufficient conditions typically involve a specific field embedding of the varied curves; also the continuity properties of the $q \in M_q$ play a role. Thus initial and final conditions have to be understood in the same vein as boundary conditions for partial differential equations that have to be specified beforehand so that the solution cannot be grown stepwise from initial data.

There are also different ways of associating the possible motions. In Hamilton's principle, *F* is the Lagrangian $L = T - V$, where *T* is the kinetic and *V* the potential energy; time is not varied such that all possible motions take equal time but correspond to higher energies than the actual motion. For the original principle of least action, *F* equals *T*, and one obtains the equations of motion only by assuming energy conservation, such that at equal total energy the varied motions take a longer time than the actual ones.

Butterfield (2004) spots three types of modality in the sense of Lewis (1986) here. While all action principles involve a modality of changed initial conditions and changed problems, Hamilton's principle also involves changed laws because the varied curves violate energy conservation. Energeticists and Mach ([1883] 1960) held that *u* is uniquely determined within M_u as compared with the other motions that appear pairwise or with higher degeneracy, but they disagreed whether one could draw conclusions about modal characteristics. Albeit well versed in their mathematical intricacies, logical empiricists treated action principles with neglect in order to prevent an intrusion of metaphysics (Stöltzner 2003).

The second-order Euler-Lagrange equations, which typically correspond to the equations established by way of Newton's laws, can be reformulated as a pair of first-order equations, Hamilton's equations, or a single partial differential equation, *viz.*, the Hamilton-Jacobi equation. Hamilton's equations emerge from the so-called Legendre transformation, which maps configuration space $(q, \dot{q})$ into phase space (q, p), thus transforming the derivative of position $\dot{q}$ into momentum *p*. If this transformation fails, constraints may be present. The core property of Hamilton's equations is their invariance under the so-called canonical transformations. The canonical transformation $(q,p) \rightarrow (Q,P)$ that renders the Hamiltonian $H(Q,P) = T + V$ equal to zero leads to the Hamilton-Jacobi equation. Its generator $W = S - E_t$ (where E_t is the total energy) can be interpreted as an action functional corresponding to moving wave fronts in ordinary space that are orthogonal to the extremals of the variational problem (cf. Lanczos 1986).

The analogy between mechanics and geometric optics is generic and played an important role in Schrödinger's justification of his wave equation. Hamilton-Jacobi theory was also a motivation for Bohm's reformulation of the de Broglie pilot wave theory (see Quantum Mechanics). For periodical motions one can use *S* to generate a canonical transformation that arrives at action and angle variables (J, ω), where $J = \oint pdq$. This integral was quantized in the older Bohr-Sommerfeld quantum theory. The space of all possible motions associated with a variational problem M_q can be considered as an ensemble of trajectories. Arguing that in quantum theory all possible motions are realized with a certain possibility provides some intuitive motivation for the Feynman path integral approach (see Quantum Field Theory).

Variational principles played a central role in several philosophically influential treatises of mechanics in the late nineteenth century. Most radical was that of Hertz ([1894] 1956), who used Gauss's (differential) principle of least constraints to dispense with the notion of force altogether. There were no single mass points but only a system of mass points connected by constraints. Hertz's problem was to obtain a geometry of straight line for this system of mass points. This task was complicated by Hertz's Kantian preference for Euclidean geometry as a basis for the non-Euclidean geometry of mass points. Contemporaries deemed Hertz's geometrization of mechanics a "God's eye view." Boltzmann (1897, 1904) praised its coherence but judged Hertz's picture as inapplicable to problems easily tractable by means of forces.

Hertz held that theoretical pictures had to be logically permissible, empirically correct, and appropriate (that is, sufficiently complete and simple). Boltzmann criticized permissibility as an unwarranted reliance upon a priori laws of thought and held instead that pictures were historically acquired and corroborated by success. Around 1900, treating theories (e.g., atomism) as pictures represented an alternative to mechanical reductionism and positivist descriptivism, until the idea of coordinative definitions between symbolic theory and empirical observations won favor (see Vienna Circle).

Mathematical Mechanics and Classical Dynamical Systems

After classical mechanics was finally dethroned as the governing paradigm of physics, its course became largely mathematical in kind and was strongly influenced by concepts and techniques developed by the new-frontier theories of physics: relativity theory and quantum physics. There was substantial progress in the variational calculus, but the main inspirations came from differential geometry, group theory, and topology, on the one hand, and probability theory and measure theory, on the other. This has led to some rigorous results on the *n*-body problem. The advent of the modern computer not only opened a new era of celestial mechanics, it also revealed that chaotic behavior, ignored by physicists in spite of its early discovery by Poincaré, occurred in a variety of simple mechanical systems.

Geometrization has drastically changed the appearance of classical mechanics. Configuration space and phase space have become the tangent and cotangent bundles on which tensors, vector fields, and differential forms are defined. Coordinates have turned into charts, and the theory of differentiable manifolds studies the relationships between the local level, where everything looks Euclidean, and the global level, where topological obstructions may arise. The dynamics acting on these bundles are expressed in terms of flows defined by vector fields, which gives a precise meaning to variations. The picture of flows continues Hamilton-Jacobi theory. This elucidates that the intricacies of variational calculus do not evaporate; they transform into obstructions of and conditions for the application of the whole geometrical machinery, e.g., how far a local flow can be extended. The constants of the motion define invariant submanifolds that restrict the flow to a manifold of lower dimension; they act like constraints. If a Hamiltonian system has, apart from total energy, enough linearly independent constants of motion and if their Poisson brackets $\left(\frac{\partial F}{\partial q_i}\frac{\partial G}{\partial p_i} - \frac{\partial G}{\partial q_i}\frac{\partial F}{\partial p_i}\right)$, where F and G are such constants, mutually vanish, the system is integrable (the equations of motion can be solved) and the invariant submanifold can be identified with a higher-dimensional torus.

The geometrical structure of Hamiltonian mechanics is kept together by the symplectic form $\omega = dq^i \wedge dp_i$ that is left invariant by canonical transformations. It plays a role analogous to the metric in relativity theory. The skew-symmetric product $\wedge$ of forms and the exterior derivative are the main tools of the Cartan calculus and permit an entirely coordinate-free formulation of dynamics. Many equations thus drastically simplify, but quantitative results still require the choice of a specific chart.

Mathematical progress went hand in hand with a shift of emphasis from quantitative to qualitative results that started with Poincaré's work on the *n*-body problem. Some deep theorems of Hamiltonian mechanics become trivialities in this new language, among them the invariance of phase space volume (Liouville's theorem) and the fact that almost all points in a phase space volume eventually—in fact, infinitely often—return arbitrarily close to their original position (Poincaré recurrence) (see also Statistical Mechanics).

For the small number of mass points characteristic of celestial mechanics, there has been major progress along the lines of the questions asked by Thirring (1992, 6): "Which configurations are stable? Will collisions ever occur? Will particles ever escape to infinity? Will the trajectory always remain in a bounded region of phase space?" In the two-body problem, periodic elliptic motions, head-on collisions, and the hyperbolic and parabolic trajectories leading to infinity are neatly separated by initial data.

For the three-body problem the situation is already complex; there do not exist sufficient constants of motion. Two types of exact solutions were quickly found: The particles remain collinear (Euler) or they remain on the vertices of an equilateral triangle (Lagrange). The equilateral configuration is realized by the Trojan asteroids, Jupiter and the sun. In general, however, even for a negative total energy, two bodies can come close enough to expel the third to infinity. In the four-body problem, there exist even collinear trajectories on which a particle might reach spatial infinity in a finite time (Mather and McGehee 1975). Five bodies, one of them lighter than the others, can even be arranged in such a manner that this happens without any previous collisions because the fifth particle oscillates faster and faster between the two escaping planes, in each of which two particles rotate around one another (Xia 1992; Saari 2005). Both examples blatantly violate special relativity. But rather than hastily resort to arguments of what is 'physical,' the philosopher should follow the mathematical physicist in analyzing the logical structure of classical mechanics, including collisions and other singularities.

Can an integrable system remain stable under perturbation? In the 1960s Kolmogorov, Arnold, and Moser (KAM) gave a very general answer that avails itself of the identification of integrable systems and invariant tori. The KAM theorem shows that sufficiently small perturbations deform only the invariant tori. If the perturbation increases, the tori in resonance with it break first; or more precisely, the more irrational the ratio of the frequency of an invariant torus (in action and angle variables) and the frequency of the perturbation, the more stable is the respective torus. If all invariant tori are broken, the system becomes chaotic (cf. Arnold 1989; Thirring 1992). The KAM theorem also provides the rigorous basis of Kepler's association of planetary orbits and platonic solids. The ratios of the radii of platonic solids to the radii of inscribed platonic solids are irrational numbers of a kind that is badly approximated by rational numbers. And, indeed, among the asteroids between Mars and Jupiter, one finds significant gaps for small ratios of an asteroid's revolution time to that of the perturbing Jupiter.

If all invariant tori have broken up, the only remaining constant of motion is energy, such that the system begins to densely cover the energy shell and becomes ergodic. No wonder that concepts of statistical physics, among them ergodicity, entropy, and mixing, have been used to classify chaotic behavior, even though chaotic behavior does not succumb to statistical laws. They are supplemented by topological concepts, such as the Hausdorff dimension, and concepts of dynamical systems theory, such as bifurcations (nonuniqueness of the time evolution) and attractors (in the case of dissipative systems where energy is no longer conserved).

MICHAEL STÖLTZNER

References

Arnold, Vladimir I (1989), *Mathematical Methods of Classical Mechanics*, 2nd ed. New York and Berlin: Springer.

Barbour, Julian B., and Herbert Pfister (eds.) (1995), *Mach's Principle: From Newton's Bucket to Quantum Gravity*. Boston and Basel: Birkhäuser.

Boltzmann, Ludwig (1897 and 1904), *Vorlesungen über die Principe der Mechanik*. 2 vols. Leipzig: Barth.

Boscovich, Ruder J. ([1763] 1966), *A Theory of Natural Philosophy*. Cambridge, MA: MIT Press.

Butterfield, Jeremy (2004), "Some Aspects of Modality in Analytical Mechanics," in Michael Stöltzner and Paul Weingartner (eds.), *Formale Teleologie und Kausalität*. Paderborn, Germany: Mentis.

Cohen, I. Bernhard (ed.) (1958), *Isaac Newton's Papers and Letters on Natural Philosophy*. Cambridge, MA: Harvard University Press.

Euler, Leonhard (1744), *Methodus Inveniendi Lineas Curvas Maximi Minimive Proprietate Gaudentes sive Solutio Problematis Isoperimetrici Latissimo Sensu Accepti*. Lausanne: Bousquet.

Grattan-Guinness, Ivor (1990), "The Varieties of Mechanics by 1800," *Historia Mathematica* 17: 313–338.

Grosser, Morton (1962), *The Discovery of Neptune*. Cambridge, MA: Harvard University Press.

Harper, William, and George E. Smith (1995), "Newton's New Way of Inquiry," in J. Leplin (ed.), *The Creation of Ideas in Physics: Studies for a Methodology of Theory Construction*. Dordrecht, Netherlands: Kluwer, 113–166.

Hertz, Heinrich ([1894] 1956), *The Principles of Mechanics: Presented in a New Form*. New York: Dover Publications.

Kepler, Johannes ([1596] 1981), *Mysterium cosmographicum*. New York: Abaris.

Kirchhoff, Gustav Robert (1874), *Vorlesungen über analytische Mechanik*. Leipzig: Teubner.

Lagrange, Joseph Louis ([1788] 1997), *Analytical Mechanics*. Dordrecht, Netherlands: Kluwer.

Lanczos, Cornelius (1986), *The Variational Principles of Mechanics*. New York: Dover.

Laplace, Pierre Simon de ([1795] 1952), *A Philosophical Essay on Probabilities*. New York: Dover.

Lewis, David (1986), *On the Plurality of Worlds*. Oxford: Blackwell.

Mach, Ernest ([1872] 1911), *History and Root of the Principle of the Conservation of Energy*. Chicago: Open Court.

——— ([1883] 1960), *The Science of Mechanics: A Critical and Historical Account of Its Development*. La Salle, IL: Open Court. German first edition published by Brockhaus, Leipzig, 1883.

Mather, John, and Richard McGehee, "Solutions of the Collinear Four-Body Problem Which Become Unbounded in Finite Time," in Jürgen Moser (ed.), *Dynamical Systems Theory and Applications*. New York: Springer, 573–587.

Newton, Isaac ([1687, 1713] 1969), *Mathematical Principles of Natural Philosophy*. Translated by Andrew Motte (1729) and revised by Florian Cajori. 2 vols. New York: Greenwood.

Planck, Max (1887), *Das Princip der Erhaltung der Energie*, 2nd ed. Leipzig: B. G. Teubner.

Pooley, Oliver, and Harvey Brown (2002), "Relationalism Rehabilitated? I: Classical Mechanics," *British Journal for the Philosophy of Science* 53: 183–204.

Saari, Donald G. (2005), *Collisions, Rings, and Other Newtonian* N-*Body Problems*. Providence, RI: American Mathematical Society.

Stöltzner, Michael (2003), "The Principle of Least Action as the Logical Empiricists' Shibboleth," *Studies in History and Philosophy of Modern Physics* 34: 285–318.

Thirring, Walter (1992), *A Course in Mathematical Physics 1: Classical Dynamical Systems*, 2nd ed. Translated by Evans M. Harrell. New York and Vienna: Springer.

von Mises, Richard (1922), "Über die gegenwärtige Krise der Mechanik," *Die Naturwissenschaften* 10: 25–29.

Wilson, Curtis (1995), "The Dynamics of the Solar System," in Ivor Grattan-Guinness (ed.), *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*. London: Routledge.

Wilson, Mark (2000), "The Unreasonable Uncooperativeness of Mathematics in the Natural Sciences," *The Monist* 83 (2000): 296–314. See also his very instructive entry "Classical Mechanics" in the *Routledge Encyclopedia of Philosophy*.

Xia, Zhihong (1992), "The Existence of Noncollision Singularities in Newtonian Systems," *Annals of Mathematics* 135: 411–468.

COGNITIVE SCIENCE

Cognitive science is a multidisciplinary approach to the study of cognition and intelligence that emerged in the late 1950s and 1960s. "Core" cognitive science holds that the mind/brain is a kind of computer that processes information in the form of mental representations. The major disciplinary participants in the cognitive science enterprise are psychology, linguistics, neuroscience, computer science, and philosophy. Other fields that are sometimes included are anthropology, education, mathematics, biology, and sociology.

There has been considerable philosophical discussion in recent years that relates, in one way or another, to cognitive science. Arguably, not all of this discussion falls within the tradition of philosophy of science. There are two fundamentally different kinds of question that philosophers of science typically raise about a specific scientific field: "external" questions and "internal" questions. If one assumes that a scientific field provides a framework of inquiry, external questions will be about that framework as a whole, from some external point of view, or about its relation to other scientific frameworks. In contrast, an internal question will be one that is asked *within* the framework, either with respect to some entity or process that constitutes part of the framework's foundations or with respect to some specific theoretical/empirical issue that scientists committed to that framework are addressing. Philosophers have dealt extensively with both external and internal questions associated with cognitive science.

Key External Questions

There are three basic groups of external questions: those concerning the nature of a scientific field *X*, those concerning the relations of *X* to other scientific fields, and those concerning the scientific merits of *X*. An interesting feature of such discussions is that they often draw on, and sometimes even contribute to, the literature on a relevant prior meta-question. For example, discussions concerning the scientific nature of a particular *X* (e.g., cognitive science) may require prior consideration of the question, What is the best way to characterize *X* (in general) scientifically (e.g., in terms of its theories, explanations, paradigm, research program)?

What Is Cognitive Science?

There are many *informal* descriptions of cognitive science in the literature, not based on any serious consideration of the relevant meta-question (e.g., Simon 1981; Gardner 1985). The only systematic formal treatment of cognitive science is that of von Eckardt (1993), although there have been attempts to describe aspects of *cognitive psychology* in terms of Kuhn's notion of a paradigm. Rejecting Kuhn's notion of a paradigm as unsuitable for the characterization of an *immature* field, von Eckardt (1993) proposes, as an alternative, the notion of a *research framework*. A research framework consists of four sets of elements *D*, *Q*, *SA*, *MA*, where *D* is a set of assumptions that provide a pretheoretical specification of the domain under study; *Q* is a set of basic empirical research questions, formulated pretheoretically; *SA* is a set of substantive assumptions that embody the approach being taken in answering the basic questions and that constrain possible answers to those questions; and *MA* is a set of methodological assumptions.

One can think of what Kuhn (1970) calls "normal science" as problem-solving activity. The fundamental problem facing the community of researchers committed to a research framework is to answer each of the basic empirical questions of that framework, subject to the following constraints:

1. Each answer is scientifically acceptable in light of the scientific standards set down by the shared and specific methodological assumptions of the research framework.
2. Each answer is consistent with the substantive assumptions of the research framework.
3. Each answer to a theoretical question makes significant reference to the entities and processes posited by the substantive assumptions of the research framework.

According to von Eckardt (1993), cognitive science consists of a *set* of overlapping research frameworks, each concerned with one or another aspect of human cognitive capacities. Arguably, the *central* research framework concerns the study of adult, normal, typical cognition, with subsidiary frameworks focused on individual differences, group differences (e.g., expert vs. novice, male vs. female), cultural variation, development, pathology, and neural realization. The description offered in von Eckardt (1993), summarized in Table 1, is intended to be of only the central research framework. In addition, it is claimed to be a rational reconstruction as well as transdisciplinary, that is, common to cognitive scientists of all disciplines. Von Eckardt claims that research frameworks can evolve and that the description in question reflects commitments of the cognitive science community at only one period of its history (specifically, the late 1980s/early 1990s).

Cognitive science is a very complex and rapidly changing field. In light of this complexity and change, von Eckardt's original reconstruction should, perhaps, be modified as follows. First, it needs to be acknowledged that there exists a fundamental disagreement within the field as to its domain. The narrow conception, embraced by most psychologists and linguists, is that the domain is *human cognition;* the broad conception, embraced by artificial intelligence researchers, is that the domain is *intelligence in general*. (Philosophers seem to be about evenly split.) A further area of disagreement concerns whether cognitive science encompasses only cognition/intelligence or includes all aspects of mind, including touch, taste, smell, emotion, mood, motivation, personality, and motor skills. One point on which there now seems to be unanimity is that cognitive science must address the phenomenon of consciousness.

Second, a modified characterization of cognitive science must describe it as evolving not only with respect to the computational assumption (from an exclusive focus on symbol systems to the inclusion of connectionist devices) but also with respect to the role of neuroscience. Because cognitive science originally emerged from cognitive psychology, artificial intelligence research, and generative linguistics, in the early years neuroscience was often relegated to a secondary role. Currently, most nonneural cognitive scientists believe that research on the mind/brain should proceed in an interactive way—simultaneously both top-down and bottom-up. Ironically, a dominant emerging view of cognitive neuroscience seems to be that it, rather than cognitive *science*, will be the locus of an interdisciplinary effort to develop, from the bottom up, a computational or information-processing theory of the mind/brain.

There are also research programs currently at the periphery of cognitive science—what might be called alternative cognitive science. These are research programs that investigate some aspect of cognition or intelligence but whose proponents reject one or more of the major guiding or methodological assumptions of mainstream cognitive science. Three such programs are research on situated cognition, artificial life, and the dynamical approach to cognition.

Interfield Relations Within Cognitive Science

The second group of external questions typically asked by philosophers of some particular scientific field concerns the relation of that field to other scientific fields. Because cognitive science is itself multidisciplinary, the most pressing interfield questions arise about the relationship of the various subdisciplines *within* cognitive science. Of the ten possible two-place relations among the five core disciplines of cognitive science, two have received the most attention from philosophers: relations between linguistics and psychology (particularly psycholinguistics) and relations between cognitive psychology and neuroscience.

Discussion of the relation between linguistics and psychology has addressed primarily Noam

Table 1. The central research framework of cognitive science

Domain-specifying assumptions

D1 (Identification assumption): The domain consists of the human cognitive capacities.

D2 (Property assumption): Pretheoretically conceived, the human cognitive capacities have a number of *basic general properties:*

- Each capacity is *intentional*; that is, it involves states that have content or are "about" something.
- Virtually all of the capacities are *pragmatically evaluable;* that is, they can be exercised with varying degrees of success.
- When successfully exercised, each of the evaluable capacities has a certain *coherence* or cogency.
- Most of the evaluable capacities are *reliable*; that is, typically, they are exercised successfully (at least to some degree) rather than unsuccessfully.
- Most of the capacities are *productive*; that is, once a person has the capacity in question, he or she is typically in a position to manifest it in a practically unlimited number of novel ways
- Each capacity involves one or more *conscious* states.[a]

D3 (Grouping assumption): The cognitive capacities of the normal, typical adult make up a theoretically coherent set of phenomena, or a *system*. This means that with sufficient research, it will be possible to arrive at a set of answers to the basic questions of the research framework that constitute a unified theory and are empirically and conceptually acceptable.

Basic questions

Q1: For the normal, typical adult, what precisely is the human capacity to ________?

Q2: In virtue of what does a normal, typical adult have the capacity to _______ (such that this capacity is intentional, pragmatically evaluable, coherent, reliable, and productive and involves consciousness?[a])

Q3: How does a normal, typical adult exercise his or her capacity to _________?

Q4: How does the capacity to __________ of the normal, typical adult interact with the rest of his or her cognitive capacities?

Substantive assumptions

SA1: The computational assumption

C1: (Linking assumption): The human, cognitive mind/brain is a computational device (computer); hence, the human cognitive capacities consist, to a large extent, of a system of computational capacities.

C2: (System assumption): A computer is a device capable of automatically inputting, storing, manipulating, and outputting information in virtue of inputting, storing, manipulating, and outputting representations of that information. These information processes occur in accordance with a finite set of rules that are effective and that are, in some sense, in the machine itself.

SA2: The representational assumption

R1 (Linking assumption): The human, cognitive mind/brain is a representational device; hence, the human cognitive capacities consist of a system of representational capacities.

R2 (System assumption): A representational device is a device that has states or that contains within it entities that are representations. Any representation will have four aspects essential to its being a representation: (1) It will be realized by a representation bearer, (2) it will represent one or more representational objects, (3) its representation relations will be grounded somehow, and (4) it will be interpretable by (will function as a representation *for*) some currently existing interpreter. In the mind/brain representational system, the nature of these four aspects is constrained as follows:

- R2.1 (The representation bearer): The representation bearer of a mental representation is a computational structure or process, considered purely formally.
  - (a) This structure or process has *constituent structure.*[b]
- R2.2 (The representational content): Mental representations have a number of semantic properties.[c]
  - (a) They are semantically *selective.*
  - (b) They are semantically *diverse.*
  - (c) They are semantically *complex.*
  - (d) They are semantically *evaluable.*
  - (e) They have *compositional* semantics.[b]

(*Continued*)

Table 1. (*Continued*)

- R2.3 (The ground): The ground of a mental representation is a property or relation that determines the fact that the representation has the object (or content) it has.
  - a) This ground is *naturalistic* (that is, nonsemantic and nonintentional).
  - b) This ground may consist of either internal or external factors. However, any such factor must satisfy the following restriction: If two token representations have different grounds, and this ground difference determines a difference in content, then they must also differ in their causal powers to produce relevant effects across nomologically possible contexts.[b]
- R2.4 (The interpretant): Mental representations are significant for the person in whose mind they "reside." The interpretant of a mental representation *R* for some subject *S* consists of the set of all possible determinate computational processes contingent upon entertaining *R* in *S*.

Methodological assumptions

M1: Human cognition can be successfully studied by focusing exclusively on the individual cognizer and his or her place in the natural environment. The influence of society or culture on individual cognition can always be explained by appealing to the fact that this influence is mediated through individual perception and representation.
M2: The human cognitive capacities are sufficiently autonomous from other aspects of mind such as affect and personality that, to a large extent, they can be successfully studied in isolation.
M3: There exists a partitioning of cognition in general into individual cognitive capacities such that each of these individual capacities can, to a large extent, be successfully studied in isolation from each of the others.
M4: Although there is considerable variation in how adult human beings exercise their cognitive capacities, it is meaningful to distinguish, at least roughly, between "normal" and "abnormal" cognition.
M5: Although there is considerable variation in how adult human beings exercise their cognitive capacities, adults are sufficiently alike when they cognize that it is meaningful to talk about a "typical" adult cognizer and it is possible to arrive at generalizations about cognition that hold (at least approximately) for all normal adults.
M6: The explanatory strategy of cognitive science is sound. In particular, answers to the original basic questions can, to a large extent, be obtained by answering their narrow information-processing counterparts (that is, those involving processes purely "in the head").
M7: In choosing among alternative hypothesized answers to the basic questions of the research framework, one should invoke the usual canons of scientific methodolology. That is, ultimately, answers to the basic questions should be justified on empirical grounds.
M8: A complete theory of human cognition will not be possible without a substantial contribution from each of the subdisciplines of cognitive science.
M9: Information-processing answers to the basic questions of cognitive science are constrained by the findings of neuroscience.[c]
M10: The optimal research strategy for developing an adequate theory of the cognitive mind/brain is to adopt a coevolutionary approach—that is, to develop information-processing answers to the basic questions of cognitive science on the basis of empirical findings from both the nonneural cognitive sciences and the neurosciences.[c]
M11: Information-processing theories of the cognitive mind/brain can explain certain features of cognition that cannot be explained by means of lower-level neuroscientific accounts. Such theories are thus, in principle, explanatorily ineliminable.[c]

Key: [a]Not in von Eckardt (1993); [b]Controversial; [c]Included on normative grounds (given the commitment of cognitive science to *X*, cognitive scientists should be committed to *Y*)

Chomsky's claim that the aims of linguistics are, first, to develop hypotheses (in the form of generative grammars) about what the native speaker of a language *knows* (tacitly) or that capture the native speaker's linguistic *competence* and, second, to develop a theory of the innate constraints on language ("universal grammar") that the child brings to bear in learning any given language. Philosophers have focused on the relation of competence models to so-called performance models, that is, models developed by psychologists to describe the information processes involved in understanding and producing language. Earlier discussion focused on syntax; more recent debates have looked at semantics.

There are several views. One, favored by Chomsky himself, is that a representation of the grammar of a language *L* constitutes a *part* of the mental apparatus involved in the psychological processes underlying language performance and, hence, is causally implicated in the production of that performance (Chomsky and Katz 1974). A second, suggested by Marr (1982) and others, is that competence theories describe a native speaker's linguistic input-output *capacity* (Marr's "computational" level) without describing the processes underlying that capacity (Marr's "algorithmic" level). For example, the capacity to understand a sentence *S* can be viewed as the capacity that maps a phonological representation of *S* onto a semantic representation *S*. Both views have been criticized. Against the first, it has been argued that because of the linguist's concern with simplicity and generality, it is unlikely that the formal structures utilized by optimal linguistic theories will be isomorphic to the internal representations posited by psycholinguists (Soames 1984). A complementary point is that because processing models are sensitive to architectural and resource constraints, the ways in which they implement syntactic knowledge have turned out to be much less transparent than followers of Chomsky had hoped (Mathews 1991). At best, they permit a psycholinguistic explanation of why a particular set of syntactic rules and principles correctly characterize linguistic competence. Against the second capacity view, it has been argued that grammars constitute idealizations (for example, there are no limitations on the length of permissible sentences or on their degree of internal complexity) and so do not describe actual speakers' linguistic capacities (Franks 1995; see Philosophy of Psychology; Neurobiology).

Internal questions have been raised about both foundational assumptions and concepts and particular theories and findings within cognitive science. Foundational discussions have focused on the core substantive assumptions: (1) The cognitive mind/brain is a computational system and (2) it is a representational system.

The Computational Assumption

Cognitive science assumes that the mind/brain is a kind of computer. But what *is* a computer? To date, two *kinds* of computer have been important to cognitive science modeling—*classical* machines ("conventional," "von Neumann," or "symbol system") and *connectionist* machines ("parallel distributed processing"). Classical machines have an architecture similar to ordinary personal computers (PCs). There are separate components for processing and memory, and processing occurs by the manipulation of data structures. In contrast, connectionist computers are more like brains, consisting of interconnected networks of neuron-like elements. Philosophers interested in the computational assumption have focused on two questions: What is a computer *in general* (such that both classical and connectionist machines count as computers), and is there any reason to believe, at the current stage of research, that human mind/brains are one sort of computer rather than another? The view that the human mind/brain is, or is importantly like, a classical computer is called classicism; the view that it is, or is importantly like, a connectionist computer is called connectionism.

The theory of computation in mathematics defines a number of different kinds of *abstract* machines in terms of the sets of functions they can compute. Of these, the most relevant to cognitive science is the universal Turing machine, which, according to the Church-Turing thesis, can compute any function that can be computed by an *effective* method, that is, a method specified by a finite number of exact instructions, in a finite number of steps, carried out by a human unaided by any machinery except paper and pencil and demanding no insight or ingenuity. Many philosophers and cognitive scientists think that the notion of a computer relevant to cognitive science can be adequately captured by the notion of a Turing machine. In fact, that is not the case. To say that a computer (and hence the mind/brain) is simply a device *equivalent* to a universal Turing machine says nothing about the device's internal structure, and to say that a computer (and hence the mind/brain) is an *implementation* of a Turing machine seems flat-out false. There are many dissimilarities. A Turing machine is in only one state at a time, while humans are, typically, in many mental or brain states simultaneously. Further, the memory of a Turing machine is a "tape" with only a simple linear structure, while human memory appears to have a complex multidimensional structure.

An alternative approach is to provide an architectural characterization, but at a fairly high level of abstraction. For example, von Eckardt (1993, 114) claims that cognitive science's computational assumption takes a computer to be "a device capable of automatically inputting, storing, manipulating, and outputting information in virtue of inputting, storing, manipulating, and outputting representations of that information," where "[t]hese information processes occur in accordance with a finite set of rules that are effective and that are, in some sense, in the machine itself." Another proposal, presented by Copeland (1996) and specifically intended to include functions that a Turing machine cannot compute, is that a computer is a device capable of solving a class of problems, that can represent both the problems and their solution, contains some number of primitive operations (some of which may not be Turing computable), can sequence these operations in some predetermined way, and has a provision for feedback (see Turing, Alan).

To the question of whether the mind/brain is a connectionist versus a classical computer, discussion has centered around Fodor and Pylyshyn's (1988) argument in favor of classicism. In their view, the claim that the mind/brain is a connectionist computer should be rejected on the grounds of the premises that

1. cognitive capacities exhibit *systematicity* (where 'systematicity' is the fact that some capacities are intrinsically connected to others—e.g., if a native speaker of English knows how to say "John loves the girl," the speaker will know how to say "The girl loves John") and
2. this feature of systematicity cannot be explained by reference to connectionist models (unless they are implementations of classical models), whereas
3. it can be explained by reference to classical models.

Fodor and Pylyshyn's reasoning is that it is "characteristic" of classical systems but not of connectionist systems to be "symbol processors," that is, systems that posit mental representations with constituent structure and then process these representations in a way that is sensitive to that structure. Given such features, classical models can explain systematicity; but without such features, as in connectionist machines, it is a mystery. Fodor and Pylyshyn's challenge has generated a host of responses, from both philosophers and computer scientists. In addition, the argument itself has evolved in two significant ways. First, it has been emphasized that to be a counterexample to premise 2, a connectionist model must not only exhibit systematicity, it must also *explain* it. Second, it has been claimed that what needs explaining is not just that human cognitive capacities are systematic; it is that they are *necessarily* so (on the basis of scientific law).

The critical points regarding premise 3 are especially important. The force of Fodor and Pylyshyn's argument rests on classical systems, being able to do something that connectionist systems (at a cognitive, nonimplementational level) cannot. However, it has been pointed out that the mere fact that a system is a symbol processer (and hence classical) does not *by itself* explain systematicity, much less the lawful necessity of it; additional assumptions must be made about the system's computational resources. Thus, if the critics are right that connectionist systems of *specific* sorts can also explain systematicity, then the explanatory asymmetry between classicism and connectionism will no longer hold (that is, neither the fact that a system is classical nor the fact that it is connectionist can, taken by itself, explain systematicity, while either fact can explain systematicity when appropriately supplemented), and Fodor and Pylyshyn's argument would be unsound.

The Representational Assumption

Following Peirce (Hartshorne, Weiss, and Burks 1931–58), one can say that any representation has four aspects essential to its being a representation: (i) It is realized by a representation bearer; (ii) it has *semantic content* of some sort; (iii) its semantic content is *grounded* somehow; and (iv) it has *significance* for (that is, it can function as a representation for) the person who has it. (Peirce's terminology was somewhat different. He spoke of a representation's bearer as the "material qualities" of the representation and the content as the representation's "object.")

In Peirce, a representation has significance for a person insofar as it produces a certain effect or "interpretant." This conception of representation is extremely useful for exploring the view of cognitive science vis-à-vis mental representation, for it leads one to ask: What is the representation bearer, semantic content, and ground of *mental* representation, and how do mental representations have significance for the people in whose mind/brains they reside?

A representation bearer is an entity or state that has semantic properties considered with respect to its nonrepresentational properties. The representation bearers for mental representations, according to cognitive science, are computational structures or states. If the mind/brain is a classical computer, its representation bearers are data structures; if it is a connectionist computer, its *explicit* representation bearers are patterns of activation of nodes in a network. It is also claimed that connectionist computers *implicitly* represent by means of their connection weights.

Much of the theoretical work of cognitive psychologists consists of claims regarding the content of the representations used in exercising one or another cognitive capacity. For example, psycholinguists posit that in understanding a sentence, people unconsciously form representations of the sentence's phonological structure, words, syntactic structure, and meaning. Although psychologists do not know enough about mental representation as a system to theorize about its semantics in the sense in which linguistics provides the formal semantics of natural language, if one reflects on what it is that mental representations are hypothesized to explain—certain features of cognitive capacities—one can plausibly infer that the semantics of human mental representation systems must have certain characteristics. In particular, a case can be made that people must be able to represent (i) specific objects; (ii) many different kinds of objects including concrete objects, sets, properties, events and states of affairs in the world, in possible worlds, and in fictional worlds, as well as abstract objects such as universals and numbers; (iii) both objects (*tout court*) and aspects of those objects (or something like extension and intension); and (iv) both correctly and incorrectly.

Although cognitive psychologists have concerned themselves primarily with the representation bearers and semantic content of mental representations, the hope is that eventually there will be an account of how the computational states and structures that function as representation bearers come by their content. In virtue of what, for example, do lexical representations represent specific words? What makes an edge detector represent edges? Such theories are sometimes described as one or another form of "semantics" (e.g., informational semantics, functional role semantics), but this facilitates a confusion between theories of content and theories of what determines that content. A preferable term is *theory of content determination*. Philosophers have been concerned both (a) to delineate the basic relation between a representation's having a certain content and its having a certain ground and (b) to sketch alternative theories of content determination, that is, theories of what that ground might be. A common view is that the basic relation is strong supervenience, as defined by Kim (1984). Others, such as Poland (1994), believe that a stronger relation of *realization* is required.

How is the content of mental representations determined? Proposals appeal either exclusively or in some combination to the structure of the representation bearer (Palmer 1978); actual historical (Devitt 1981) or counterfactual causal relations (Fodor 1987) between the representation bearer and phenomena in the world; actual and counterfactual (causal, computational, inferential) relations between the representation-bearer state and other states in the mind/brain (Harman 1987; Block 1987); or the information-carrying or other *functions* of the representation-bearer state and associated components (based on what they were selected for in evolution or learning) (Millikan 1984; Papineau 1987).

Arguably, any adequate theory of content determination will be able to account for the full range of semantic properties of mental representational systems. On this criterion, all current theories of content determination are inadequate. For example, theories that ground representational content in an isomorphism between some aspect of a representation (usually, either its formal structure or its functional role) and what it represents do not seem to be able to explain how one can represent *specific* objects, such as a favorite coffee mug. In contrast, theories that rely on actual, historical causal relations can easily explain the representation of specific objects but do not seem to be able to explain the representation of sets or kinds of objects (e.g., all coffee mugs). It is precisely because single-factor theories of content determination do not seem to have the resources to explain all aspects of people's representational capacities that many philosophers have turned to two-factor theories, such as theories that combine causal relations with function (so-called teleofunctional theories) and theories that combine causal relations with functional role (so-called two-factor theories).

The fourth aspect of a mental representation, in the Peircian view, is that the content of a representation must have significance *for* the representer. In the information-processing paradigm, this amounts to the fact that for each representation, there will be a set of computational consequences of that representation being "entertained" or "activated," and in particular a set of computational consequences that are *appropriate* given the content of the representation (von Eckardt 1993, Ch. 8).

The Viability of Cognitive Science

Cognitive science has been criticized on several grounds. It has been claimed that there are phenomena within its domain that it does not have the conceptual resources to explain; that one or more of its foundational assumptions are problematic and, hence, that the research program grounded in those assumptions can never succeed; and that it can, in principle, be eliminated in favor of pure neuroscience.

The list of mental phenomena that, according to critics, cognitive science will never be able to explain include that of people "making sense" in their actions and speech, of their having sensations, emotions, and moods, a self and a sense of self, consciousness, and the capabilities of insight and creativity and of interacting closely and directly with their physical and social environments. Although no impossibility proofs have been offered, when cognitive science consisted simply of a top-down, "symbolic" computational approach, this explanatory challenge had a fair amount of intuitive plausibility. However, given the increasing importance of neuroscience within cognitive science and the continuing evolution of the field, it is much less clear today that no cognitive science explanation of such phenomena will be forthcoming. The biggest challenge seems to be the "explanatory gap" posed by phenomenal consciousness (Levine 1983) (see Consciousness).

The major challenge against the computational assumption is the claim that the notion of a computer employed within cognitive science is vacuous. Specifically, Putnam (1988) has offered a proof that every ordinary open system realizes every abstract finite automaton, and Searle (1990) has claimed that implementation is not an objective relation and, hence, that any given system can be seen to implement any computation if interpreted appropriately. Both claims have been disputed. For example, it has been pointed out that Putnam's result relies on employing both an inappropriate computational formalism (the formalism of finite state automata, which have only *monadic* states), an inappropriately weak notion of implementation (one that does not require the mapping from computational to physical states to satisfy *counterfactual* or lawful state–state transitions), and an inappropriately permissive notion of a physical state (one that allows for rigged disjunctions). When these parameters of the problem are made more restrictive, implementations are much harder to come by (Chalmers 1996). However, in defense of Putnam, it has been suggested that this response does not get to the heart of Putnam's challenge, which is to develop a theory that provides necessary and sufficient criteria to determine whether a class of computations is implemented by a class of physical systems (described at a given level). An alternative approach to implementation might be based on the notion of realization of a (mathematical) function by a "digital system" (Scheutz 1999).

The major challenge to the representational assumption is the claim that the project of finding an adequate theory of content determination is doomed to failure. One view is that the attempt by cognitive science to explain the intentionality of cognition by positing mental representations is fundamentally confused (Horst 1996). The argument is basically this: According to cognitive science, a mental state, as ordinarily construed, has some content property in virtue of the fact that it is identical to (supervenes on) a representational state that has some associated semantic property. There are four ways of making sense of this posited semantic property. The semantic property in question is identical to one of the following options:

1. The original content property,
2. An interpreter-dependent semiotic-semantic property,
3. A pure semiotic-semantic property of the sort posited in linguistics, or
4. Some new theoretical "naturalized" property.

However, according to Horst (1996), none of these options will work. Options 1 and 2 are circular and hence uninformative. Option 3 is ruled out on the grounds that there is no reason to believe that there are such properties. And option 4 is ruled out on the grounds that naturalized theories of content determination do not have the conceptual resources to deliver the kind of explanation required.

In response, it has been argued that Horst's case against option 3 is inadequate and that his argument against option 4 shows a misunderstanding of what theories of content determination are trying to achieve (von Eckardt 2001). Horst's reason for ruling out option 3 is that he thinks that people who believe in the existence of pure semantic properties or relations do so because there are formal semantic theories that posit such properties and such theories have met with a certain degree of success. But (he argues) when scientists develop models (theories) in science, they do so by a process of abstraction from the phenomena *in vivo* that they wish to characterize. Such abstractions can be viewed either as models *of* the real-world phenomena they were abstracted from or as descriptions of the mathematical properties of those phenomena. Neither view provides a license for new ontological claims. Thus, Horst's argument rests on a nonstandard conception of both theory construction (as nothing but abstraction) and models or theories (as purely mathematical). If, *contra* Horst, a more standard conception is substituted, then linguists have as much right to posit pure semantic properties as physicists have to posit strange subatomic particles. Furthermore, the existence of the posited entities will depend completely on how successful they are epistemically.

Horst's case against option 4, again, exhibits a misunderstanding of the cognitive science project. He assumes that the naturalistic ground N of the content of a mental representation C will be such that the truth of the statement "X is N" conjoined with the necessary truths of logic and mathematics will be *logically sufficient* for the truth of the statement "X has C" and that, further, this entailment will be epistemically transparent. He then argues against there being good prospects for a naturalistic theory of content on the grounds that naturalistic discourse does not have the conceptual resources to build a naturalistic theory that will entail, in an epistemically transparent way, the truths about intentionality. However, as von Eckardt (2001) points out, Horst's conception of naturalization is much stronger than what most current theory of content determination theorists have in mind, *viz.*, strong supervenience or realization (see Supervenience). As a consequence, his arguments that naturalization is implausible given the conceptual resources of naturalistic discourse are seriously misguided.

Barbara von Eckardt

References

Block, N. (1987), "Functional Role and Truth Conditions," *Proceedings of the Aristotelian Society* 61: 157–181.

Chalmers, D. (1996), "Does a Rock Implement Every Finite-State Automaton?" *Synthese* 108: 309–333.

Chomsky, N., and J. Katz (1974), "What the Linguist Is Talking About," *Journal of Philosophy* 71: 347–367.

Copeland, B. (1996), "What Is Computation?" *Synthese* 108: 336–359.

Devitt, M. (1981), *Designation*. New York: Columbia University Press.

Fodor, J. (1987), *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J., and Z. W. Pylyshyn (1988), "Connectionism and Cognitive Architecture: A Critical Analysis," in S. Pinker and J. Miller (eds.), *Connections and Symbols*. Cambridge, MA: MIT Press, 1–71.

Franks, B. (1995), "On Explanation in the Cognitive Sciences: Competence, Idealization and the Failure of the Classical Cascade," *British Journal for the Philosophy of Science* 46: 475–502.

Gardner, H. (1985), *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books.

Harman, G. (1987), "(Non–Solipsistic) Conceptual Role Semantics," in E. Lepore (ed.), *New Directions in Semantics*. London: Academic Press.

Hartshorne, C., P. Weiss, and A. Burks (eds.) (1931–58), *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Horst, S. W. (1996), *Symbols, Computation, and Intentionality*. Berkeley and Los Angeles: University of California Press.

Kim, J. (1984), "Concepts of Supervenience," *Philosophical and Phenomenological Research* 45: 153–176.

Kuhn, T. (1970), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Levine, J. (1983), "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64: 354–361.

Marr, D. (1982), *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.

Mathews, R. (1991), "Psychological Reality of Grammars," in A. Kasher (ed.), *The Chomskyan Turn*. Oxford and Cambridge, MA: Basil Blackwell.

Millikan, R. (1984), *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Palmer, S. (1978), "Fundamental Aspects of Cognitive Representation," in E. Rosch and B. Lloyd (eds.), *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.

Papineau, D. (1987), *Reality and Representation*. Oxford: Blackwell.

Poland, J. (1994), *Physicalism: The Philosophical Foundations*. Oxford: Oxford University Press.

Putnam, H. (1988), *Representation and Reality*. Cambridge, MA: MIT Press.

Scheutz, M. (1999), "When Physical Systems Realize Functions," *Minds and Machines* 9: 161–196.

Searle, J. (1990), "Is the Brain a Digital Computer?" *Proceedings and Addresses of the American Philosophical Association* 64: 21–37.

Simon, H. (1981), "Cognitive Science: The Newest Science of the Artificial," in D. A. Norman (ed.), *Perspectives on Cognitive Science*. Norwood, NJ: Ablex Publishing, 13–26.

Soames, S. (1984), "Linguistics and Psychology," *Linguistics and Philosophy* 7: 155–179.

von Eckardt, B. (2001), "In Defense of Mental Representation," in P. Gardenfors, K. Kijania–Placek, and J. Wolenski (eds.), *Proceedings of the 11th International Congress of Logic, Methodology and Philosophy of Science*. Dordrecht, Netherlands: Kluwer.

——— (1993), *What Is Cognitive Science?* Cambridge, MA: MIT Press.

See also **Consciousness; Physicalism; Psychology, Philosophy of; Supervenience**

COGNITIVE SIGNIFICANCE

One of the main objectives of logical empiricism was to develop a formal criterion by which cognitively significant statements, which are true or false, could be delineated from meaningless ones, which are neither. The desired criterion would specify, and in some way justify, the logical empiricists' conviction that scientific statements were exemplars of significance and metaphysical ones were decidedly not (see Logical Empiricism). Finding such a criterion was crucial to logical empiricism. Without it there seemed to be no defensible way to distinguish metaphysics from science and, consequently, no defensible way to exclude metaphysics from subjects that deserved serious philosophical attention (see Demarcation, Problem of). Accordingly, several logical empiricists devoted attention to developing a criterion of cognitive significance, including Carnap, Schlick, Ayer, Hempel, and, to a lesser degree, Reichenbach.

Scientific developments also motivated the project in two related ways. First, physics and biology were demonstrating that a priori metaphysical speculations about empirical matters were usually erroneous and methodologically misguided. Hans Driesch's idea of an essential entelechy was no longer considered scientifically respectable, and the intuitive appeal of the concept of absolute simultaneity was shown to be misleading by Albert Einstein (Feigl 1969). Scientific results demonstrated both the necessity and the fruitfulness of replacing intuitive convictions with precise, empirically testable hypotheses, and logical empiricists thought the same methodology should be applied to philosophy. Formulating a defensible criterion that ensured the privileged epistemological status of science, and revealed the vacuity of metaphysics, was thought crucial to the progress and respectability of philosophy.

Second, many scientific discoveries and emerging research programs, especially in theoretical physics, were considerably removed from everyday observable experience and involved abstract, mathematically sophisticated theories. The logical empiricists felt there was a need for a formal systematization of science that could clarify theoretical concepts, their interrelations, and their connection with observation. The emerging tools of modern mathematical logic made this task seem imminently attainable. With the desire for clarity came the pursuit of a criterion that could sharply distinguish these scientific developments, which provided insights about the world and constituted advances in knowledge, from the obfuscations of metaphysics.

Formulation of a cognitive significance criterion requires an empirical significance criterion to delineate empirical from nonempirical statements and a criterion of analyticity to delineate analytic from synthetic statements (see Analyticity). Most logical empiricists thought analytically true and false statements were meaningful, and most metaphysicians thought their claims were true but not analytically so. In their search for a cognitive significance criterion, as the principal weapon of their antimetaphysical agenda, the logical empiricists focused on empirical significance.

The Verifiability Requirement

The first attempts to develop the antimetaphysical ideas of the logical empiricists into a more rigorous criterion of meaningfulness were based on the verifiability theory of meaning (see Verifiability). Though of auxiliary importance to the rational reconstruction in the *Aufbau*, Carnap (1928a) claimed that a statement was verifiable and thereby meaningful if and only if it could be translated into a constructional system; for instance, by reducing it (at least in principle) to a system about basic physical objects or elementary experiences (§179) (see Carnap, Rudolf). Meaningful questions have verifiable answers; questions that fail this requirement are pseudo-questions devoid of cognitive content (§180).

The first explicit, semiformal criterion originated with Carnap in 1928. With the intention of demonstrating that the realism/idealism debate, and many other philosophical controversies, were devoid of cognitive significance, Carnap (1928b) presented a criterion of factual content:

> If a statement *p* expresses the content of an experience *E*, and if the statement *q* is either the same as *p* or can be derived from *p* and prior experiences, either through deductive or inductive arguments, then we say that *q* is 'supported by' the experience *E*. . . . A statement *p* is said to have 'factual content' if experiences which would support *p* or the contradictory of *p* are at least conceivable, and if their characteristics can be indicated. (Carnap 1967, 327)

Only statements with factual content are empirically meaningful. Notice that a fairly precise inferential method is specified and that a statement has factual content if there are conceivable experiments that could support it. Thus, the earliest formal significance criterion already emphasized that possible, not necessarily actual, connection to experience made statements meaningful.

Carnap ([1932] 1959) made three significant changes to his proposal. First, building on an earlier example (1928b, §7), he developed in more detail the role of syntax in determining the meaningfulness of words and statements in natural languages. The "elementary" sentence form for a word is the simplest in which it can occur. For Carnap, a word had to have a fixed mode of occurrence in its elementary sentence form to be significant. Besides failing to connect with experience in some way, statements could also be meaningless because they contained sequences of words that violated the language's syntactic rules, or its "logical syntax." According to Carnap, "Dog boat in of" is meaningless because it violates grammatical syntax, and "Our president is definitely a finite ordinal," is meaningless because it violates logical syntax, 'president' and 'finite ordinal' being members of different logical categories. The focus on syntax led Carnap to contextualize claims of significance to specific languages. Two languages that differ in syntax differ in whether words and word sequences are meaningful.

Second, Carnap ([1932] 1959) no longer required statements to be meaningful by expressing conceivable states of affairs. Rather, statements are meaningful because they exhibit appropriate deducibility relations with protocol statements whose significance was taken as primitive and incorrigible by Carnap at that time (see Protocol Sentences). Third, Carnap did not specify exactly how significant statements must connect to protocol statements, as he had earlier (1928b). In 1932, Carnap would ascertain a word's meaning by considering the elementary sentence in which it occurred and determining what statements entailed and were entailed by it, the truth conditions of the statement, or how it was verified—considerations Carnap then thought were equivalent. The relations were probably left unspecified because Carnap came to appreciate how difficult it was to formalize the significance criterion, and realized that his earlier criterion was seriously flawed, as was shown of Ayer's first formal criterion (see below).

In contrast to antimetaphysical positions that evaluated metaphysical statements as false, Carnap believed his criterion justified a radical elimination of metaphysics as a vacuous enterprise. The defensibility of this claim depended upon the status of the criterion—whether it was an empirical hypothesis that had to be supported by evidence or a definition that had to be justified on other grounds. Carnap ([1932] 1959) did not address this issue, though he labels the criterion as a stipulation. Whether this stipulation was defensible in relation to other possible criteria or whether the statement of the criterion satisfied the criterion itself were questions left unanswered.

In his popularization of the work of Carnap ([1932] 1959) and Schlick ([1932] 1979), Ayer (1934) addressed these questions and stated that a significance criterion should not be taken as an empirical claim about the linguistic habits of the class of people who use the word 'meaning' (see Ayer, Alfred Jules; Schlick, Moritz). Rather, it is a different kind of empirical proposition, which, though conventional, has to satisfy an adequacy condition. The criterion is empirical because, to be adequate, it must classify "propositions which by universal agreement are given as significant" as significant, and propositions that are universally agreed to be nonsignificant as nonsignificant (Ayer 1934, 345).

Ayer developed two formalizations of the criterion. The first edition of *Language, Truth, and Logic* contained the proposal that "a statement is verifiable ... if some observation-statement can be deduced from it in conjunction with certain other premises, without being deducible from those other premises alone," where an observation statement is one that records any actual or possible observation (Ayer 1946, 11).

Following criticisms (see the following section) of his earlier work, a decade later Ayer (1946) proposed a more sophisticated criterion by distinguishing between directly verifiable statements and indirectly verifiable ones. In conjunction with a set of observation statements, *directly* verifiable statements entail at least one observation statement that does not follow from the set alone. *Indirectly* verifiable statements satisfy two requirements: (1) In conjunction with a set of premises, they entail at least one directly verifiable statement that does not follow from the set alone; and (2) the premises can include only statements that are either analytic or directly verifiable or can be indirectly verified on independent grounds. Nonanalytic statements that are directly or indirectly verifiable are meaningful, whereas analytic statements are meaningful but do not assert anything about the world.

Early Criticisms of the Verifiability Criterion

The verifiability criterion faced several criticisms, which took two general forms. The first, already mentioned in the last section, questioned its status—specifically, whether the statement of the criterion satisfies the criterion. The second questioned its adequacy: Does the criterion ensure that obviously meaningful statements, especially scientific ones, are labeled as meaningful and that obviously meaningless statements are labeled as meaningless?

Criticisms of the first form often mistook the point of the criterion, construing it as a simple empirical hypothesis about how the concept of meaning is understood or a dogmatic stipulation about how it should be understood (Stace 1935). As mentioned earlier, Ayer (1934, 1946) clearly recognized that it was not this type of empirical claim, nor was it an arbitrary definition. Rather, as Hempel (1950) later made clear, the criterion was intended to clarify and explicate the idea of a meaningful statement. As an explication it must accord with intuitions about the meaningfulness of common statements and suggest a framework for understanding how theoretical terms of science are significant (see Explication). The metaphysician can deny the adequacy of this explication but must then develop a more liberal criterion that classifies metaphysical claims as significant while evaluating clearly meaningless assertions as meaningless (Ayer 1934).

Criticisms of the second form often involved misinterpretations of the details of the criterion, due partially to the ambiguity of what was meant by 'verifiability.' For example, in a criticism of Ayer (1934), Stace (1935) argued that the verifiability criterion made all statements about the past meaningless, since it was in principle impossible to access the past and therefore verify them. His argument involved two misconceptions. First, Stace construed the criterion to require the possibility of conclusive verification, for instance a complete reduction of any statement to (possible) observations that could be directly verified. Ayer (1934) did not address this issue, but Schlick ([1932] 1979), from whose work Ayer drew substantially, emphasized that many meaningful propositions, such as those concerning physical objects, could never be verified conclusively. Accepting Neurath's criticisms in the early 1930s, Carnap accepted that no statement, including no protocol statement, was conclusively verified (see Neurath, Otto). Recall also that Carnap (1928b) classified statements that were "supported by" conceivable experiences—not conclusively verified—as meaningful.

Second, Stace's argument depended on the ambiguity of "possible verification," which made early formulations of the criterion misleadingly unclear (Lewis 1934). The possibility of verification can have three senses: practical possibility, empirical possibility, and logical possibility. Practical possibility was not the intended sense: "There are 10,000-foot mountains on the moon's far side" was meaningful in the 1930s, though its verification was practically impossible (Schlick [1932] 1979).

However, Carnap (1928a, 1928b, [1932] 1959), Schlick ([1932] 1979), and Ayer (1934) were silent on whether empirical or logical possibility divided the verifiable from the unverifiable. Stace thought time travel was empirically impossible. The question was therefore whether statements about past events were meaningful for which no present evidence was available, and no future evidence would be.

In the first detailed analysis of the verifiability criterion, Schlick ([1936] 1979) stated that the logical impossibility of verification renders a statement nonsignificant. Empirical impossibility, which Schlick understood as contradicting the "laws of nature," does not entail non-verifiability. If it did, Schlick argued, the meaningfulness of a putative statement could be established only by empirical inquiry about the laws of nature. For Schlick, this conflated a statement's meaning with its truth. The meaning of a statement is determined ("bestowed") by logical syntax, and only with meaning fixed a priori can its truth or falsity be assessed. Furthermore, since some lawlike generalizations are yet to be identified and lawlike generalizations are never established with absolute certainty, it seems that a sharp boundary between the empirically impossible and possible could never be determined. Hence, there would be no sharp distinction between the verifiable and unverifiable, which Schlick found unacceptable.

For Schlick ([1936] 1979), questions formulated according to the rules of logical grammar are meaningful if and only if it is logically possible to verify their answers. A state of affairs is logically possible for Schlick if the statement that describes it conforms to the logical grammar of language. Hence, meaningful questions may concern states of affairs that contradict well-supported lawlike generalizations. Schlick's position implies that the set of meaningful questions is an extension of the set of questions for which verifiable answers can be imagined. Questions about velocities greater than light are meaningful according to Schlick, but imagining how they could be verified surpasses our mental capabilities.

Schlick's emphasis on logical possibility was problematic because it was unclear that the verification conditions of most metaphysical statements are, or entail, logical impossibilities. In contrast, Carnap ([1936–1937] 1965) and Reichenbach (1938) claimed that metaphysical statements were nonsignificant because no *empirically* possible process of confirmation could be specified for them (see Reichenbach, Hans). Furthermore, if only the *logical* possibility of verification were required for significance, then the nonsignificance of metaphysical statements could no longer be demonstrated by demanding an elucidation of the circumstances in which they could be verified. Metaphysicians can legitimately respond that such circumstances may be difficult or impossible to conceive because they are not empirically possible. Nevertheless, the circumstances may be logically possible, and hence the metaphysical statements may be significant according to Schlick's position.

Faced with the problematic vagueness of the early criteria, a formal specification of the criterion was thought to be crucial. Berlin (1939) pointed out that the early verifiability criteria were open to objections from metaphysicians because the details of the experiential relevance required of meaningful statements were left unclear: "Relevance is not a precise logical category, and fantastic metaphysical systems may choose to claim that observation data are 'relevant' to their truth" (233).

With formalizations of the criterion, however, came more definitive criticisms. Ayer's (1946, 39) first proposal was seriously flawed because it seemed to make almost all statements verifiable. For any grammatical statement S—for instance "The Absolute is peevish"—any observation statement O, and the conditional $S \rightarrow O$, S and $S \rightarrow O$ jointly entail O, though neither of them alone usually does. According to Ayer's criterion, therefore, S and $S \rightarrow O$ are meaningful except in the rare case that $S \rightarrow O$ entails O (Berlin 1939).

Church (1949) presented a decisive criticism of Ayer's (1946, 13) second proposal. Consider three logically independent observation statements O_1, O_2, and O_3 and any statement S. The disjunction $(\neg O_1 \wedge O_2) \vee (\neg S \wedge O_3)$ is directly verifiable, since in conjunction with O_1 it entails O_3. Also, $(\neg O_1 \wedge O_2) \vee (\neg S \wedge O_3)$ and S together entail O_2. Hence, by Ayer's criterion, S is indirectly verifiable, unless $(\neg O_1 \wedge O_2) \vee (\neg S \wedge O_3)$ alone entails O_2, which implies $\neg S$ and O_3 entail O_2 so that $\neg S$ is directly verifiable. Thus, according to Ayer's criterion, any statement is indirectly verifiable, and therefore significant, or its negation is directly verifiable, and thereby meaningful.

Nidditch (1961) pointed out that Ayer's (1946) proposal could be amended to avoid Church's (1949) criticism by specifying that the premises could only be analytic, directly verifiable, or indirectly verifiable on independent grounds *and* could only be composed of such statements. Thus that $(\neg O_1 \wedge O_2) \vee (\neg S \wedge O_3)$ and S together entail O_2 does not show that S is indirectly verifiable because $(\neg O_1 \wedge O_2) \vee (\neg S \wedge O_3)$ contains a statement (S) that has not been shown to be analytic, directly verifiable, or independently verifiable on independent grounds. Unfortunately, Scheffler (1963) pointed out that according to Nidditch's (1961) revised criterion, an argument similar to Church's (1949) with the disjunction $\neg O_2 \vee (S \wedge O_1)$ shows that any statement S is significant, unless it is a logical consequence of an observation statement. Scheffler (1963) also pointed out that Ayer's second proposal makes any statement of the form $S \wedge (O_1 \rightarrow O_2)$ significant, where O_1, O_2 are logically independent observation sentences and S is any statement. $S \wedge (O_1 \rightarrow O_2)$ entails $O_2$when conjoined with O_1 and neither the conjunction nor O_1 entails O_2 alone.

Beyond Verifiability: Carnap and Hempel

While Ayer first attempted to formalize the verifiability criterion, Carnap ([1936–7] 1965) recognized the obvious weaknesses of verifiability-based significance criteria. At roughly the same time, in the light of Tarski's rigorous semantic account of truth, Carnap was coming to accept that a systematic (that is, nonpragmatic) account might be possible for other concepts, such as 'confirmation.' He subsequently refocused the question of cognitive significance away from verifiability, which seemed to connote the possibility of definitive establishment of truth, to confirmability—the possibility of obtaining evidence, however partial, for a statement. In particular, Carnap thought a justifiable significance criterion could be formulated if an adequate account of the confirmation of theory by observation were available. A better understanding of the latter would provide a clearer grasp of how scientific terms are significant due to their connection to observation and prediction and how metaphysical concepts are not, because they lack this connection. Yet, insights into the nature of confirmation of theory by observation do not alone determine the form of an adequate significance criterion. Rather, these insights were important because Carnap ([1936–7] 1965) radically changed the nature of the debate over cognitive significance.

Carnap reemphasized (from his work in 1932) that what expressions are cognitively significant depends upon the structure of language, and hence a criterion could be proposed relative to only a specific language. He distinguished two kinds of questions about cognitive significance: those concerning "historically given language system[s]" and those concerning constructible ones (Carnap [1932] 1959, 237). Answers to the two kinds of questions are evaluated by different standards. To be meaningful in the first case, an expression E must be a sentence of L, which is determined by the language's syntax, and it must "fulfill the empiricist criterion of meaning" (167) for L. Carnap does not disclose the exact form of the criterion—verifiability, testability, or confirmability—for a particular language, such as English.

The reason for Carnap's silence, however, was his belief that the second type of question posed a more fruitful direction for the debate. The second type of question is practical, and the answers are proposals, not assertions. Carnap ([1936–7] 1965) remarked that he was no longer concerned with arguing directly that metaphysical statements are not cognitively significant (236). Rather, his strategy was to construct a language L in which every nonanalytic statement was confirmable by some experimental procedure. Given its designed structure, L will clearly indicate how theoretical statements can be confirmed by observational ones, and it will not permit the construction of metaphysical statements. If a language such as L can be constructed that accords with intuitions about the significance of common statements and is sufficient for the purposes of science, then the onus is on the metaphysician to show why metaphysical statements are significant in anything but an emotive or attitude-expressing way.

In a review paper more than a decade later, Hempel (1950) construed Carnap's ([1936–1937] 1965) position as proposing a translatability criterion—a sentence is cognitively significant if and only if it is translatable into an empiricist language (see Hempel, Carl). The vocabulary of an empiricist language L contains observational predicates, the customary logical constants, and any expression constructible from these; the sentence formation rules of L are those of *Principia Mathematica*. The problem Carnap ([1936–1937] 1965) attempted to rectify was that many theoretical terms of science cannot be defined in L.

Hempel's interpretation, however, slightly misconstrued Carnap's intention. Carnap ([1936–7] 1965) did not try to demonstrate how theoretical terms could be connected to observational ones in order to *assert* translatability as a criterion of cognitive significance. Rather, in accord with the *principle of tolerance* (Carnap 1934) Carnap's project in 1936–1937 was to construct an alternative to metaphysically infused language. The features of the language are then evaluated with respect to the purposes of the language user on pragmatic grounds. Although it seems to conflict with his position in 1932, Carnap (1963) clarified that a "neutral attitude toward various forms of language based on the principle that everyone is free to use the language most suited to his purposes, has remained the same throughout my life" (18–19). Carnap ([1936–1937] 1965) tried to formulate a replacement for metaphysics, rather than directly repudiate it on empiricist grounds.

Three definitions were important in this regard. The forms presented here are slightly modified from those given by Carnap ([1936–1937] 1965):

1. The confirmation of a sentence S is completely reducible to the confirmation of a class of sentences C if S is a consequence of a finite subclass of C.
2. The confirmation of S directly incompletely reduces to the confirmation of C if (a) the confirmation of S is not completely reducible to C and (b) there is an infinite subclass C' of mutually independent sentences of C such that S entails, by substitution alone, each member of C'.
3. The confirmation of a predicate P reduces to the confirmation of a class of predicates Q if the confirmation of every full sentence of P with a particular argument (e.g., $P(a)$, in which a is a constant of the language) is reducible to the confirmation of a consistent set of predicates of Q with the same argument, together with their negations.

With these definitions Carnap ([1936–1937] 1965) showed how dispositional predicates (for instance, "is soluble in water") S could be introduced into an empiricist language by means of reduction postulates or finite chains of them. These postulates could take the simple form of a reduction pair:

$$(\forall x)(Wx \rightarrow (Dx \rightarrow Sx));$$
$$(\forall x)(Fx \rightarrow (Rx \rightarrow \neg Sx));$$

in which W, D, F, and R designate observational terms and S is a dispositional predicate. (In the solubility example, $Sx =$ "x is soluble in water"; $Dx =$ "x dissolves in water"; $Wx =$ "x is placed in water"; and R and F are other observational terms.) If (Vx) $(Dx \leftrightarrow \neg Rx)$ and (Vx) $(Wx \leftrightarrow Fx)$, then the reduction pair is a bilateral reduction sentence:

$$(\forall x)(Wx \rightarrow (Dx \leftrightarrow Sx)).$$

The reduction postulates introduce, but do not explicitly define, terms by specifying their logical relations with observational terms. They also provide confirmation relations between the two types of terms. For instance, the above reduction pair entails that the confirmation of S reduces to that of the confirmation of the set $\{W, D, F, R\}$. Carnap ([1936–1937] 1965) defined a sentence or a predicate to be confirmable (following definitions 1–3 above) if its confirmation reduces to that of a class of observable predicates (156–157). Reduction postulates provide such a reduction for disposition terms such as S. The reduction pair does not define S in terms of observational terms. If $\neg Wx$ and $\neg Fx$, then Sx is undetermined. However, the conditions in which S or its negation hold can be extended by adding other reduction postulates to the language. Carnap thought that supplementing an empiricist language to include terms that could be introduced by means of reduction postulates or chains of them (for example, if Wx is introduced by a reduction pair) would adequately translate all theoretical terms of scientific theories.

Although it set a more rigorous standard for the debate, Carnap's ([1936–1937] 1965) proposal encountered difficulties. Carnap believed that bilateral reduction sentences were analytic, since all the consequences of individual reduction sentences that contained only observation terms were tautologies. Yet, Hempel (1951) pointed out that two bilateral reduction sentences together sometimes entailed synthetic statements that contained only observation terms. Since the idea that the conjunction of two analytic sentences could entail synthetic statements was counterintuitive, Hempel made the important suggestion that analyticity and cognitive significance must be relativized to a specific language *and* a particular theoretical context. A bilateral reduction sentence could be analytic in one context but synthetic in a different context that contained other reduction postulates.

Hempel (1950) also argued that many theoretical terms, for instance "gravitational potential" or "electric field," could not be translated into an empiricist language with reduction postulates or chains of them. Introducing a term with reduction postulates provides some sufficient and necessary observation conditions for the term, but Hempel claimed that this was possible only in simple cases, such as electric fields of a simple kind. Introducing a theoretical term with reduction sentences also unduly restricted theoretical concepts to observation conditions. The concept of length could not be constructed to describe unobservable intervals, for instance 1×10^{-100} m, and the principles of calculus would not be constructible in such a language (Hempel 1951). Carnap's ([1936–1937] 1965) proposal could not accommodate most of scientific theorizing.

Although ultimately untenable, adequacy conditions for a significance criterion were included in Carnap's ([1936–1937] 1965) papers, generalized by Hempel (1951) as: If *N* is a nonsignificant sentence, then all truth-functional compound sentences that nonvacuously contain *N* must be nonsignificant. It follows that the denial of a nonsignificant sentence is nonsignificant and that a disjunction, conjunction, or conditional containing a nonsignificant component sentence is also nonsignificant. Yet Hempel (1951) was pessimistic that any adequate criterion satisfying this condition and yielding a sharp dichotomy between significance and nonsignificance could be found. Instead, he thought that cognitive significance was a matter of degree:

> Significant systems range from those whose entire extralogical vocabulary consists of observational terms, through theories whose formulation relies heavily on theoretical constructs, on to systems with hardly any bearing on potential empirical findings. (74).

Hempel suggested that it may be more fruitful to compare theoretical systems according to other characteristics, such as clarity, predictive and explanatory power, and simplicity. On these bases, the failings of metaphysical systems would be more clearly manifested.

Of all the logical empiricists' criteria, Carnap's (1956) criterion was the most sophisticated. It attempted to rectify the deficiencies of his 1936–7 work and thereby avoid Hempel's pessimistic conclusions. Scientific languages were divided into two parts, a theoretical language L_T and an observation language L_O. Let V_O be the class of descriptive constants of L_O, and V_T be the class of *primitive* descriptive constants of L_T. Members of V_O designate observable properties and relations such as 'hard,' 'white,' and 'in physical contact with.' The logical structure of L_O contains only an elementary logic, such as a simple first-order predicate calculus.

The descriptive constants of L_T, called theoretical terms, designate unobservable properties and relations such as 'electron' or 'magnetic field.' L_T contains the mathematics required by science along with the "entities" referred to in scientific physical, psychological, and social theories, though Carnap stressed that this way of speaking does not entail any ontological theses. A theory was construed as a finite set of postulates within L_T and represented by the conjunction of its members T. A finite set of correspondence rules, represented by the conjunction of its members C, connects terms of V_T and V_O.

Within this framework Carnap (1956) presented three definitions, reformulated as:

D1. A theoretical term M is *significant relative* to a class K with respect to L_T, L_O, T, and $C =_{\mathrm{df}}$ if (i) $K \subset V_T$, (ii) $M \notin K$, and (iii) there are three sentences S_M, $S_K \in L_T$, and $S_O \in L_O$ such that:
  (a) S_M contains M as the only descriptive term.
  (b) The descriptive terms in S_K belong to K.
  (c) $(S_M \wedge S_K \wedge T \wedge C)$ is consistent.
  (d) $(S_M \wedge S_K \wedge T \wedge C)$ logically implies S_O.
  (e) $\neg[(S_K \wedge T \wedge C)$ logically implies $S_O]$.

D2. A theoretical term M_n is *significant* with respect to L_T, L_O, T, and $C =_{\mathrm{df}}$ if there is a sequence of theoretical constants $<M_1, \ldots M_n>$ $(M_i \in V_T)$ such that every M_i is significant relative to $\{M_1, \ldots M_{i-1}\}$ with respect to L_T, L_O, T, and C.

D3. An expression A of L_T is a significant sentence of $L_T =_{\mathrm{df}}$ if (i) A satisfies the rules of formation of L_T and (ii) every descriptive term in A is significant, as in D2.

These definitions, especially D1 (d) and (e), are intended to explicate the idea that a significant term must make a predictive difference. Carnap was aware that observation statements can often be deduced only from theoretical statements containing several theoretical terms. With D2 Carnap implicitly distinguishes between theoretical terms whose significance depends on other theoretical terms and those that acquire significance independently of others. In contrast to his work in 1936–7, and in accord with Hempel's (1951) relativization of analyticity and cognitive significance, Carnap (1956) specified that the significance of theoretical terms is relativized to a particular language *and* a particular theory T.

With the adequacy of his proposal in mind, Carnap (1956, 54–6) proved an interesting result. Consider a language in which V_T is divided into empirically meaningful terms V_1 and empirically meaningless terms V_2. Assume that C does not permit any implication relation between those sentences that contain only V_1 or V_O terms and those sentences that contain only V_2 terms. For a given theory T that can be resolved into a class of statements T_1 that contain only terms from V_1, and T_2 that contain only terms of V_2, then a simple but adequate significance criterion can be given. Any theoretical term that occurs only in isolated sentences, which can be omitted from T without affecting the class of sentences of L_O that it entails in conjunction with C, is meaningless.

The problem is that this criterion cannot be utilized for a theory T' equivalent to T that cannot be similarly divided. Carnap (1956), however, showed by indirect proof that his criterion led to the desired conclusion that the terms of V_2 were not significant relative to T' (L_O, L_T, and C) and that therefore the criterion was not too liberal.

The Supposed Failure of Carnap

Kaplan (1975) raised two objections to Carnap's (1956) criterion that were designed to show that it was too liberal and too restrictive. Kaplan's first objection utilized the "deoccamization" of $T \wedge C$. The label is appropriate, since the transformation of $T \wedge C$ into its deoccamization $T' \wedge C'$ involves replacing all instances of some theoretical terms with disjunctions or conjunctions of new terms of the same type: an Occam-unfriendly multiplication of theoretical terms. Kaplan proved that any deductive systematization of L_O by $T \wedge C$ is also established by any of its deoccamizations. This motivates his intuition that deoccamization should preserve the empirical content of a theory and, therefore, not change the significance of its theoretical terms.

The objection is as follows: If any members of V_T are significant with respect to T, C, L_T, and L_O, then there must be at least one M_1 that is significant relative to an empty K (D2). Yet, if $T \wedge C$ is deoccamized such that M_1 is resolved into two new terms M_{11} and M_{12} that are never found apart, then the original argument that satisfied D1 can no longer be used, since $T' \wedge C'$ do not provide similar logical relationships for M_{11} and M_{12} individually. Hence, the sequence of theoretical terms required by D2 will have no first member. Subsequently, no chain of implications that establishes the significance of successive theoretical terms exists. Although deoccamization preserves the deductive systematization of L_O, according to Carnap's criterion it may render every theoretical term of $T' \wedge C'$ meaningless and therefore render $T' \wedge C'$ devoid of empirical content.

Creath (1976) vindicated the core of Carnap's (1956) criterion by generalizing it to accommodate *sets* of terms, reformulated as:

D1′. A theoretical term M is *significant relative* to a class K with respect to L_T, L_O, T, and $C =_{\text{df}}$ if (i) $K \subset V_T$, (ii) $M \notin K$, (iii) there is a class J such that $J \subset V_T$, $M \in J$, but J and K do not share any members, and (iv) there are sentences S_J, $S_K \in L_T$, and $S_O \in L_O$ such that:

- (a) S_J contains members of J as the only descriptive terms.
- (b) The descriptive terms in S_K belong to K.
- (c) $(S_J \wedge S_K \wedge T \wedge C)$ *is* consistent.
- (d) $(S_J \wedge S_K \wedge T \wedge C)$ logically implies S_O.
- (e) $\neg[(S_K \wedge T \wedge C)$ logically implies $S_O]$.
- (f) It is not the case that $(\exists J')(J' \subset J)$ and sentences $S_{J'}$, $S_{K'} \in L_T$, and $S_{O'} \in L_O$ such that:
  - (f1) $S_{J'}$ contains only terms of J' as its descriptive terms.
  - (f2) The descriptive terms of $S_{K'}$ belong to K.
  - (f3) $(S_{J'} \wedge S_{K'} \wedge T \wedge C)$ is consistent.
  - (f4) $(S_{J'} \wedge S_{K'} \wedge T \wedge C)$ logically implies $S_{O'}$.
  - (f5) $\neg[(S_{K'} \wedge T \wedge C)$ logically implies $S_{O'}]$.

D2′. A theoretical term M_n is *significant* with respect to L_T, L_O, T and $C =_{\text{df}}$ if there is a sequence of sets $<J_1, \ldots J_n>$ ($M_n \in J_n$ and $J_i \subset V_T$) such that every member of every set J_i is significant relative to the union of

J_1 through J_{i-1} with respect to L_T, L_O, T and C.

Condition (f) ensures that each member of J is required for the significance of the entire set. Creath (1976) points out that any term made significant by D1 and D2 of Carnap (1956) is made significant by D1′ and D2′ and that according to the generalized criterion, Kaplan's (1975) deoccamization criticism no longer holds.

Kaplan (1975) and Rozeboom (1960) revealed an apparent second flaw in Carnap's (1956) proposal: As postulates (definitions for Kaplan's criticism) are added to $T \wedge C$, the theoretical terms it contains may change from cognitively significant to nonsignificant or vice versa. Consider an example from Kaplan (1975) in which $V_O = \{J_O, P_O, R_O\}$; L_O is the class of all sentences of first-order logic with identity that contain no descriptive constants or only those from V_O; $V_T = \{B_T, F_T, G_T, H_T, M_T, N_T\}$; and L_T is the class of all sentences of first-order logic with identity that contain theoretical terms from V_T. Let T be:

$$(T)[(\forall x)(H_T x \rightarrow F_T x)] \wedge [(\forall x)(H_T x \rightarrow (B_T x \vee \neg G_T x))] \wedge [(\forall x)(M_T x \leftrightarrow N_T x)];$$

and let C be:

$$(C)[(\forall x)(R_O x \rightarrow H_T x)] \wedge [(\forall x)(F_T x \rightarrow J_O x)] \wedge [(\forall x)(G_T x \rightarrow P_O x)].$$

G_T, F_T, and H_T are significant with respect to $T \wedge C$ relative to the empty set (see Carnap [1956] D1) and, hence, significant with respect to L_O, L_T, T, and C (see Carnap [1956] D2). R_O is significant relative to $K = \{G_T\}$; M_T and N_T are not significant.

Consider a definitional extension T' of T in an extended vocabulary V_T' and language L_T'. After adding two definitions to T:

$$(\text{DEF1})(\forall x)(D1_T x \leftrightarrow (M_T x \wedge (\exists x) F_T x))$$

and

$$(\text{DEF2})(\forall x)(D2_T x \leftrightarrow (M_T x \rightarrow (\exists x) G_T x)),$$

$D1_T$ is significant relative to the empty set and therefore significant with respect to T', C, L_O, and L_T' (D2). $D2_T$ is significant relative to $K = \{D1_T\}$, and therefore significant with respect to T', C, L_O, and L_T' (D2). M_T, which failed to be significant with respect to T, C, L_O, and L_T, is now significant with respect to T', C, L_O, and L_T'. A similar procedure makes N_T significant. Kaplan thought this showed that Carnap's (1956) criterion was too liberal. The procedure seems able to make any theoretical term significant with respect to some extended language and definition-extended theory, but "definitional extensions are ordinarily thought of as having no more empirical content than the original theory" (Kaplan 1975, 90).

Using the same basic strategy, Rozeboom (1960) demonstrated that extending $T \wedge C$ can transform an empirically significant term into an insignificant one. Consider a term M that is significant with respect to T, C, L_T, and L_O. Rozeboom showed that if postulates (not necessarily definitions) are added to T or to C to form T' or C', in some cases D1(e) will no longer be satisfied, and no other sentences S_M', S_K', S_O' exist by which M could be independently shown to be significant. Furthermore, if $T \wedge C$ is maximally L_O consistent, no theoretical term of L_T is significant, since D1(e) is never satisfied; for any S_O, if $T \wedge C$ is maximally L_O consistent then it alone implies S_O. Rozeboom (1960) took the strength of his criticism to depend upon the claim that for a criterion to be "intuitively acceptable," theoretical terms must retain significance if T or C is extended.

Carnap (1956) can be defended in at least two ways. First, as Kaplan (1975) notes, the criterion was restricted to primitive, nondefined theoretical terms. It was explicitly formulated to avoid criticisms derived from definitional extensions. Defined terms often play an important role in scientific theories, and it could be objected that any adequate criterion should apply directly to theories that contain them. Yet the amendment that any theoretical term within the definiens of a significant defined term must be antecedently shown significant quells these worries (Creath 1976).

Second, Carnap (1956) insisted that terms are significant only *within a particular language* and *for a particular T and C*. He did not intend to formulate a criterion of cognitive significance that held under theory or language change. If Carnap's (1956) work on a significance criterion was an explication of the idea of meaningfulness (Hempel 1950), the explicandum was the idea of a meaningful statement of a particular language in a particular theoretical context, not meaningfulness per se. Hence, Kaplan and Rozeboom's objections, which rely on questionable intuitions about the invariance of significance as $T \wedge C$ changes, are not appropriately directed at Carnap (1956). The fact that Carnap did not attempt such an account is not merely the result of a realization that so many problems would thwart the project. Rather, it is a consequence of the external/internal framework that he believed was the most fruitful approach to the philosophical questions (Carnap 1947).

Furthermore, Rozeboom's acceptability condition is especially counterintuitive, since changes in

T or *C* designate changes in the connections between theoretical terms themselves or theoretical terms and observation terms. Additional postulates that specify new connections, or changes in the connections, between these terms can obviously change the significance of a theoretical term. Scientific advances are sometimes made when empirical or theoretical discoveries render a theoretical term nonsignificant.

James Justus

References

Ayer, A. J. (1934), "Demonstration of the Impossibility of Metaphysics," *Mind* 43: 335–345.

——— (1946), *Language, Truth, and Logic*, 2nd ed. New York: Dover Publications.

Berlin, I. (1939), "Verification," *Proceedings of the Aristotelian Society* 39: 225–248.

Carnap, R. (1928a), *Der Logische Aufbau der Welt*. Berlin-Schlachtensee: Weltkreis-Verlag.

——— (1928b), *Scheinprobleme in der Philosophie: Das Fremdpsychische und der Realismusstreit*. Berlin-Schlachtensee: Weltkreis-Verlag.

——— ([1932] 1959), "The Elimination of Metaphysics Through Logical Analysis of Language," in A. J. Ayer (ed.), *Logical Positivism*. Glencoe, IL: Free Press, 60–81.

——— (1934), *The Logical Syntax of Language*. London: Routledge Press.

——— ([1936–7] 1965), "Testability and Meaning," in R. R. Ammerman (ed.), *Classics of Analytic Philosophy*. New York: McGraw-Hill, 130–195.

——— (1947), "Empiricism, Semantics, and Ontology," in *Meaning and Necessity*. Chicago: University of Chicago Press, 205–221.

——— (1956), "The Methodological Character of Theoretical Concepts," in H. Feigl and M. Scriven (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minneapolis: University of Minnesota Press, 38–76.

——— (1963), "Intellectual Autobiography," in P. Schlipp (ed.), *The Philosophy of Rudolf Carnap*. Peru, IL: Open Court Press, 3–86.

——— (1967), *Logical Structure of the World and Pseudoproblems in Philosophy*. Berkeley and Los Angeles: University of California Press.

Church, A. (1949), "Review of the Second Edition of *Language, Truth and Logic*, " *Journal of Symbolic Logic* 14: 52–53.

Creath, R. (1976), "Kaplan on Carnap on Significance," *Philosophical Studies* 30: 393–400.

Feigl, H. (1969), "The Origin and Spirit of Logical Positivism," in P. Achinstein and S. F. Barker (eds.), *The Legacy of Logical Positivism*. Baltimore, MD: Johns Hopkins Press, 3–24.

Hempel, C. (1950), "Problems and Changes in the Empiricist Criterion of Meaning," *Revue Internationale de Philosophie* 11: 41–63.

——— (1951), "The Concept of Cognitive Significance: A Reconsideration," *Proceedings of the American Academy of Arts and Sciences* 80: 61–77.

Kaplan, D. (1975), "Significance and Analyticity," in J. Hintikka (ed.), *Rudolf Carnap, Logical Empiricist*. Dordrecht, Netherlands: D. Reidel Publishing Co., 87–94.

Lewis, C. I. (1934), "Experience and Meaning," *Philosophical Review* 43: 125–146.

Nidditch, P. (1961), "A Defense of Ayer's Verifiability Principle Against Church's Criticism," *Mind* 70: 88–89.

Reichenbach, H. (1938), *Experience and Prediction*. Chicago: University of Chicago Press.

Rozeboom, W. W. (1960), "A Note on Carnap's Meaning Criterion," *Philosophical Studies* 11: 33–38.

Scheffler, I. (1963), *Anatomy of Inquiry*. New York: Alfred A. Knopf.

Schlick, M. ([1932] 1979), "Positivism and Realism," in H. L. Mulder and B. F. B. van de Velde-Schlick (eds.), *Moritz Schlick: Philosophical Papers (1926–1936)*, vol. 2. Dordrecht, Netherlands: D. Reidel Publishing Co., 259–284.

———([1936] 1979), "Meaning and Verification," in H. L. Mulder and B. F. B. van de Velde-Schlick (eds.), *Moritz Schlick: Philosophical Papers (1926–1936)*, vol. 2. Dordrecht, Netherlands: D. Reidel Publishing Co., 456–481.

Stace, W. T. (1935), "Metaphysics and Meaning," *Mind* 44: 417–438.

See also **Analyticity; Ayer, Alfred Jules; Carnap, Rudolf; Corroboration; Demarcation, Problem of; Explication; Feigel, Herbert; Hempel, Carl; Logical Empiricism; Neurath, Otto; Popper, Karl; Rational Reconstruction; Reichenbach, Hans; Schlick, Moritz; Verifiability; Vienna Circle**

COMPLEMENTARITY

The existence of indivisible interaction quanta is a crucial point that implies the *impossibility of any sharp separation between the behavior of atomic objects and the interaction with the measuring instruments that serve to define the conditions under which the phenomena appear*. In fact, the individuality of the typical quantum effects finds its proper expression in the circumstance that any attempt at subdividing the phenomena will demand a change in the experimental arrangement, introducing new possibilities of interaction between objects and measuring instruments, which in principle cannot be controlled. Consequently, evidence obtained under different experimental conditions cannot be comprehended within a single picture but must be regarded as "complementary" in the sense that only the totality of the phenomena exhausts the possible information about the objects (Bohr 2000, 209–10).

Complementarity is distinctively associated with the Danish physicist Niels Bohr and his attempt to understand the new quantum mechanics (QM) during the heyday of its invention, 1920–1935 (see Quantum Mechanics). Physicists know in practice how to extract very precise and accurate predictions and explanations from QM. Yet, remarkably, even today no one is confident about how to interpret it metaphysically ("M-interpret" it). That is, there is no compellingly satisfactory account of what sort of objects and relations make up the QM realm, and so, ultimately, no one knows why the precise answers are as they are. In classical mechanics (CM), physicists thought they had a lucidly M-interpretable theory: There was a collection of clearly specified entities, particles, or waves that interacted continuously according to simple laws of force so that the system state was completely specified everywhere and at all times—indeed, specification of just instantaneous position and momenta sufficed (see Classical Mechanics). Here the dynamic process specified by the laws of force, expressed in terms of energy and momentum (Bohr's "causal picture"), generated a uniquely unfolding system configuration expressed in terms of position and time (Bohr's "space-time picture"). To repeat this for QM, what is needed is a collection of equivalent quantum objects whose interactions and movements in space-time generate the peculiar QM statistical results in ways that are as intuitively clear as they are for CM. (However, even this appealing concept of CM proves too simplistic; there is continuing metaphysical perplexity underlying physics; see e.g., Earman 1986; Hooker 1991, note 13.)

That M-interpreting QM is not easy is nicely illustrated by the status of the only agreed-on general "interpretation" to which physicists refer, Born's rule. It specifies how to extract probabilities from the QM wave function (ψ ["psi"]-function). These are normally associated with particle-like events. But without further M-interpretive support, Born's rule becomes merely part of the recipe for extracting numbers from the QM mathematics. That it does not M-interpret the QM mathematics is made painfully clear by the fact that the obvious conception of the QM state it suggests—a statistical ensemble of particles, each in a definite classical state—is provably not a possible M-interpretation of QM. (For example, no consistent sense can then be made of a superposition of ψ-states, since it is a mathematical theorem that the QM statistics of a superposed state cannot all be deduced from any single product of classical statistical states.)

Bohr does not offer an M-interpretation of QM. He came to think the very idea of such an interpretation incoherent. (In that sense, the term "Copenhagen interpretation" is a misnomer; Bohr does not use this label.) Equally, however, Bohr does not eschew all interpretive discussion of QM, as many others do on the (pragmatic or positivist-inspired) basis that confining the use of QM strictly to deriving statistics will avoid error while allowing science to continue. Bohr's position is that this too is profoundly wrong, and ultimately harmful to physics. Instead he offers the doctrine of complementarity as a "framework" for understanding the "epistemological lesson" of QM (not its ontological lesson) and for applying QM consistently and as meaningfully as possible. (see Folse 1985 for a general introduction. For extensive, more technical analyses, see Faye and Folse 1994; Honner 1987; Murdoch 1987. For one of many critiques, and the opposing Bohrian M-interpretation, see Cushing 1994.)

Although Bohr considered it necessary ("unavoidable") to continue using the key descriptive concepts of CM, the epistemological lesson of QM was that the basic conditions for their well-defined use were altered by the quantization of QM interactions into discrete unanalyzable units. This, he argued, divided the CM state in two, the conditions for the well-defined use of (1) causal (energy-momentum) concepts and (2) configurational (space-time) concepts being now mutually exclusive, so that only one kind of description could be coherently provided at a time. Both kinds of description were necessary to capture all the aspects of a QM system, but they could not be simply conjoined as in CM. They were now complementary.

Heisenberg's uncertainty relations of QM, such as $\Delta x \Delta p \geq h/2\pi$, where Δx is the uncertainty in the position, Δp is the uncertainty in the momentum, and h is Planck's constant (or more generally the commutation relations, such as $[x, p_x] = ih/2\pi$ where h is again Planck's constant, the magnitude of quantization), are not themselves statements of complementarity. Rather, they specify the corresponding quantitative relationships between complementary quantities.

These complementary exclusions are implicit in QM ontologies. For instance, single-frequency ("pure") waves must, mathematically, occupy all space, whereas restricting their extent involves superposing waves of different frequencies, with a point-size wave packet requiring use of all frequencies; thus, frequency and position are not uniquely cospecifiable. And since the wavelength λ is the wave velocity V (a constant) divided by the frequency v ($\lambda = V/v$), uniquely specifying wavelength and uniquely specifying position equally are mutually exclusive. QM associates wavelength with momentum ($p = h/\lambda$) and energy with frequency ($E = hv$) in all cases with discrete values and for both radiation (waves) and matter (particles), yielding the QM exclusions. (Note, however, that wave/particle complementarity is but one aspect of causal/configurational complementarity, the aspect concerned with physical conditions that frame coherent superposition versus those that frame localization.)

Precisely why these particular associations (and similar QM associations) should follow from quantization of interaction is not physically obvious, despite Bohr's confident assertion. Of course such associations follow from the QM mathematics, but that presupposes rather than explains complementarity, and Bohr intended complementarity to elucidate QM. The physical and mathematical roots of quantization are still only partially understood. However, it is clear that discontinuity leads to a constraint, in principle, on joint precise specification. Consider initially any quantity that varies with time (t), for example, energy (E), so that $E = f(t)$.Then across some time interval, $t_1 \rightarrow t_2$, E will change accordingly: $E_1(t_1) \rightarrow E_2(t_2)$.

If E varies continuously, then both E and t are everywhere jointly precisely specifiable because for every intermediate value of t between t_1 and t_2 (say, t_{1+n}) there will be a corresponding value for E: $E_{1+n} = f(t_{1+n})$. Suppose, however, that E (but not t) is quantized, with no allowed value between E_1 and E_2. Then no intermediate E value is available, and energy must remain undefined during at least some part of the transition period. This conclusion can be generalized to any two or more related quantities. The problem is resolved if both quantities are quantized, but there is as yet no satisfactory quantization of space and time (Hooker 1991).

Such inherent mutual exclusions should not be mistaken for merely practical epistemic exclusions (some of Heisenberg's pronouncements notwithstanding). Suppose that the position of an investigated particle i is determined by bouncing ("scattering") another probe particle p off of it, determining the position of i from the intersection of the initial and final momenta of p. However, i will have received an altered momentum in the interaction, and it is tempting to conclude that we are thus excluded from knowing both the position and the momentum of i immediately after the interaction. But in CM the interaction may be retrospectively analyzed to calculate the precise change in momentum introduced by p to i, using conservation of momentum, and so establish both the position and the momentum of i simultaneously. More generally, it is in this manner possible to correct for all measurement interactions and arrive at a complete classical state specified independently of its method of measurement. This cannot, in principle, be done in QM, because of quantization.

Faye (1991) provides a persuasive account of the origins of Bohr's ideas about the applicability of physical concepts in the thought of the Danish philosopher Harald Høffding (a family friend and early mentor of Bohr's) and sets out Bohr's consequent approach. According to Høffding, objective description in principle required a separation between describing a subject and describing a known object (Bohr's "cut" between them) in a way that always permitted the object to be ascribed a unique (Bohr's "unambiguous") spatiotemporal location, state, and causal interaction. These ideas in turn originated in the Kantian doctrine that an objective description of nature requires a well-defined distinction between the knower and the object of knowledge, permitting the unique construction of a well-defined object state, specified in applications of concepts from the synthetic a priori (essentially Newtonian) construction of the external world (see Friedman 1992). We have just noted how CM satisfies this requirement.

Contrarily, Bohr insisted, the quantum of action creates an "indissoluble bond" between the measurement apparatus (m-apparatus, including the sentient observer) and the measured (observed) system, preventing the construction of a well-defined system state separate from observing interactions. This vitiates any well-defined, global cut between m-apparatus and system. Creating a set of complementary partial cuts is the best that can now be done. In fact, these circumstances are generalized to all interactions between QM systems; the lack of a global separation is expressed in their superposition, which defies reduction to any combination of objectively separate states. Consequently, Bohr regarded CM as an idealized physics (achieved, imperfectly, only in the limit $h \rightarrow 0$) and QM as a "rational generalization" of it, in the sense of the principle of affinity, the Kantian methodological requirement of continuity.

Bohr's conception of what is required of a physically intelligible theory T can thus be summarized as follows (Hooker 1991, 1994):

BI1. Each descriptive concept A of T has a set of well-defined, epistemically accessible conditions C_A under which it is unambiguously applicable.
BI2. The set of such concepts collectively exhausts, in a complementary way, the epistemically accessible features of the phenomena in the domain of T.
BI3. There is a well-defined, unified, and essentially unique formal structure $S(T)$ that structures and coordinates descriptions of phenomena so that each description is well defined (the various conditions C_A are consistently combined), $S(T)$ is formally complete (Bub 1974), and BI2 is met.
BI4. Bohr objectivity (BO) satisfies BI1–3 in the most empirically precise and accurate way available across the widest domain of phenomena while accurately specifying the interactive conditions under which such phenomena are accessible to us.

An objective representation of nature thus reflects the interactive access ("point of view") of the knowing subject, which cannot be eliminated. In coming to know nature, we also come to know ourselves as knowers—not fundamentally by being modeled in the theory as *objects* (although this too happens, in part), but by the way the very form of rational generalization reflects our being as knowing *subjects.*

A very different ideal of scientific intelligibility operates in classical physics, and in many proposed M-interpretations of QM. Contrary to BI1, descriptive concepts are taken as straightforwardly characterizing external reality (describing an M, even if it is a strange one). Hence, contrary to BI2, these concepts apply conjointly to describe reality completely. Contrary to BI3 and BI4, an objective theory completely and accurately describes the physical state at each moment in time and provides a unique interactive dynamic history of states for all systems in its domain. Accordingly, measurements are analyzed similarly as the same kinds of dynamic interactions, and statistical descriptions reflect (only) limited information about states and are not fundamental (contrary to common readings of QM). Here the objective representation of nature through invariances eliminates any inherent reference to any subject's point of view. Rather, in coming to know nature we also come to know ourselves as knowers by being modeled in the theory as some *objects* among others so as to remove ourselves from the form of the theory, disappearing as *subjects*. This shift in ideals of intelligibility and objectivity locates the full depth of Bohr's doctrine of complementarity.

References

Bohr, N. (2000), "Discussion with Einstein on Epistemological Problems in Atomic Physics," in P. A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*, 3rd ed. La Salle, IL: Open Court, 199–242.

——— (1961), *Atomic Theory and the Description of Nature*. Cambridge: Cambridge University Press.

Bub, J. (1974), *The Interpretation of Quantum Mechanics*. Dordrecht, Netherlands: Reidel.

Cushing, J. T. (1994), "A Bohmian Response to Bohr's Complementarity," in J. Faye and H. J. Folse (eds.), *Niels Bohr and Contemporary Philosophy*. Dordrecht, Netherlands: Kluwer.

Earman, J. (1986), *A Primer on Determinism*. Dordrecht, Netherlands: Reidel.

Faye, J. (1991), *Niels Bohr: His Heritage and Legacy*. Dordrecht, Netherlands: Kluwer.

Faye, J., and H. J. Folse (eds.) (1994), *Niels Bohr and Contemporary Philosophy*. Dordrecht, Netherlands: Kluwer.

Folse, H. J. (1985), *The Philosophy of Niels Bohr*. Amsterdam: North-Holland.

Friedman, M. (1992), *Kant and the Exact Sciences*. Cambridge, MA: Harvard University Press.

Honner, J. (1987), *The Description of Nature*. Oxford: Clarendon.

Hooker, C. A. (1972), "The Nature of Quantum Mechanical Reality: Einstein versus Bohr," in R. G. Colodny (ed.), *Pittsburgh Studies in the Philosophy of Science*, vol. 5. Pittsburgh, PA: University of Pittsburgh Press.

——— (1991), "Physical Intelligibility, Projection, and Objectivity: The Divergent Ideals of Einstein and Bohr," *British Journal for the Philosophy of Science* 42, 491–511.

——— (1994), "Bohr and the Crisis of Empirical Intelligibility: An Essay on the Depth of Bohr's Thought and Our Philosophical Ignorance," in J. Faye and H. J. Folse (eds.), *Niels Bohr and Contemporary Philosophy*. Dordrecht, Netherlands: Kluwer.

Murdoch, D. (1987), *Niels Bohr's Philosophy of Physics*. Cambridge: Cambridge University Press.

C. A. Hooker

See also **Classical Mechanics; Quantum Measurement Problem; Quantum Mechanics**

COMPLEXITY

See **Unity and Disunity of Science**

CONFIRMATION THEORY

When evidence does not conclusively establish (or refute) a hypothesis or theory, it may nevertheless provide *some* support for (or against) the hypothesis or theory. Confirmation theory is concerned almost exclusively with the latter, where conclusive support (or "countersupport") are limiting cases of confirmation (or disconfirmation). (Included also, of course, is concern for the case in which the evidence is confirmationally irrelevant.) Typically, confirmation theory concerns *potential* support, the impact that evidence *would* have on a hypothesis or theory if learned, where whether the evidence is actually learned or not is not the point; for this reason, confirmation theory is sometimes called the *logic* of confirmation. (For simplicity of exposition for now, theories will be considered separately below and not explicitly mentioned until then.)

It is relevant here to point out the distinction between deductive logic (or deductive evaluation of arguments) and inductive logic (or inductive evaluation of arguments). In deductive logic, the question is just *whether or not* the supposed truth of all the premises of an argument gives an *absolute guarantee* of truth to the conclusion of the argument. In inductive logic, the question is whether the supposed truth of all the premises of an argument gives *significant support* for the truth of the conclusion, where, ideally, some measure of *to what degree* the premises support the conclusion (which is sometimes called the inductive probability of an argument) would be provided (see Inductive Logic; Induction, Problem of; and Verisimilitude. As in each of these topics also, the question is one of either qualitative or quantitative support that premises or evidence provides to a conclusion or that a hypothesis has. See Carnap, Rudolf, for an idea of degree of confirmation based on his proposed "logical" interpretation of probability and degree of support.) In the theory of the logic of support, confirmation theory is concerned primarily with inductive support, where the theory of deductive support is supposed to be more fully understood.

The concept of confirmation can be divided into a number of subconcepts, corresponding to three distinctions. First, absolute confirmation and incremental confirmation may be distinguished. In the absolute sense, a hypothesis is confirmed by evidence if the evidence makes (or would make) the hypothesis highly supported; absolute confirmation is about how the evidence "leaves" the hypothesis. In the incremental sense, evidence confirms a hypothesis if the evidence makes the hypothesis *more* highly confirmed (in the absolute sense) than it is (in the absolute sense) without the evidence; incremental confirmation involves a comparison. Second, confirmation can be thought of either *qualitatively* or *quantitatively*. So, in the absolute sense of confirmation, a hypothesis can be, qualitatively, left *more or less* confirmed by evidence, where quantitative confirmation theory attempts to make sense of assigning *numerical degrees* of confirmational support ("inductive probabilities") to hypotheses in light of the evidence. In the incremental sense of confirmation, evidence E may, qualitatively, either confirm, disconfirm, or be evidentially irrelevant to a hypothesis H, where in the quantitative sense, a numerical magnitude (which can be measured, "inductive probabilistically," in different ways; see below) is assigned to the "boost" (positive or negative, if any) that E gives H. Finally, confirmation can be considered to be either comparative or noncomparative. Noncomparative confirmation concerns just one hypothesis/evidence pair. In comparative confirmation, one can compare how well an E supports an H with how well an E' supports the same H; or one may compare how well an E supports an H with how well the same E supports an H'; or one may compare how well an E supports an H with how well an E' supports an H'.

The exposition below will be divided into two main parts. The first part, "Nonprobabilistic Approaches," will concern different aspects of qualitative confirmation; and the second part, "Probabilistic Approaches," will consider some major quantitative approaches. Almost exclusively, as in the literature, the issue will be incremental confirmation rather than absolute confirmation. Both noncomparative and various kinds of comparative approaches will be described.

Nonprobabilistic Approaches

A simple and natural idea about the confirmation of a general hypothesis of the form "All *F*s are *G*s" is that an object confirms the hypothesis if and only if it is both an F and a G (a "positive instance" of the hypothesis), disconfirms the hypothesis if and only if it is an F but not a G (a "negative instance"), and is evidentially irrelevant if and only if it is not even an F (no kind of instance). Hempel ([1945] 1965) calls this Nicod's criterion (Nicod 1930). Another natural idea about the confirmation of hypotheses is that if hypotheses H and H' are logically equivalent, then evidence E confirms, disconfirms, or is irrelevant to H if and only if E confirms, disconfirms, or is irrelevant to H', respectively. Hempel calls this the *equivalence condition*, and distinguishes between criteria (definitions or partial definitions) of confirmation and the conditions of adequacy that the criteria should satisfy. Hempel points out that Nicod's criterion does not satisfy the equivalence condition (as long as confirmation, disconfirmation, and evidential irrelevance are mutually exclusive). For example, a hypothesis "All *F*s are *G*s" is logically equivalent to "All non-*G*s are non-*F*s," but Nicod's criterion implies that an object that is an F and a G would confirm the former but be irrelevant to the latter, thus violating the equivalence condition. Also, "All *F*s are *G*s" is logically equivalent to "Anything that is both an F and a non-G is both an F and a non-F," which Nicod's criterion implies that nothing can confirm (a positive instance would have to be both an F and a non-F).

Thus, Hempel suggests weakening Nicod's criterion. The idea that negative instances disconfirm (i.e., are *sufficient* to disconfirm) is retained. Further, Hempel endorses the positive-instance criterion, according to which positive instances are *sufficient* for confirmation. Nicod's criterion can be thought of as containing six parts: necessary and sufficient conditions for all three of confirmation, disconfirmation, and irrelevance. The positive-instance criterion is said to be one-sixth of Nicod's criterion, and it does not lead to the kind of contradiction that Nicod's full criterion does when conjoined with the equivalence condition.

However, the combination of the positive-instance criterion and the equivalence condition (i.e., the proposition that the positive-instance criterion satisfies the equivalence condition) does lead to what Hempel called *paradoxes of confirmation*, also known as the Ravens paradox and Hempel's paradox. Hempel's famous example is the hypothesis H: "All ravens are black." Hypothesis H is logically equivalent to hypothesis H': "All nonblack things are nonravens." According to the equivalence condition, anything that confirms H' confirms H. According to the positive-instance criterion, nonblack nonravens (positive instances of H') confirm H'. Examples of nonblack nonravens (positive instances of H') include white shoes, yellow pencils, transparent tumblers, etc. So it follows from the positive-instance criterion plus the equivalence condition that objects of the kinds just listed confirm the hypothesis H that all ravens are black. These conclusions seem incorrect or counterintuitive, and the paradox is that the two seemingly plausible principles, the positive-instance criterion and the equivalence condition, lead, by valid reasoning, to these seemingly implausible conclusions. Further paradoxical consequences can be obtained by noting that the hypothesis H is logically equivalent also to H'', "All things that are either a raven or not a raven (i.e., all things) are either black or not a raven," which has as positive instances any objects that are black and any objects that are not ravens.

Since the equivalence condition is so plausible (if H and H' are logically equivalent, they can be thought of as simply different formulations of the same hypothesis), attention has focused on the positive-instance criterion. Hempel defended the criterion, arguing that the seeming paradoxicalness of the consequences of the criterion is more of a psychological illusion than a mark of a logical flaw in the criterion:

> In the seemingly paradoxical cases of confirmation, we are often not actually judging the relation of the given evidence *E* alone to the hypothesis *H* . . . . [I]nstead, we tacitly introduce a comparison of *H* with a body of evidence which consists of *E* in conjunction with additional information that we happen to have at our disposal. (Hempel [1945] 1965, 19)

So, for example, if one is just given the information that an object is nonblack and a nonraven (where it may happen to be a white shoe or a yellow pencil, but this is not included in the evidence), then the idea is that one should intuitively judge the evidence as confirmatory, "and the paradoxes vanish" (20). To assess properly the seemingly paradoxical cases for their significance for the logic of confirmation, one must observe the "*methodological* fiction" (as Hempel calls it) that one is in a position to judge the relation between the given evidence *alone* (e.g., that an object is a positive instance of the contrapositive of a universalized conditional) and the hypothesis in question and that there is no other information. This approach has been challenged by some who have argued that confirmation should be thought of as a relation among three things: evidence, hypothesis, and background knowledge (see the section below on Probabilistic Approaches).

Given the equivalence condition, the Ravens paradox considers the question of which of *several* kinds of evidence confirm(s) what one can consider to be a single hypothesis. There is another kind of paradox, or puzzle, that arises in a case of a single body of evidence and multiple hypotheses. In Nelson Goodman's (1965) well-known Grue paradox or puzzle, a 'new' predicate is defined as follows. Let's say that an object A is "grue" if and only if *either* (i) A has been observed before a certain time t (which could be now or some time in the future) and A is green *or* (ii) A has not been observed before that time t and A is blue. Consider the hypothesis H that all emeralds are green and the hypothesis H' that all emeralds are grue. And consider the evidence E to be the observation of a vast number of emeralds, all of which have been green. Given that t is now or some time in the future, E is equivalent to E', that the vast number of emeralds observed have all been grue. It is taken that E (the "same" as E') confirms H (this is natural enough) but not H'—for in order for H' to be true, exactly all of the unobserved (by t) emeralds would have to be blue, which would seem to be disconfirmed by the evidence. Yet, the evidence E' (or E) consists of positive instances of H'.

Since the positive-instance criterion can be formulated purely syntactically—in terms of simply the logical forms of evidence sentences and hypotheses and the logical relation between their forms—a natural lesson of the Grue example is that confirmation cannot be characterized purely syntactically. (It should be noted that an important feature of Hempel's ([1945] 1965) project was the attempt to characterize confirmation purely syntactically, so that evidence E should, strictly speaking, be construed as evidence statements or sentences, or "observation reports," as he put it, rather than as observations or the objects of observation.) And a natural response to this has been to try to find nonsyntactical features of evidence and hypothesis that differentiate cases in which positive instances confirm and cases in which they do not. And a natural idea here is to distinguish between *predicates* that are "projectible" (in Goodman's terminology) and those that are not. Goodman suggested "entrenchment" of predicates as the mark of projectibility—where a predicate is entrenched to the extent to which it has been used in the past in hypotheses that have been successfully confirmed. Quine (1969) suggested drawing the distinction in terms of the idea of natural kinds. A completely different approach would be to point out that the reason why one thinks the observation of grue emeralds (E' or E) disconfirms the grue hypothesis (H') is because of background knowledge about constancy of color (in our usual concept of color) of many kinds of objects, and to argue that the evidence in this case should be taken as actually confirming the hypothesis H, given the Hempelian methodological fiction. It should be pointed out that the positive-instance criterion applies to a limited, though very important, kind of hypothesis and evidence: universalized conditionals for the hypothesis and positive instances for the evidence. And it supplies only a sufficient condition for confirmation. Hempel ([1945] 1965) generalized, in a natural way, this criterion to his satisfaction criterion, which applies to different and more complex logical structures for evidence and hypothesis and provides explicit definitions of confirmation, disconfirmation, and evidential irrelevance. Without going into any detail about this more general criterion, it is worth pointing out what Hempel took to be evidence for its adequacy. It is the satisfaction, by the satisfaction criterion, of what Hempel took to be some intuitively obvious conditions of adequacy for definitions, or *criteria*, of confirmation. Besides the equivalence condition, two others are the entailment condition and the special-consequence condition. The entailment condition says that evidence that logically entails a hypothesis should be deemed as confirming the hypothesis. The special-consequence condition says that if evidence E confirms hypothesis H and if H logically entails hypothesis H', then E confirms H'. This last condition will be considered further in the section below on probabilistic approaches.

The criteria for confirmation discussed above apply in cases in which evidence reports and hypotheses are stated in the "same language," which Hempel took to be an observational language. Statements of evidence, for example, are usually referred to as observational reports in Hempel ([1945] 1965). What about confirmation of *theories*, though, which are often thought of as containing two kinds of vocabulary, observational and theoretical? Hypothetico-deductivism (HD) is the idea that theories and hypotheses are confirmed by their observational deductive consequences. This is different from the positive-instance criterion and the satisfaction criterion. For example, "A is an F and A is a G," which is a report of a positive instance of the hypothesis that all Fs are Gs, is not a deductive consequence of the hypothesis. The positive-instance and satisfaction criteria are formulations of the idea, roughly, that observations that are *logically consistent with* a hypothesis confirm the hypothesis, while HD says that *deductive consequences of* a hypothesis or theory confirm the hypothesis or theory.

As an example, Edmund Halley in 1705 published his prediction that a comet, now known as Halley's comet, would be visible from Earth sometime in December of 1758; he deduced this using Newtonian theory. The prediction was successful, and the December 1758 observation of the comet was taken by scientists to provide (further) very significant confirmation of Newtonian theory. Of course, the prediction was not deduced from Newtonian theory alone. In general, other needed premises include statements of *initial conditions* (in the case of the example, reports of similar or related observations at approximately 75-year intervals) and *auxiliary assumptions* (that the comet would not explode before December 1758; that other bodies in the solar system would have only an insignificant effect on the path of the comet; and so on). In addition, when the theory and the observation report share no nonlogical vocabulary (say, the theory is highly theoretical, containing no observational terms), then so-called bridge principles are needed to establish a deductive connection between theory and observation. An example of such a principle would be, "If there is an excess of electrons [theoretical] on the surface of a balloon [observational], then a sheet of paper [observational] will cling [observational] to it, in normal circumstances [auxiliary assumption]." Of course, if the prediction fails (an observational deductive consequence of a theory turns out to be false), then this is supposed to provide disconfirmation of the theory.

Two of the main issues or difficulties that have been discussed in connection with the HD idea have to do with what might be called distribution of credit and distribution of blame. The first has also been called the *problem of irrelevant conjunction*. If a hypothesis H logically implies an observation report E, then so does the conjunction, $H \wedge G$, where G can be any sentence whatsoever. So the basic idea of HD has the consequence that whenever an E confirms an H, the E confirms also $H \wedge G$, where G can be any (even irrelevant) hypothesis whatsoever. This problem concerns the distribution of credit. A natural response would be to refine the basic HD idea in a way to make it sensitive to the possibility that logically weaker parts of a hypothesis may suffice to deductively imply the observation report. The second issue has to do with the possibility of the failure of the prediction, of the observational deductive consequence of the hypothesis turning out to be false. This is also known as Duhem's problem (Duhem 1914) (see Duhem Thesis). If a hypothesis H plus statements of initial conditions I plus auxiliary assumptions A plus bridge principles B logically imply an observation report E $((H \wedge I \wedge A \wedge B) \Rightarrow E)$,and E turns out to be false, then what one can conclude is that the four-part conjunction $H \wedge I \wedge A \wedge B$ is false. And the problem is how in general to decide whether the evidence should be counted as telling against the hypothesis H, the statements of initial conditions I, the auxiliary assumptions A, or the bridge principles B.

Clark Glymour (1980) catalogues a number of issues relevant to the assessment of HD (and accounts of confirmation in general) and proposes an alternative deductivist approach to confirmation called "the bootstrap strategy," which attempts to clarify the idea of different parts of a theory and how evidence can bear differently on them. In Glymour's bootstrap account, confirmation is a relation among a theory T, a hypothesis H, and evidence expressed as a sentence E, and Glymour gives an intricate explication of the idea that "E confirms H with respect to T," an explication that is supposed to be sensitive especially to the idea that evidence can be differently confirmationally relevant to different hypotheses that are parts of a complex theory.

Probabilistic Approaches

One influential probabilistic approach to various issues in confirmation theory is called Bayesian confirmation theory (see Bayesianism). The basic idea, on which several refinements may be based, is that evidence E confirms hypothesis H if and only if the conditional probability $P(H \mid E)$ (defined as $P(H \wedge E) \mid P(E)$) is greater than the unconditional probability $P(H)$ and where $P(S)$ is the probability that the statement, or proposition, or claim, or assertion, or sentence S is true (the status of what kind of entity S might be is a concern in metaphysics or the philosophy of language, as well as in the philosophy of science). Disconfirmation is defined by reversing the inequality, and evidential irrelevance is defined by changing the inequality to an equality.

Typically in Bayesian confirmation theory, the function P is taken to be a measure of an agent's subjective probabilities—also called degrees of belief, partial beliefs, or degrees of confidence. Much philosophical work has been done in the area of foundations of subjective probability, the intent being to clarify or operationalize the idea that agents (e.g., scientists) have (or should have only) more or less strong or weak beliefs in propositions, rather than simply adopting the attitudes of acceptance or rejection of them. One approach, called the *Dutch book argument*, attempts to clarify the idea of subjective probability in terms of odds that one is willing to accept when betting for or against the truth or falsity of propositions (see Dutch Book Argument). Another approach, characterized as a decision theoretical approach, assumes various axioms regarding rational preference (transitivity, asymmetry, etc.) and some structural conditions (involving the richness of the set of propositions, or acts, states, and outcomes, considered by an agent) and derives, from preference data, via representation theorems, a probability assignment P and a desirability (or utility) assignment DES such that an agent prefers an item A to an item B if and only if some kind of expected utility of A is numerically greater than the expected utility of B, when the expected utilities are calculated in terms of the derived P and DES functions (see Decision Theory). Various formulas for expected utility have been proposed. (Important work in foundations of subjective probability include Ramsey 1931; de Finetti 1937; Savage [1954] 1972; Jeffrey [1965] 1983; and Joyce 1999.)

Where P measures an agent's subjective degrees of belief, $P(H|E)$ is supposed to be the agent's degree of belief in H on the assumption that E is true, or the degree of belief that the agent would have in H were the agent to learn that E is true. If a person's degree of belief in H would increase if E were learned, then it is natural to say that for this agent, E is positively evidentially relevant to H, even when E is not in fact learned. Of course, different people will have different subjective probabilities, or degrees of belief, even if the different people are equally rational, this being due to different bodies of background knowledge or beliefs possessed (albeit possibly equally justifiable or excusable, depending on one's experience), so that in this approach to confirmation theory, confirmation is a relation among three things: evidence, hypothesis, and background knowledge. The reason this approach is called Bayesian is because of the use that is sometimes made of a mathematical theorem discovered by Thomas Bayes (Bayes 1764), a simple version of which is $P(H \mid E) = P(E \mid H)P(H)/P(E)$. This is significant because it is sometimes easier to figure out the probability of an evidence statement conditional on a hypothesis than it is to figure out the probability of a hypothesis on the assumption that the evidence statement is true (for example, when the hypothesis is statistical and the evidence statement reports the outcome of an experiment to which the hypothesis applies). Bayes's theorem can be used to link these two converse conditional probabilities when the priors, $P(H)$ and $P(E)$, are known (see Bayesianism).

Bayesian confirmation theory not only provides a qualitative definition of confirmation, disconfirmation, and evidential irrelevance, but also suggests measures of degree of evidential support. The most common is the *difference measure*: $d(H,E) = P(H \mid E) - P(H)$, where confirmation, disconfirmation, and evidential irrelevance correspond to whether this measure is greater than, less than, or equal to 0, and the degree is measured by the magnitude of the difference. Another commonly used measure is the *ratio measure*: $r(H,E) = P(H \mid E)/P(H)$ where confirmation, disconfirmation, and evidential irrelevance correspond to whether this measure is greater than, less than, or equal to 1, and the degree is measured by the magnitude of the ratio. Other measures have been defined as functions of likelihoods, or converse conditional probabilities, $P(E|H)$. One application of the idea of degree of evidential support has been in the Ravens paradox, discussed above. Such definitions of degree of evidential support provide a framework within which one can clarify intuitions that under certain conditions (contrapositive instances of the hypothesis that all ravens are black [i.e., nonblack nonravens]) confirm the hypothesis that all ravens are black, but to a minuscule degree compared with positive instances (black ravens).

What follows is a little more detail about the application of Bayesian confirmation theory to the positive-instance criterion and the Ravens paradox. (See Eells 1982 for a discussion and references.) Let H be the hypothesis that all ravens are black; let R_A symbolize the statement that object A is a raven; and let B_A symbolize the statement that object A is black. It can be shown that if H is probabilistically independent of R_A, (i.e., $P(H \mid R_A) = P(H)$), then a positive instance (or report of one), $R_A \wedge B_A$, of H confirms H in the Bayesian sense (i.e., $P(H \mid R_A \wedge B_A) > P(H)$) if and only if $P(B_A \mid R_A) < 1$ (which latter inequality can naturally be interpreted as saying that it was not *already* certain that A would be black if a raven). Further, on the same independence assumption, it can be shown that a positive instance, $R_A \wedge B_A$, confirms H more than a contrapositive instance, $\neg B_A \wedge \neg R_A$ if and only if $P(B_A \mid R_A) < P(\neg R_A \mid \neg B_A)$.

What about the assumption of probabilistic independence of H from R_A? I. J. Good (1967) has proposed counterexamples to the positive-instance criterion like the following. Suppose it is believed that *either* (1) there are just a few ravens in the world and they are all black *or* (2) there are lots and lots of ravens in the world, a very, very few of which are nonblack. Observation of a raven, even a black one (hence a positive instance of H), would tend to support supposition (2) against supposition 1 and thus undermine H, so that a positive instance would disconfirm H. But in this case the independence assumption does not hold, so that Bayesian confirmation theory can help to isolate the kinds of situations in which the positive-instance criterion holds and the kinds in which it may not.

Bayesian confirmation theory can also be used to assess Hempel's proposed conditions of adequacy for criteria of confirmation. Recall, for example, his special-consequence condition, discussed above: If E confirms H and H logically entails H', then E must also confirm H'. It is a theorem of probability theory that if H logically entails H', then $P(H')$ is at least as great as $P(H)$. If the inequality is strict, then an E can increase the probability of H while decreasing the probability of H', consistent with H logically entailing H'. This fact can be used to construct intuitively compelling examples of an H entailing an H' and an E confirming the H while disconfirming the H', telling against the special-consequence condition and also in favor of Bayesian confirmation theory (e.g., Eells 1982).

Some standard objections to Bayesian confirmation theory are characterized as the *problem of the priors* and the *problem of old evidence*. As to the first, while it is sometimes admitted that it makes sense to assign probabilities to evidence statements E conditional on some hypothesis H (even in the absence of much background knowledge), it is objected that it often does not make sense to assign unconditional, or "prior," probabilities to hypothesis H or to evidence statements (reports of observation) E. If H is a newly formulated physical hypothesis, for example, it is hard to imagine what would *justify* an assignment of probability to it prior to evidence—but that is just what the suggested criterion of confirmation, Bayes's theorem, and the measures of confirmation described above require. Such issues make some favor a likelihood approach to the evaluation of evidence—Edwards (1972) and Royall (1997), for instance, who represent a different approach and tradition in the area of statistical inference. According to one formulation of the likelihood account, an E confirms an H more than the E confirms an H' if and only if $P(E|H)$ is greater than $P(E|H')$. This is a comparative principle, an approach that separates the question of which hypothesis it is more justified to believe given the evidence (or the comparative acceptability of hypotheses given the evidence) from the question of what the comparative significance is of evidence for one hypothesis compared with the evidence's significance for another hypothesis. It is the latter question that the likelihood approach actually addresses, and it is sometimes suggested that the degree to which an E supports an H compared with the support of E for an H' is measured by the likelihood ratio, $P(E \mid H)/P(E \mid H')$. Also, likelihood measures of *degree* of confirmation of *single* hypotheses have been proposed, such as $L(H, E) = P(E \mid H)/P(E \mid \neg H)$, or the log of this ratio. (see Fitelson 2001 and Forster and Sober 2002 for recent discussion and references.)

Another possible response to the problem of priors is to point to convergence theorems (as in de Finetti 1937) and argue that initial settings of priors does not matter in the long run. According to such theorems, if a number of agents set different priors, are exposed to the same series of evidence, and update their subjective probabilities (degrees of belief) in certain ways, then, almost certainly, their subjective probabilities will eventually converge on each other and, under certain circumstances, upon the truth.

The problem of old evidence (Glymour 1980; Good 1968, 1985) arises in cases in which $P(E) = 1$. It is a theorem of probability theory that in such cases $P(H|E) = P(H)$, for any hypothesis H, so that in the Bayesian conception of confirmation as formulated above, an E with probability 1 cannot confirm any hypothesis H. But this seems to run against intuition in some cases. An often-cited such case is the confirmation that Albert Einstein's general theory of relativity apparently was informed by already known facts about the behavior of the perihelion of the planet Mercury. One possible Bayesian solution to the problem, suggested by Glymour (1980), would be to say that it is not the already known E that confirms the H after all, but rather a newly discovered logical or explanatory relation between the H and the E. Other solutions have been proposed, various versions of the problem have been distinguished (see Earman 1992 for a discussion), and the problem remains one of lively debate.

ELLERY EELLS

References

Bayes, Thomas (1764), "An Essay Towards Solving a Problem in the Doctrine of Chance," *Philosophical Transactions of the Royal Society of London* 53: 370–418.

de Finetti, Bruno (1937), "La prévision: Ses lois logiques, ses sources subjectives," *Annales de l'Institute Henri Poincaré* 7: 1–68.

Duhem, Pierre (1914), *The Aim and Structure of Physical Theory*. Princeton, NJ: Princeton University Press.

Earman, John (1992), *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA, and London: MIT Press.

Edwards, A. W. F. (1972), *Likelihood*. Cambridge: Cambridge University Press.

Eells, Ellery (1982), *Rational Decision and Causality*. Cambridge and New York: Cambridge University Press.

Fitelson, Branden (2001), *Studies in Bayesian Confirmation Theory*. Ph.D. dissertation, University of Wisconsin–Madison.

Forster, Malcolm, and Elliott Sober (2002), "Why Likelihood?" in M. Taper and S. Lee (eds.), *The Nature of Scientific Evidence*. Chicago: University of Chicago Press.

Glymour, Clark (1980), *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Good, I. J. (1967), "The White Shoe Is a Red Herring," *The British Journal for the Philosophy of Science* 17: 322.

——— (1968), "Corroboration, Explanation, Evolving Probability, Simplicity, and a Sharpened Razor," *British Journal for the Philosophy of Science* 19: 123–143.

——— (1985), "A Historical Comment Concerning Novel Confirmation," *British Journal for the Philosophy of Science* 36: 184–186.

Goodman, Nelson (1965), *Fact, Fiction, and Forecast*. New York: Bobbs-Merrill.

Hempel, Carl G. ([1945] 1965), "Studies in the Logic of Confirmation," in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, 3–51. Originally published in *Mind* 54: 1–26 and 97–121.

Jeffrey, Richard C. ([1965] 1983), *The Logic of Decision*. Chicago and London: University of Chicago Press.

Joyce, James M. (1999), *The Foundations of Causal Decision Theory*. Cambridge and New York: Cambridge University Press.

Nicod, Jean (1930), *Foundations of Geometry and Induction*. New York and London: P. P. Wiener.

Quine, Willard Van (1969), "Natural Kinds," in *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Ramsey, Frank Plumpton (1931), "Truth and Probability," in R. B. Braithwaite (ed.), *The Foundations of Mathematics and Other Logical Essays*. London: Routledge and Kegan Paul, 156–198.

Royall, Richard (1997), *Statistical Evidence: A Likelihood Paradigm*. Boca Raton, FL: Chapman and Hall.

Savage, Leonard ([1954] 1972), *The Foundations of Statistics*. New York: Dover Publications.

See also **Bayesianism; Carnap, Rudolf; Decision Theory; Dutch Book Argument; Epistemology; Hempel, Carl Gustav; Induction, Problem of; Inductive Logic; Probability**

CONNECTIONISM

Connectionist models, also known as models of parallel distributed processing (PDP) and artificial neural networks (ANN), have merged into the mainstream of cognitive science since the mid-1980s. Connectionism currently represents one of two dominant approaches (symbolic modeling is the other) within artificial intelligence used to develop computational models of mental processes.

Unlike symbolic modeling, connectionist modeling also figures prominently in computational neuroscience (where the preferred term is *neural network modeling*).

Connectionist modeling first emerged in the period 1940–1960, as researchers explored possible approaches to using networks of simple neurons to perform psychological tasks, but it fell into decline when limitations of early network designs became apparent around 1970. With the publication of the PDP Research Group volumes (Rumelhart, McClelland, et al. 1986; McClelland, Rumelhart, et al. 1986), connectionism was rescued from over a decade of popular neglect, and the way was opened for a new generation to extend the approach to fresh explanatory domains. (For a collection of historically significant papers from these neglected years and before, see Anderson and Rosenfeld 1988; Anderson, Pellionisz, and Rosenfeld 1990.) Despite some early claims that connectionism constitued a new, perhaps revolutionary, way of understanding cognition, the veritable flood of network-based research has ultimately occurred side by side with other, more traditional, styles of modeling and theoretical frameworks.

The renaissance in connectionist modeling is a result of convergence from many different fields. Mathematicians and computer scientists attempt to describe the formal, mathematical properties of abstract network architectures. Psychologists and neuroscientists use networks to model behavioral, cognitive, and biological phenomena. Roboticists also make networks the control systems for many kinds of embodied artificial agents. Finally, engineers employ connectionist systems in many industrial and commercial applications. Research has thus been driven by a broad spectrum of concerns, ranging from the purely theoretical to problem-solving applications for problems in various scientific domains to application-based or engineering needs.

Given these heterogeneous motivations, and the recent proliferation of network models, analytic techniques, applications, and theories, it is appropriate to ask whether connectionism constitutes a coherent research program or is instead primarily a modeling tool. Following Lakatos, connectionism could be construed as a research program involving a set of core theoretical principles about the role of networks in explaining and understanding cognition, a set of positive and negative heuristics that guide research, an ordering of the important commitments of connectionist modeling, and a set of principles and strategies dictating how recalcitrant empirical results are to be accounted for (see Lakatos, Imre; Research Programs). The greater the disunity in these factors, the less connectionism resembles a research program, and the more it appears to be a convenient tool for modeling certain phenomena. If it is a modeling tool, connectionism need not commit modelers to having anything in common beyond their use of the particular mathematical and formal apparatus itself.

This article will briefly describe the features of prevalent connectionist architectures and discuss a number of challenges to the use of these models. One challenge comes from symbolic models of cognition, which present an alternative representational framework and set of processing assumptions. Another comes from a purportedly nonrepresentational framework, that of nonlinear dynamical systems theory. Finally, there is the neuroscientific challenge to the disciplinary boundaries drawn around "cognitive" modeling by some connectionist psychologists. The status of connectionism is assessed in light of these challenges.

The Properties of Connectionist Models

Connectionist networks are built up from basic computational elements called units or nodes, which are linked to each other via weighted connections, called simply weights. Units take on a variable numerical level of activation, and they pass activation to each other via the weights. Weights determine how great an effect one unit has on other units. This effect may be positive (excitatory) or negative (inhibitory). The net input a unit receives at a time is the weighted sum of the activations on all of the units that are active and connected to it. Given the net input, an activation function (often nonlinear or imposing a threshold) determines the activation of the unit. In this way, activation is passed in parallel throughout the network. Connectionist networks compute functions by mapping vectors of activation values onto other such vectors.

Multilayer feedforward networks are the most intensively studied and widely used class of contemporary models. Units are arranged into layers, beginning with an input layer and passing through a number of intermediate hidden layers, terminating with an output layer. There are no reciprocal connections, so activation flows unidirectionally through the network. The modeler assigns representational significance to the activation vectors at the input and output layers of a network, thereby forging the link between the model and the cognitive task to be explained.

For example, Sejnowski and Rosenberg's NETtalk architecture is designed to map graphemic inputs (letters) onto phonemic outputs. It consists of an input layer of seven groups of 29 units, a hidden layer of 80 units, and an output layer of 26 units. Each layer is completely connected to the next one. Vectors of activity in the network represent aspects of the letter-reading task. The input groups are used to represent the seven letters of text that the network is perceiving at a time, and the output layer represents the phoneme corresponding to the fourth letter in the input string. The task of the network is to pronounce the text correctly, given the relevant context. When the values of the weights are set correctly, the network can produce the appropriate phoneme-representing vectors in response to the text-representing vectors.

Simple networks can be wired by hand to compute some functions, but in networks containing hundreds of units, this is impossible. Connectionist systems are therefore usually not programmed in the traditional sense, but are trained by fixing a learning rule and repeatedly exposing the network to a subset of the input-output mappings it is intended to learn. The rule then systematically adjusts the network's weights until its outputs are near the target output values. One way to classify learning rules is according to whether they require an external trainer. Unsupervised learning (e.g., Hebbian learning) does not require an external trainer or source of error signals. Supervised learning, on the other hand, requires that something outside the network indicate when its performance is incorrect. The most popular supervised learning rule currently in use is the backpropagation rule.

In backpropagation learning, the network's weights are initially set to random values (within certain bounds). The network is then presented with patterns from the training environment. Given random weights, the network's response will likely be far from the intended mapping. The difference between the output and the target is computed by the external trainer and used to send an error signal backward through the network. As the signal propagates, the weights between each layer are adjusted by a slight amount. Over many training cycles, the network's performance gradually approaches the target. When the output is within some criterion distance of the target, training ceases. Since the error is being reduced gradually, backpropagation is an instance of a gradient-descent learning algorithm.

Backpropagation-trained networks have been successful at performing in many domains, including past-tense transformation of verbs, generation of prototypes from exemplars, single word reading, shape-from-shading extraction, visual object recognition, modeling deficits arising in deep dyslexia, and more. Their formal properties are well known. However, they suffer from a number of problems. Among these is the fact that learning via backpropagation is extremely slow, and increasing the learning-rate parameter typically results in overshooting the optimum weights for solving the task.

Another problem facing feedforward networks generally is that individual episodes of processing inputs are independent of each other except for changes in weights resulting from learning. But often a cognitive agent is sensitive not just to what it has learned over many episodes, but to what it processed recently (e.g., the words prior to the preceding one). The primary way sensitivity to context has been achieved in feedforward networks has been to present a constantly moving window of input. For example, in NETtalk, the input specified the three phonemes before and three phonemes after the one to be pronounced. But this solution is clearly a kludge and suffers from the fact that it imposes a fixed window. If sensitivity to the item four back is critical to correct performance, the network cannot perform correctly.

An alternative architecture that is increasingly being explored is the simple recurrent network (SRN) (Elman 1991). SRNs have both feedforward and recurrent connections. In the standard model, an input layer sends activity to a hidden layer, which has two sets of outgoing connections: one to other hidden layers and eventually on to the output layer, and another to a specialized context layer. The weights to the units in the context layer enable it to construct a copy of the activity in the hidden units. The activation over these units is then treated as an additional input to the same hidden units at the next temporal stage of processing. This allows for a limited form of short-term memory, since activity patterns that were present during the previous processing cycle have an effect on the next cycle. Since the activity on the previous cycle was itself influenced by that on a yet earlier cycle, this allows for memory extending over several previous processing epochs (although sensitivity to more than one cycle back will be diminished).

Once trained in a variation of backpropagation, many SRNs are able to discover patterns in temporally ordered sets of events. Elman (1991) has trained SRNs on serially presented words in an attempt to teach them to predict the grammatical category of the next word in a sentence. The networks can achieve fairly good performance at this task. Since the networks were never supplied information about grammatical categories, this suggests that they induced a representation of a more abstract similarity among words than was present in the raw training data. There are many other kinds of neural network architecture. (For further details on their properties and applications, see Anderson 1995; Bechtel and Abrahamsen 2002.)

Connectionism and Symbolic Models

Within cognitive science, symbolic models of cognition have constituted the traditional alternative to connectionism. In symbolic models, the basic representational units are symbols having both syntactic form and typically an intuitive semantics that corresponds to the elements picked out by words of natural language. The symbols are discrete and capable of combining to form complex symbols that have internal syntactic structure. Like the symbol strings used in formal logic, these complex symbols exhibit variable binding and scope, function-and-argument structure, cross-reference, and so on. The semantics for complex symbols is combinatorial: The meaning of a complex symbol is determined by the meanings of its parts plus its syntax. Finally, in symbolic models the dynamics of the system are governed by rules that transform symbols into other symbols by responding to their syntactic or formal properties. These rules are intended to preserve the truth of the structures manipulated. Symbolic models are essentially proof-theoretic engines.

Connectionist models typically contain units that do not individually represent lexicalized semantic contents. (What are called *localist* networks are an exception. In these, individual units are interpretable as expressing everyday properties or propositions. See Page 2000.) More commonly, representations with lexicalized content are *distributed* over a number of units in a network (Smolensky 1988). In a distributed scheme, individual units may stand for repeatable but nonlexicalized microfeatures of familiar objects, which are themselves represented by vectors of such features. In networks of significant complexity, it may be difficult, if not impossible, to discern what content a particular unit is carrying.

In networks, there is no clear analog to the symbolicist's syntactic structures. Units acquire and transmit activation values, resulting in larger patterns of coactivation, but these patterns of units do not themselves syntactically compose. Also, there is no clear program/data distinction in connectionist systems. Whether a network is hand-wired or trained using a learning rule, the modifications are changes to the weights between units. The new weight settings determine the future course of activation in the network and simultaneously constitute the data stored in the network. There are no explicitly represented rules that govern the system's dynamics.

Classical theorists (Fodor and Pylyshyn 1988) have claimed that there are properties of cognition that are captured naturally in symbolic models but that connectionist models can capture them only in an ad hoc manner, if at all. Among these properties are the productivity and the systematicity of thought. Like natural language, thought is productive, in that a person can think a potentially infinite number of thoughts. For example, one can think that Walt is an idiot, that Sandra believes that Walt is an idiot, that Max wonders whether Sandra believes that Walt is an idiot, and so on. Thought is also systematic, in that a person who can entertain a thought can also entertain many other thoughts that are semantically related to it (Cummins 1996). If a person can think that Rex admires the butler's courage, that person can also think that the butler admires Rex's courage. Anyone who can think that dogs fear cats can think that cats fear dogs, and so on. Unless each thought is to be learned anew, these capacities need some finite basis.

Symbolicists argue that this basis is compositionality: Thought possesses a combinatorial syntax and semantics, according to which complex thoughts are built up from their constituent concepts, and those concepts completely determine the meaning of a complex thought. The compositionality of thought would explain both productivity and systematicity. Grasping the meaning of a set of primitive concepts and grasping the recursive rules by which they can be combined is sufficient for grasping the infinite number of thoughts that the concepts and rules can generate. Similarly, grasping a syntactic schema and a set of constituent concepts explains why the ability to entertain one thought necessitates the ability to entertain others: Concepts may be substituted into any permissible role slot in the schema.

The challenge symbolicists have put to connectionists is to explain productivity and systematicity in a principled fashion, without merely implementing a symbolic architecture on top of the connectionist network (for one such implementation, see Franklin and Garzon 1990). One option that some connectionists pursue is to deny that thought is productive or systematic, as symbolicists characterize these properties. For example, one might deny that it is possible to think any systematically structured proposition. Although one can compose the symbol string "The blackberry ate the bear," that does not entail that one is able to think such a thought. But clearly much of thought exhibits some degree of productivity and systematicity, and it is incumbent on connectionists to offer some account of how it is achieved.

Connectionists have advanced a number of proposals for explaining productivity and systematicity. Two of the most widely discussed are Pollack's recursive auto-associative memories (RAAMs) and Smolensky's tensor product networks (Pollack 1990; Smolensky, 1991). RAAMs are easier to understand. An auto-associative network is one trained to produce the same pattern on the output layer as is present on the input layer. If the hidden layer is smaller than the input and output layers, then the pattern produced on the hidden layer is a compressed representation of the input pattern. If the input layer contains three times the number of units as the hidden layer, then one can treat the input activation as consisting of three parts constituting three patterns (e.g., for different words) and recursively compose the hidden pattern produced by three patterns with two new ones. In such an encoding, one is implicitly ignoring the output units, but one can also ignore the input units and supply patterns to the hidden units, allowing the RAAM to generate a pattern on the output units. What is interesting is that even after several cycles of compression, one can, by recursively copying the pattern on one-third of the output units back onto the hidden units, re-create with a fair degree of accuracy the original input patterns.

If RAAMs are required to perform many cycles of recursive encoding, the regeneration of the original pattern may exhibit errors. But up to this point, the RAAM has exhibited a degree of productivity by composing representations of complex structures from representations of their parts. One can also use the compressed patterns in other processing (e.g., to train another feedforward network to construct the compressed representation of a passive sentence from the corresponding active sentence). RAAMs thus exhibit a degree of systematicity. But the compressed representations are not composed according to syntactic principles. Van Gelder (1990), accordingly, construes them as manifesting functional, not explicit, compositionality.

Symbolicists have rejected such functional compositionality as inadequate for explaining cognition. In many respects, this debate has reached a standoff. In part its resolution will turn on the issue posed earlier: To what degree do humans exhibit productivity and systematicity? But independently of that issue, there are serious problems in scaling up from connectionist networks designed to handle toy problems to ones capable of handling the sort of problems humans deal with regularly (e.g., communicating in natural languages). Thus, it is not clear whether solutions similar to those employed in RAAMs (or in SRNs, which also exhibit a degree of productivity and systematicity) will account for human performance. (Symbolic models have their own problems in scaling, and so are not significantly better off in practice.)

Connectionism and Dynamical Systems Theory

Although the conflict between connectionists and symbolicists reached a stalemate in the 1990s, within the broader cognitive science community a kind of accord was achieved. Connectionist approaches were added to symbolic approaches as parts of the modeling toolkit. For some tasks, connectionist models proved to be more useful tools than symbolic models, while for others symbolic models continued to be preferred. For yet other tasks, connectionist models were integrated with symbolic models into hybrids.

As this was happening, a new competitor emerged on the scene, an approach to cognition that challenged both symbolic and connectionist modeling insofar as both took seriously that cognitive activity involved the use of some form of representation. Dynamical systems theory suggested that rather than construing cognition as involving syntactic manipulation of representations or processing them through layers of a network, one should reject the notion of representation altogether. The alternative that these critics advanced was to characterize cognitive activity in terms of a (typically small) set of variables and to formulate (typically nonlinear) equations that would relate the values of different variables in terms of how they changed over time. In physics, dynamical systems theory provides a set of tools for understanding the changes over time of such systems of variables. For example, each variable can be construed as defining a dimension in a multidimensional *state space*, and many systems, although starting at different points in this space, will settle onto a fixed point or into a cycle of points. These points are known as *attractors*, and the paths to them as the *transients*. The structure of the state space can often be represented geometrically—for example, by showing different basins, each of which represents starting states that will end up in a different attractor.

Connectionist networks, especially those employing recurrent connections, are dynamical systems, and some connectionist modelers have embraced the tools of dynamical systems theory for describing their networks. Elman (1991), for example, employs such tools to understand how networks manage to learn to respect syntactic categories in processing streams of words. Others have made use of some of the more exotic elements in dynamical systems theory, such as activation functions that exhibit deterministic chaos, to develop new classes of networks that exhibit more complex and interesting patterns of behavior than simpler networks (e.g., jumping intermittently between two competing interpretations of an ambiguous perceptual figure such as the duck-rabbit (see van Leeuwen, Verver, and Brinkers 2000)). Some particularly extreme dynamicists, however, contend that once one has characterized a cognitive system in this way, there is no further point to identifying internal states as representations and characterizing changes as operations on these representations. Accordingly, they construe dynamical systems theory as a truly radical paradigm shift for cognitive science (van Gelder 1995).

This challenge has been bolstered by some notable empirical successes in dynamical systems modeling of complex cognition and behavior. For example, Busemeyer and Townsend (1993) offer a model of decision under uncertainty, decision field theory, that captures much of the empirical data about the temporal course of decision making using difference equations containing only seven parameters for psychological quantities such as attention weight, valence, preference, and so on. Connectionist models, by contrast, have activation state spaces of as many dimensions as they have units. There are dynamical systems models of many other phenomena, including coordination of finger movement, olfactory perception, infants' stepping behavior, the control of autonomous robot agents, and simple language processing. The relative simplicity and comprehensibility of their models motivates the antirepresentational claims advanced by dynamical systems theorists.

There is reason, though, to question dynamical systems theory's more radical challenge to cognitive modeling. One can accept the utility of characterizing a system in terms of a set of equations and portraying its transitions through a multidimensional state space without rejecting the utility of construing states within the system as representational and the trajectories through state space in terms of transitions between representations. This is particularly true when the system is carrying out what one ordinarily thinks of as a complex reasoning task such as playing chess, where the task is defined in terms of goals and the cognizer can be construed as considering different possible moves and their consequences. To understand why a certain system is able to play chess successfully, rather than just recognizing that it does, the common strategy is to treat some of its internal states as representing goals or possible moves. Moreover, insofar as these internal states are causally connected in appropriate ways, they do in fact carry information (in what is fundamentally an informational-theoretic sense, in which the state covaries with referents external to the system), which is then utilized by other parts of the system. In systems designed to have them, these information-carrying states may arise without their normal cause (as when a frog is subjected to a laboratory with bullets on a string moving in front of its eyes). In these situations the system responds as it would if the state were generated by the cause for which the response was designed or selected. Such internal states satisfy a common understanding of what a representation is, and the ability to understand why the system works successfully appeals to these representations. If this construal is correct, then dynamical systems theory is not an alternative to connectionist modeling, but should be construed as an extension of the connectionist toolkit (Bechtel and Abrahamsen 2002).

Connectionism and Neuroscience

Connectionist models are often described as being *neurally inspired*, as the term "artificial neural networks" implies. More strongly, many connectionists have claimed that their models enjoy a special sort of *neural plausibility*. If there is a significant similarity between the processing in ANNs and the activity in real networks of neurons, this might support connectionism for two reasons. First, since the cognitive description of the system closely resembles the neurobiological description, it seems that connectionist models in psychology have a more obvious account about how the mind might supervene on the brain than do symbolic models (see Supervenience). Second, connectionist models might provide a fairly direct characterization of the functioning of the neural level itself (e.g., the particular causal interactions among neurons). This functioning is not easily revealed by many standard neuroscience methods. For example, localization of mental activity via functional magnetic resonance imaging can indicate which brain regions are preferentially activated during certain tasks, but this alone does not give information about the specific computations being carried out within those regions. Connectionist modeling of brain function thus might supplement other neurobiological techniques.

This strategy can be illustrated within the domain of learning and memory. Neuropsychological studies suggest that the destruction of structures in the medial temporal lobes of the brain, especially the hippocampus, results in a characteristic pattern of memory deficits. These include (i) profound anterograde amnesia for information presented in declarative or verbal form, such as arbitrary paired associates, as well as memory for particular experienced events more generally (episodic memory); (ii) preserved implicit memory for new information, such as gradually acquired perceptual-motor skills; and (iii) retrograde amnesia for recently acquired information, with relative preservation of memories farther back in time. This triad of deficits was famously manifested by H.M., a patient who underwent bilateral removal of sections of his medial temporal lobes in the early 1950s in order to cure intractable epilepsy. Since H.M., this pattern of deficits has been confirmed in other human and animal studies.

McClelland, McNaughton, and O'Reilly (1995) offer an explanation of these deficits based on connectionist principles. One feature of backpropagation-trained networks is that once they are trained on one mapping, they cannot learn another mapping without "unlearning" the previously stored knowledge. This phenomenon is known as catastrophic interference. However, catastrophic interfence can be overcome if, rather than fully training a network on one mapping, then training it on another, the training sets are interleaved so that the network is exposed to both mappings in alternation. When the training environment is manipulated in this way, the network can learn both mappings without overwriting either one.

As a model of all learning and memory, this technique suffers from being slow and reliant on a fortunate arrangement of environmental contingencies. However, McClelland et al. (1995) suggest that learning in the neocortex may be characterized by just such a process, if there is a neural mechanism that stores, organizes, and presents appropriately interleaved stimuli to it. They conjecture that this is the computational function of the hippocampus. Anatomically, the hippocampus receives convergent inputs from many sensory centers and has wide-ranging efferent connections to the neocortex. The pattern of deficits resulting from hippocampal lesions could be explained on the assumption that the hippocampus has a method of temporarily storing associations among stimuli without catastrophic interference. Ablation of the hippocampus results in anterograde amnesia for arbitrary associations and declarative information because the neocortex alone is incapable of learning these without the appropriately interleaved presentation. Implicit learning is preserved because it does not require rapid integration of many disparate representations into a single remembered experience; further, it typically takes many trials for mastery, as is also the case with backpropagation learning. Finally, the temporally graded retrograde amnesia is explained by the elimination of memory traces that are temporarily stored in the hippocampus itself. Older memories have already been integrated into the neocortex, and hence are preserved.

McClelland et al. (1995) did not model the hippocampus directly; rather, they implemented it as a black box that trained the neocortical network according to the interleaving regimen. Others have since presented more elaborate models. Murre, Graham, and Hodges (2001) have implemented a system called TraceLink that features a network corresponding to a simplified single-layer hippocampus, the neocortex, and a network of neuromodulatory systems (intended to correspond to the basal forebrain nuclei). TraceLink accounts for the data reviewed by McClelland et al. (1995), and also predicts several phenomena associated with semantic dementia. Other models incorporating a more elaborate multilayer hippocampal network have been presented by Rolls and Treves (1998) and O'Reilly and Rudy (2001). These models collectively support the general framework set out by McClelland et al. (1995) concerning the computational division of labor between the hippocampus and the neocortex in learning and memory.

These studies of complementary learning systems suggest a useful role for network-based modeling in neuroscience. However, this role is presently limited in several crucial respects. The models currently being offered are highly impoverished compared with the actual complexity of the relevant neurobiological structures. Assuming that these networks are intended to capture neuron-level interactions, they are several orders of magnitude short of the number of neurons and connections in the brain. Further, the backpropagation rule itself is biologically implausible if interpreted at the neuronal level, since it allows individual weights to take on either positive or negative values, while actual axonal connections are either excitatory or inhibitory, but never both. There is no network model that captures all of the known causal properties of neurons, even when the presence of glial cells, endocrine regulators of neural function, and other factors are abstracted away.

A common response to these objections is to interpret networks as describing only select patterns of causal activity among large populations of neurons. In many cases, this interpretation is appropriate. However, there are many kinds of network models in neuroscience, and they can be interpreted as applying to many different levels of organization within the nervous system, including individual synaptic junctions on dendrites, particular neurons within cortical columns, and interactions at the level of whole neural systems. No single interpretation appears to have any special priority over the others. The specific details of the network architecture are dictated in most cases by the particular level of neural analysis being pursued, and theorists investigate multiple levels simultaneously. There are likely to be at least as many distinct kinds of possible connectionist models in neuroscience as there are distinct levels of generalization within the nervous system.

Conclusions

At the beginning of this article, the question was asked whether connectionism is best thought of as a research program, a modeling tool, or something in between. Considering the many uses to which networks have been put, and the many disciplines that have been involved in cataloguing their properties, it seems unlikely that there will be any common unity of methods, heuristics, principles, etc., among them. Ask whether a neuroscientist using networks to model the development of receptive fields in the somatosensory system would have anything in common with a programmer training a network to take customers' airline reservations. Each of these might be a neural network theorist, despite having nothing significant in common besides the formal apparatus they employ. Across disciplines, then, connectionism lacks the characteristic unity one would expect from a research program.

The question may be asked again at the level of each individual discipline. This article has not surveyed every field in which networks have played a significant role but has focused on their uses in artificial intelligence, psychology, and neuroscience. Even within these fields, the characteristic unity of a research program is also largely absent, if one takes into account the diverse uses that are made of networks.

In psychology, for instance, there are some connectionists who conceive of their models as providing a theory of how mental structures might functionally resemble, and therefore plausibly supervene on, the organization of large-scale neuronal structures (see Psychology). However, there are just as many theorists who see their work as being only, in some quite loose sense, neurally inspired. The organization of the NETtalk network is not particularly neurally plausible, since it posits a simple three-layered linear causal process leading from the perception of letters to the utterance of phonemes. Being a connectionist in psychology does not appear to require agreement on the purpose of the models used or the possible data (e.g., neurobiological) that might confirm or disconfirm them. This is what one might expect of a tool rather than a research program.

It is perhaps ironic that this state of affairs was predicted by Rumelhart, one of the theorists who revitalized connectionism during the 1980s. In a 1993 interview, Rumelhart claimed that as networks become more widely used in a number of disciplines, "there will be less and less of a core remaining for neural networks per se and more of, 'Here's a person doing good work in [her] field, and [she's] using neural networks as a tool' " (Anderson and Rosenfeld 1998, 290). Within the fields, in turn, network modeling will "[d]isappear as an identifiable separate thing" and become "part of doing science or doing engineering" (291). Such a disappearance, however, may not be harmful. Connectionist networks, like other tools for scientific inquiry, are to be evaluated by the quality of the results they produce. In this respect, they have clearly proven themselves a worthy, and sometimes indispensible, component of research in an impressive variety of disciplines.

Dan Wieskopf
William Bechtel

References

Anderson, James A. (1995), *An Introduction to Neural Networks*. Cambridge, MA: MIT Press.

Anderson, James A., and Edward Rosenfeld (eds.) (1988), *Neurocomputing: Foundations of Research*, vol. 1. Cambridge, MA: MIT Press.

——— (1998), *Talking Nets: An Oral History of Neural Networks*. Cambridge, MA: MIT Press.

Anderson, James A., Andras Pellionisz, and Edward Rosenfeld (eds.) (1990), *Neurocomputing: Directions for Research*, vol. 2. Cambridge, MA: MIT Press.

Bechtel, William, and Adele Abrahamsen (2002), *Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks*. Oxford: Basil Blackwell.

Busemeyer, Jerome R., and James T. Townsend (1993), "Decision Field Theory: A Dynamic-Cognitive

Approach to Decision Making in an Uncertain Environment," *Psychological Review* 100: 423–459.
Cummins, Robert (1996), "Systematicity," *Journal of Philosophy* 93: 591–614.
Elman, Jeffrey L. (1991), "Distributed Representations, Simple Recurrent Networks, and Grammatical Structure," *Machine Learning* 7: 195–225.
Fodor, Jerry A., and Zenon Pylyshyn (1988), "Connectionism and Cognitive Architecture: A Critical Analysis," *Cognition* 28: 3–71.
Franklin, Stan, and Max Garzon (1990), "Neural Computability," in O. M. Omidvar (ed.), *Progress in Neural Networks*, vol. 1. Norwood, NJ: Ablex, 127–145.
McClelland, James L., Bruce L. McNaughton, and Randall C. O'Reilly (1995), "Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory," *Psychological Review* 102: 419–457.
McClelland, James L., David E. Rumelhart, and the PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 2. Cambridge, MA: MIT Press.
Murre, Jacob, Kim S. Graham, and John R. Hodges (2001), "Semantic Dementia: Relevance to Connectionist Models of Long-Term Memory," *Brain* 124: 647–675.
O'Reilly, Randall C., and J. W. Rudy (2001), "Conjunctive Representations in Learning and Memory: Principles of Cortical and Hippocampal Function," *Psychological Review* 108: 311–345.
Page, M. (2000), "Connectionist Modeling in Psychology: A Localist Manifesto," *Behavioral and Brain Sciences* 23: 443–512.
Pollack, Jordan B (1990), "Recursive Distributed Representations," *Artificial Intelligence* 46: 77–105.
Rolls, Edmund T., and Alessandro Treves (1998), *Neural Networks and Brain Function*. Oxford: Oxford University Press.
Rumelhart, David E., James L. McClelland, and the PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1. Cambridge, MA: MIT Press.
Smolensky, Paul (1988), "On the Proper Treatment of Connectionism," *Behavioral and Brain Sciences* 11: 1–74.
——— (1991), "Connectionism, Constituency, and the Language of Thought," in Barry Loewer and Georges Rey (eds.), *Meaning in Mind: Fodor and His Critics*. Oxford: Basil Blackwell.
van Gelder, Timothy (1990), "Compositionality: A Connectionist Variation on a Classical Theme," *Cognitive Science* 14: 355–384.
——— (1995), "What Might Cognition Be, If Not Computation?" *Journal of Philosophy* 111: 345–381.
van Leeuwen, Cees, S. Verver, and M. Brinkers (2000), "Visual Illusions, Solid/Outline Invariance and Non-Stationary Activity Patterns," *Connection Science* 12: 279–297.

See also **Artificial Intelligence; Cognitive Science; Neurobiology; Psychology, Philosophy of; Supervenience**

CONSCIOUSNESS

Consciousness is extremely familiar, yet it is at the limits—beyond the limits, some would say—of what one can sensibly talk about or explain. Perhaps this is the reason its study has drawn contributions from many fields, including psychology, neuroscience, philosophy, anthropology, cultural and literary theory, artificial intelligence, physics, and others. The focus of this article is on the varieties of consciousness, different problems that have been raised about these varieties, and prospects for progress on these problems.

Varieties of Consciousness

Creature Versus State Consciousness

One attributes consciousness both to people and to their psychological states. An agent can be conscious (as opposed to unconscious), and that agent's desire for a certain emotional satisfaction might be unconscious (as opposed to conscious). Rosenthal (1992) calls the former creature consciousness and the latter state consciousness. Most, but not all, discussion of consciousness in the contemporary literature concerns state consciousness rather than creature consciousness. Rosenthal (1992) goes on to propose an explanation of state consciousness in terms of creature consciousness, according to which a state is conscious just in case an agent who is in the state is conscious of it—but this proposal has proved controversial (e.g., Dretske 1993).

Essential Versus Nonessential Consciousness

Focusing on conscious states, one may distinguish those that are essentially conscious from those that are (or might be) conscious but not essentially so. The distinction is no doubt vague, but, to a first approximation, a state is essentially conscious just in case being in the state entails that it is conscious, and is not essentially conscious just in case this is not so.

Sensations are good candidates for states that are essentially conscious. If an agent is in pain, this state would seem to be conscious. (This is not to deny that the agent might fail to attend to it.) Beliefs, knowledge, and other cognitive states are good candidates for states that might be conscious but not essentially so. One may truly say that an agent knows the rules of his language even though this knowledge is unconscious. Perception presents a hard case, as is demonstrated by the phenomenon of blindsight, in which subjects report that they do not see anything in portions of the visual field and yet their performance on forced-choice tasks suggests otherwise (Weiskrantz 1986). Clearly *some* information processing is going on in such cases, but it is not obvious that what is going on is properly described as perception, or at least as perceptual experience. It is plausible to suppose that indecision about how to describe matters here derives in part from indecision about whether perceptual states or experiences are essentially conscious.

Transitive Versus Intransitive Consciousness

In the case of creature consciousness, one may speak of someone's being conscious *simpliciter* and of someone's being conscious *of* something or other. Malcolm calls the first "intransitive" and the second "transitive" consciousness (e.g., Armstrong and Malcolm 1984). To say that a person is conscious *simpliciter* is a way of saying that the person is awake or alert. So the study of creature intransitive consciousness may be assimilated to the study of what it is to be alert. The denial of consciousness *simpliciter* does not entail a denial of psychological states altogether. If an agent is fast asleep on a couch, one may truly say both that the agent is unconscious *and* that the agent believes that snow is white. Humphrey (1992) speculates that the *notion* of intransitive consciousness is a recent one, perhaps about 200 years old, but presumably people *were* on occasion intransitively conscious (i.e., alert or awake) prior to that date.

To say that a person is conscious of something seems to be a way of saying that the person knows or has beliefs about that thing. To say that one is conscious of a noise overhead is to say that one knows there is a noise overhead, though perhaps with the accompanying implication that one knows this only vaguely. So the study of creature transitive consciousness may be assimilated to the study of knowledge or beliefs. It is sometimes suggested that "consciousness" and "awareness" are synonyms. This is true only on the assumption that what is intended is creature transitive consciousness, since awareness is always *by* someone *of* something.

Intentional Versus Nonintentional Consciousness

While the transitive/intransitive distinction has no obvious analogue in the case of state consciousness—a state is not *itself* awake or alert, nor is it aware of anything—a related distinction is that between intentional and nonintentional conscious states. An intentional conscious state is *of* something in the sense that it represents the world as being a certain way—such states exhibit "intentionality," to adopt the traditional word. A nonintentional conscious state is a state that does not represent the world as being in some way. It is sometimes suggested that bodily sensations (itches, pains) are states of this second kind, while perceptual experiences (seeing a blue square on a red background) are cases of the first. But the matter is controversial given that to have a pain in one's foot seems to involve among other things representing one's foot as being in some condition or other, a fact that suggests that even here there is an intentional element in consciousness (see Intentionality).

Phenomenal Versus Access Consciousness

Block (1995) distinguishes two kinds of state consciousness: phenomenal consciousness and access consciousness. The notion of a phenomenally conscious state is usually phrased in terms of "What is it like ... ?" (e.g., Nagel 1974, "What is it like to be a bat?"). For a state to be phenomenally conscious is for there to be something it is akin to being, in that state. In the philosophical literature, the terms "qualia," "phenomenal character," and "experience" are all used as rough synonyms for phenomenal consciousness in this sense, though unfortunately there is no terminological consensus here.

For a state to be access conscious is, roughly, for the state to control rationally, or be poised to control rationally, thought and behavior. For creatures who have language, access consciousness closely correlates with reportability, the ability to express the state at will in language. Block suggests, among other things, that a state can be phenomenally conscious without its being access conscious.

For example, suppose one is engaged in intense conversation and becomes aware only at noon that there is a jackhammer operating outside—at five to twelve, one's hearing the noise is phenomenally conscious, but it is not access conscious. Block argues that many discussions both in the sciences and in philosophy putatively about phenomenal consciousness are in fact about access consciousness. Block's distinction is related to one made by Armstrong (1980) between experience and consciousness. For Armstrong, consciousness is attentional: A state is conscious just in case one attends to it. However, since one can have an experience without attending to it, it is possible to divorce experience and consciousness in this sense.

Within the general concept of access consciousness, a number of different strands may be distinguished. For example, an (epistemologically) normative interpretation of the notion needs to be separated from a nonnormative one. In the former case, the mere fact that one is in an access-conscious state puts one in a position to *know* or *justifiably* believe that one is; in the latter, being in an access-conscious state prompts one to think or believe that one is—here there is no issue of epistemic appraisal (see Epistemology). Further, an actualist interpretation of the notion needs to be separated from a counterfactual one. In the former case, what is at issue is whether one *does* know or think that one is in the state; in the latter, what is at issue is whether one *would*, provided other cognitive conditions were met. Distinguishing these various notions leads naturally into other issues. For example, consider the claim that it is essentially true of all psychological states that if one is in them, one would know that one is, provided one reflects and has the relevant concepts. That is one way of spelling out the Cartesian idea (recently defended by Searle [1992]) that the mind is "transparent" to itself.

There are hints both in Block (1995) and in related discussion (e.g., Davies and Humphreys 1993) that the phenomenal/access distinction is in (perhaps rough) alignment with both the intentional/nonintentional and the essential/nonessential distinction. The general idea is that phenomenally conscious states are *both* essentially so *and* nonintentional, while access-conscious states are neither. In view of the different interpretations of access consciousness, however, it is not clear that this is so. Psychological states might well be both essentially access conscious and phenomenally conscious. And, as indicated earlier, perhaps all conscious states exhibit intentionality in some form or other.

Self-Consciousness

Turning back from state consciousness to creature consciousness, a notion of importance here is self-consciousness, i.e., one's being conscious of oneself as an agent or self or (in some cases) a person. If to speak of a creature's being "conscious" of something is to speak about knowledge or beliefs, to attribute self-consciousness is to attribute to a creature knowledge or beliefs that the creature is a self or an agent. This would presumably require significant psychological complexity and perhaps cultural specificity. Proposals like those of Jaynes (1976) and Dennett (1992)—that consciousness is a phenomenon that emerges only in various societies—are best interpreted as concerning self-consciousness in this sense, which becomes more natural the more one complicates the underlying notion of self or agent. (Parallel remarks apply to any notion of group consciousness, assuming such a notion could be made clear.)

Problems of Consciousness

If these are the varieties of consciousness, it is easy enough to say in general terms what the problems of consciousness are, i.e., to explain or understand consciousness in all its varieties. But demands for explanation mean different things to different people, so the matter requires further examination.

To start with, one might approach the issue from an unabashedly scientific point of view. Consciousness is a variegated phenomenon that is a pervasive feature of the mental lives of humans and other creatures. It is desirable to have an explanation of this phenomenon, just as it is desirable to have an explanation of the formation of the moon, or the origin of HIV/AIDS. Questions that might be raised in this connection, and indeed have been raised, concern the relation of consciousness to neural structures (e.g., Crick and Koch 1998), the evolution of consciousness (e.g., Humphrey 1992), the relation of consciousness to other psychological capacities (e.g., McDermott 2001), the relation of consciousness to the physical and social environment of conscious organisms (e.g., Barlow 1987), and relations of unity and difference among conscious states themselves (e.g., Bayne and Chalmers 2003).

The attitude implicit in this approach—that consciousness might be studied like other empirical phenomena—is attractive, but there are at least four facts that need to be confronted before it can be completely adopted:

1. No framework of ideas has as yet been worked out within which the study of consciousness can proceed. Of course, this does not exclude the possibility that such a framework might be developed in the future, but it does make the specific proposals (such as that of Crick and Koch 1998) difficult to evaluate.
2. In the past, one of the main research tasks in psychology and related fields was to study psychological processes that were not conscious. This approach yielded a number of fruitful lines of inquiry—for example, Noam Chomsky's (1966) idea that linguistic knowledge is to be explained by the fact that people have unconscious knowledge of the rules of their language—but will presumably have to be abandoned when it comes to consciousness itself.
3. As Block (1995) notes, the standard concept of consciousness seems to combine a number of different concepts, which in turn raises the threat that consciousness in one sense will be confused with consciousness in another.
4. The issue of consciousness is often thought to raise questions of a philosophical nature, and this prompts further questions about whether a purely scientific approach is appropriate.

What are the philosophical aspects of the issue of consciousness? In the history of the subject, the issue of consciousness is usually discussed in the context of another, the traditional mind–body problem. This problem assumes a distinction between two views of human beings: the materialist or physicalist view, according to which human beings are completely physical objects; and the dualist view, according to which human beings are a complex of both irreducibly physical and irreducibly mental features. Consciousness, then, emerges as a central test case that decides the issue. The reason is that there are a number of thought experiments that apparently make it plausible to suppose that consciousness is distinct from anything physical. An example is the inverted spectrum hypothesis, in which it is imagined that two people might be identical except for the fact that the sensation provoked in one when looking at blood is precisely the sensation provoked in the other when looking at grass (Shoemaker 1981). If this hypothesis represents a genuine possibility, it is a very short step to the falsity of physicalism, which, setting aside some complications, entails that if any two people are identical physically, they are identical psychologically. On the other hand, if the inverted spectrum is possible, then two people identical physically may yet differ in respect of certain aspects of their conscious experience, and so differ psychologically. In short, physicalists are required to argue that the inverted spectrum hypothesis does not represent a genuine possibility. And this places the issue of consciousness at the heart of the mind–body problem.

In contemporary philosophy of mind, the traditional mind–body problem has been severely criticized. First, most contemporary philosophers do not regard the falsity of physicalism as a live option (Chalmers [1996] is an exception), so it seems absurd to debate something one already assumes to be true. Second, some writers argue that the very notions within which the traditional mind–body problem is formed are misguided (Chomsky 2000). Third, there are serious questions about the legitimacy of supposing that reflection of possible cases such as the inverted spectrum could even in principle decide the question of dualism or materialism, which are apparently empirical, contingent claims about the nature of the world (Jackson 1998).

As a result of this critique, many philosophers reject the mind–body problem in its traditional guise. However, it is mistaken to infer from this that concern with the inverted spectrum and related ideas has likewise been rejected. Instead, the theoretical setting of these arguments has changed. For example, in contemporary philosophy, the inverted spectrum often plays a role not so much in the question of whether physicalism is true, but rather in questions about whether phenomenal consciousness is in principle irreducible or else lies beyond the limits of rational inquiry. The impact of philosophical issues of this kind on a possible science of consciousness is therefore straightforward.

In the philosophical debates just alluded to, the notion of consciousness at issue is phenomenal consciousness. In other areas of philosophy, other notions are more prominent. In epistemology, for example, an important question concerns the intuitive difference between knowledge of one's own mental states—which seems in a certain sense privileged or direct—and knowledge of the external world, including the minds of others. This question has been made more acute by the impact of externalism, the thesis that one's psychological states depend constitutively on matters external to the subject, factors for which direct knowledge is not plausible (Davies 1997). Presumably these issues will be informed by the study of access consciousness. Similarly, in discussions of the notion of personal identity and related questions about how and why persons are objects of special moral concern, the notion of self-consciousness plays an important role. One might also regard both access consciousness and self-consciousness as topics for straightforward scientific study.

Prospects for Progress

Due to the influence of positivist and postpositivist philosophy of science in the twentieth century, it was at one time common to assume that some or all questions of consciousness were pseudo-questions. Recently it has been more common to concede that the questions are real enough. But what are the chances of progress here?

In light of the multifariousness of the issues, a formulaic answer to this question would be inappropriate. Access consciousness seems to be a matter of information processing, and there is reason to suppose that such questions might be addressed using contemporary techniques. Hence, many writers (e.g., Block 1995) find grounds for cautious optimism here, though this might be tempered depending on whether access consciousness is construed as involving a normative element. In the case of self-consciousness, the issue of normativity is also present, and there is the added complication that self-consciousness is partly responsive to questions of social arrangements and their impact on individual subjects.

But it is widely acknowledged that the hardest part of the issue is phenomenal consciousness. Here the dominant strategy has been an indirect one of attempting to reduce the overall number of problems. One way to implement this strategy is to attempt to explain the notion of phenomenal consciousness in terms of another notion, say, access consciousness or something like it. Some philosophers suggest that puzzlement about phenomenal consciousness is a cognitive illusion, generated by a failure to understand the special nature of concepts of phenomenal consciousness, and that once this puzzlement is dispelled, the way will be clear for a straightforward identification of phenomenal and access consciousness (e.g., Tye 1999). Others argue that discussions of phenomenal consciousness neglect the extent to which conscious states involve intentionality, and that once this is fully appreciated there is no bar to adopting the view that phenomenal consciousness is just access consciousness (e.g., Carruthers 2000).

The attractive feature of these ideas is that, if successful, they represent both philosophical and scientific progress. But the persistent difficulty is that the proposed explanations are unpersuasive. It is difficult to rid oneself of the feeling that what is special about concepts of phenomenal consciousness derives from only what it is that they are concepts of, and this makes it unlikely that the puzzles of phenomenal consciousness are an illusion. And, while it is plausible that phenomenally conscious states are intentional, emphasizing this fact will not necessarily shed light on the issue, for the intentionality of phenomenal consciousness might be just as puzzling as phenomenal consciousness itself.

However, even if one agrees that phenomenal consciousness represents a phenomenon distinct from these other notions, and therefore requires a separate approach, there is still a way in which one might seek to implement the strategy of reducing the number of problems, for, as noted earlier, phenomenal consciousness is thought to present both a philosophical and a scientific challenge. But what is the relation between these two issues? It is common to assume that the philosophical problem needs to be removed before one can make progress on the science. But perhaps the reverse is true. If the philosophical problems can be seen to be a reflection partly of ignorance in the scientific domain, there is no reason to regard them as a further impediment to scientific study. This might not seem like much of an advance. But the study of consciousness has been hampered by the feeling that it presents a problem of a different order from more straightforward empirical problems. In this context, to combat that assumption is to move forward, though slowly.

Daniel Stoljar

References

Armstrong, D. (1980), "What Is Consciousness?" in *The Nature of Mind and Other Essays*. Ithaca, NY: Cornell University Press, 55–67.

Armstrong, D., and N. Malcolm (1984), *Consciousness and Causality*. Oxford: Blackwell.

Barlow, H. (1987), "The Biological Role of Consciousness," in C. Blakemore and S. Greenfield (eds.), *Mindwaves*. Oxford: Basil Blackwell.

Bayne, T., and D. Chalmers (2003), "What Is the Unity of Consciousness?" in A. Cleeremans (ed.), *The Unity of Consciousness: Binding, Integration and Dissociation*. Oxford: Oxford University Press.

Block, N. (1995), "On a Confusion About a Function of Consciousness," *Behavioral and Brain Sciences* 18: 227–247.

Chalmers, D. (1996), *The Conscious Mind*. New York: Oxford University Press.

Chomsky, N. (1966), *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

——— (2000), *New Horizons in the Study of Mind and Language*. Cambridge: Cambridge University Press.

Crick, F., and C. Koch (1990), "Towards a Neurobiological Theory of Consciousness," *Seminars in the Neurosciences* 2: 263–275.

Carruthers, P. (2000), *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge: Cambridge University Press.

Davies, M. (1997), "Externalism and Experience," in N. Block, O. Flanagan, and G. Guzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press, 309–328.

Davies, M., and G. Humphreys (1993), "Introduction," in M. Davies and G. Humphreys (eds.), *Consciousness: Psychological and Philosophical Essays*. Oxford: Blackwell.

Dennett , D. (1992), *Consciousness Explained*. Boston: Little, Brown and Co.

Dretske, D. (1993), "Conscious Experience," *Mind* 102: 263–283.

Humphrey, N. (1992), *A History of the Mind*. New York: Simon and Schuster.

Jackson, J. (1998), *From Metaphysics to Ethics*. Oxford: Clarendon.

Jaynes, J. (1976), *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Boston: Houghton Mifflin Co.

McDermott, D. (2001), *Mind and Mechanism*. Cambridge, MA: MIT Press.

Nagel, T. (1974), "What Is It Like to Be a Bat?" *Philosophical Review* 83: 7–22.

Rosenthal, D. (1992), "A Theory of Consciousness," in N. Block, O. Flanagan, and G. Guzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press, 729–754.

Searle, R. (1992), *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Shoemaker, S. (1981), "The Inverted Spectrum," *Journal of Philosophy* 74: 357–381.

Tye, M. (1999), "Phenomenal Consciousness: The Explanatory Gap as a Cognitive Illusion," *Mind* 108: 432.

Weiskrantz, L. (1986), *Blindsight*. Oxford: Oxford University Press.

See also **Cognitive Science; Connectionism; Intentionality; Psychology, Philosophy of**

CONSERVATION BIOLOGY

Conservation biology emerged in the mid-1980s as a science devoted to the conservation of biological diversity, or *biodiversity*. Its emergence was precipitated by a widespread concern that anthropogenic development, especially deforestation in the tropics (Gómez-Pompa, Vázquez-Yanes, and Guevera 1972), had created an extinction crisis: a significant increase in the rate of species extinction (Soulé 1985). From its beginning, the primary objective of conservation biology was the design of conservation area networks (CANs), such as national parks, nature reserves, and managed-use zones that protect areas from anthropogenic transformation.

Conservation biology, then, is a normative discipline, in that it is defined in terms of a practical goal in addition to the accumulation of knowledge about a domain of nature. In this respect, it is analogous to medicine (Soulé 1985). Like medicine, conservation biology performs its remedial function in two ways: through intervention (for example, when conservation plans must be designed for species at risk of imminent extinction) and through prevention (for example, when plans are designed to prevent decline in species numbers long before extinction is imminent).

The normative status of conservation biology distinguishes it from ecology, which is not defined in terms of a practical goal (see Ecology). Moreover, besides using the models and empirical results of ecology, conservation biology also draws upon such disparate disciplines as genetics, computer science, operations research, and economics in designing and implementing CANs. Each of these fields contributes to a comprehensive framework that has recently emerged about the structure of conservation biology (see "The Consensus Framework" below).

Different views about the appropriate target of conservation have generated distinct methodologies within conservation biology. How 'biodiversity' is defined and, correspondingly, what conservation plans are designed will partly reflect ethical views about what features of the natural world are valuable (Norton 1994; Takacs 1996). There exists, therefore, a close connection between conservation biology and environmental ethics.

The Concept of Biodiversity

'Biodiversity' is typically taken to refer to diversity at all levels of biological organization: molecules, cells, organisms, species, and communities (e.g., Meffe and Carroll 1994). This definition does not, however, provide insight into the fundamental goal of conservation biology, since it refers to all biological entities (Sarkar and Margules 2002). Even worse, this definition does not exhaust all items of biological interest that are worth preserving: Endangered biological phenomena, such as the migration of the monarch butterfly, are not included within this definition (Brower and Malcolm 1991). Finally, since even a liberally construed notion of biodiversity does not capture the ecosystem processes that sustain biological diversity, some have argued that a more general concept of biological *integrity*, incorporating both diversity of entities and the ecological processes that sustain them, should be recognized as the proper focus of conservation biology (Angermeier and Karr 1994).

In response to these problems, many conservation biologists have adopted a pluralistic approach to biodiversity concepts (Norton 1994; Sarkar and Margules 2002; see, however, Faith [2003], who argues that this ready acceptance of pluralism confuses the [unified] concept of biodiversity with the plurality of different conservation strategies). Norton (1994), for example, points out that any measure of biodiversity presupposes the validity of a specific model of the natural world. The existence of several equally accurate models ensures the absence of any uniquely correct measure. Thus, he argues, the selection of a biodiversity measure should be thought of as a normative political decision that reflects specific conservation values and goals. Sarkar and Margules (2002) argue that the concept of biodiversity is implicitly defined by the specific procedure employed to prioritize places for conservation action (see "Place Prioritization" below). Since different contexts warrant different procedures, biodiversity should similarly be understood pluralistically.

Two Perspectives

Throughout the 1980s and early 90s, the discipline was loosely characterized by two general approaches, which Caughley (1994) described as the "small-population" and "declining-population" paradigms of conservation. Motivated significantly by the legal framework of the Endangered Species Act of 1973 and the National Forest Management Act of 1976, the small-population paradigm originated and was widely adopted in the United States (Sarkar 2005). It focused primarily on individual species threatened by extinction. By analyzing their distributions and habitat requirements, conservation biologists thought that "minimum viable populations" could be demonstrated, that is, population sizes below which purely stochastic processes would significantly increase the probability of extinction (see "Viability Analysis" below). It quickly became clear, however, that this methodology was inadequate. It required much more data than could be feasibly collected for most species (Caughley 1994) and, more importantly, failed to consider the interspecies dynamics essential to most species' survival (Boyce 1992).

Widely followed in Australia, the declining-population paradigm focused on deterministic, rather than stochastic, causes of population decline (Sarkar 2005). Unlike the small-population paradigm, its objective was to identify and eradicate these causes before stochastic effects became significant. Since the primary cause of population decline was, and continues to be, habitat loss, conservation biologists, especially in Australia, became principally concerned with protecting the full complement of regional species diversity and required habitat within CANs. With meager monetary resources for protection, conservation biologists concentrated on developing methods that identified representative CANs in minimal areas (see "Place Prioritization" below).

The Consensus Framework

Recently, a growing consensus about the structure of conservation biology has emerged that combines aspects of the small-population and declining-population paradigms into a framework that emphasizes the crucial role computer-based place prioritization algorithms play in conservation planning (Margules and Pressey 2000; Sarkar 2005). The framework's purpose is to make conservation planning more systematic, thereby replacing the ad hoc reserve design strategies often employed in real-world planning in the past. It focuses on ensuring the adequate representation and persistence of regional biodiversity within the socioeconomic and political constraints inherent in such planning.

Place Prioritization

The first CAN design methods, especially within the United States, relied almost exclusively on island biogeography theory (MacArthur and Wilson 1967) (see Ecology). It was cited as the basis for geometric design principles intended to minimize extinction rates in CANs as their ambient regions were anthropogenically transformed (Diamond 1975). Island biogeography theory entails that the particular species composition of an area is constantly changing while, at an equilibrium between extinction and immigration, its species richness remains constant. The intention behind design principles inspired by the theory, therefore, was to ensure persistence of the maximum number of species, not the specific species the areas currently contained. However, incisive criticism of the theory (Gilbert 1980; Margules, Higgs, and Rafe 1982) convinced many conservation biologists, especially in Australia, that *representation* of the *specific species* that areas now contain, rather than the *persistence* of the *greatest number of species* at some future time, should be the first goal of CAN design. Computer-based place prioritization algorithms supplied a defensible methodology for achieving the first goal and an alternative to the problematic reliance on island biogeography theory in CAN design.

Place prioritization involves solving a resource allocation problem. Conservation funds are usually significantly limited and priority must be given to protecting some areas over others. The Expected Surrogate Set Covering Problem (ESSCP) and the Maximal Expected Surrogate Covering Problem (MESCP) are two prioritization problems typically encountered in biodiversity conservation planning (Sarkar 2004). Formally, consider a set of individual places called cells $\{c_j : j = 1, \ldots, n\}$; cell areas $\{a_j : j = 1, \ldots, n\}$; biodiversity surrogates $\Lambda = \{s_i : i = 1, \ldots, m\}$; representation targets, one for each surrogate $\{t_i : i = 1, \ldots, m\}$; probabilities of finding s_i at c_j $\{p_{ij} : i = 1, \ldots, m;\ j = 1, \ldots, n\}$; and two indicator variables $X_j (j = 1, \ldots, n)$ and $Y_i (i = 1, \ldots, m)$ defined as follows:

$$X_j = \begin{cases} 1, & \text{if } c_j \in \Gamma; \\ 0, & \text{otherwise;} \end{cases}$$

$$Y_i = \begin{cases} 1, & \text{if } \sum\limits_{c_j \in \Gamma} p_{ij} > t_i; \\ 0, & \text{otherwise.} \end{cases}$$

ESSCP is the problem:

$$\text{Minimize } \sum_{j=1}^{n} a_j X_j \text{ such that } \sum_{j=1}^{n} X_j p_{ij} \geq t_i \text{ for } \forall s_i \in \Lambda.$$

Informally, find the set of cells Γ with the smallest area such that every representation target is satisfied. MESCP is the problem:

$$\text{Maximize } \sum_{i=1}^{m} Y_i \text{ such that } \sum_{j=1}^{n} X_j = \mathbf{M},$$

where **M** is the number of protectable cells. Informally, given the opportunity to protect **M** cells, find those cells that maximize the number of representation targets satisfied.

The formal precision of these problems allows them to be solved computationally with heuristic or exact algorithms (Margules and Pressey 2000; Sarkar 2005). Exact algorithms, using mixed integer linear programming, guarantee optimal solutions: the smallest number of areas satisfying the problem conditions. Since ESSCP and MESCP are NP hard, however, exact algorithms are computationally intractable for practical problems with large datasets. Consequently, most research focuses on heuristic algorithms that are computationally tractable for large datasets but do not guarantee minimal solutions.

The most commonly used heuristic algorithms are "transparent," so called because the exact criterion by which each solution cell is selected is known. These algorithms select areas for incorporation into CANs by iteratively applying a hierarchical set of conservation criteria, such as rarity, richness, and complementarity. Rarity, for example, requires selecting the cell containing the region's most infrequently present surrogates, which ensures that endemic taxa are represented. The criterion most responsible for the efficiency of transparent heuristic algorithms is complementarity: Select subsequent cells that complement those already selected by adding the most surrogates not yet represented (Justus and Sarkar 2002). In policymaking contexts, transparency facilitates more perspicuous negotiations about competing land uses, which constitutes an advantage over nontransparent heuristic prioritization procedures such as those based on simulated annealing.

Methodologically, the problem of place prioritization refocused theoretical research in conservation biology from general theories to algorithmic procedures that require geographically explicit data (Sarkar 2004). In contrast to general theories, which abstract from the particularities of individual areas, place prioritization algorithms demonstrated that adequate CAN design critically depends upon these particularities.

Surrogacy

Since place prioritization requires detailed information about the precise distribution of a region's biota, devising conservation plans for specific areas would ideally begin with an exhaustive series of field surveys. However, owing to limitations of time, money, and expertise, as well as geographical and sociopolitical boundaries to fieldwork, such surveys are usually not feasible in practice. These limitations give rise to the problem of discovering a (preferably small) set of biotic or abiotic land attributes (such as subsets of species, soil types, vegetation types, etc.) the precise distribution of which is realistically obtainable and that adequately represents biodiversity as such. This problem is referred to as that of finding surrogates or "indicators" for biodiversity (Sarkar and Margules 2002).

This challenge immediately gives rise to two important conceptual problems. Given the generality of the concept of biodiversity as such, the first problem concerns the selection of those entities that should be taken to represent concretely biodiversity in a particular planning context. These entities will be referred to as the *true surrogate*, or objective parameter, as it is the parameter that one is attempting to estimate in the field. Clearly, selection of the true surrogate is partly conventional and will depend largely on pragmatic and ethical concerns (Sarkar and Margules 2002); in most conservation contexts the true surrogate will typically be species diversity, or a species at risk. Once a set of true surrogates is chosen, the set of biotic and/or abiotic land attributes that will be tested for their capacity to represent the true surrogates adequately must be selected; these are *estimator surrogates*, or indicator parameters.

The second problem concerns the nature of the relation of representation that should obtain between the estimator and the true surrogate: What does 'representation' mean in this context and under what conditions can an estimator surrogate be said to represent adequately a true surrogate? Once the true and estimator surrogates are selected and a precise (operational) interpretation of representation is determined, the adequacy with which the estimator surrogate represents the true surrogate becomes an empirical question. (Landres, Verner, and Thomas [1988] discuss the import of subjecting one's choice of estimator surrogate to stringent empirical testing.)

Very generally, two different interpretations of representation have been proposed in the conservation literature. The more stringent interpretation is that the distribution of estimator surrogates should allow one to *predict* the distribution of true surrogates (Ferrier and Watson 1997). The satisfaction of this condition would consist in the construction of a well-confirmed model from which correlations between a given set of estimator surrogates and a given set of true surrogates can be derived. Currently such models have met with only limited predictive success; moreover, since such models can typically be used to predict the distribution of only one surrogate at a time, they are not computationally feasible for practical conservation planning, which must devise CANs to sample a wide range of regional biodiversity. Fortunately, the solution to the surrogacy problem does not in practice require predictions of true-surrogate distributions.

A less stringent interpretation of representation assumes only that the set of places prioritized on the basis of the estimator surrogates adequately captures true-surrogate diversity up to some specified target (Sarkar and Margules 2002). The question of the *adequacy* of a given estimator surrogate then becomes the following: If one were to construct a CAN that samples the full complement of estimator-surrogate diversity, to what extent does this CAN also sample the full complement of true-surrogate diversity? Several different quantitative measurements can be carried out to evaluate this question (Sarkar 2004). At present, whether adequate surrogate sets exist for conservation planning remains an open empirical question.

Viability Analysis

Place prioritization and surrogacy analysis help identify areas that *currently* represent biodiversity. This is usually not, however, sufficient for successful biodiversity conservation. The problem is that the biodiversity these areas contain may be in irreversible decline and unlikely to persist. Since principles of CAN design inspired by island biogeography theory were, by the late 1980s, no longer believed to ensure persistence adequately, attention subsequently turned, especially in the United States, to modeling the probability of population extinction. These principles became known as population viability analysis (PVA).

PVA models focus primarily on factors affecting small populations, such as inbreeding depression, environmental and demographic stochasticity, genetic drift, and spatial structure. Drift and inbreeding depression, for example, may reduce the genetic variation required for substantial evolvability, without which populations may be more susceptible to disease, predation, or future environmental change. PVA modeling has not, however, provided a clear understanding of the general import of drift and inbreeding depression to population persistence in nature (Boyce 1992). Unfortunately, this kind of problem pervades PVA. In a trenchant review, for instance, Caughley (1994) concluded that the predominantly theoretical work done thus far in PVA had not been adequately tested with field data and that many of the tests that had been done were seriously flawed.

Models used for PVA have a variety of different structures and assumptions (see Beisinger and McCullough 2002). As a biodiversity conservation methodology, however, PVA faces at least three general difficulties:

1. Precise estimation of parameters common to PVA models requires enormous amounts of quality field data, which are usually unavailable (Fieberg and Ellner 2000) and cannot be collected, given limited monetary resources and the imperative to take conservation action quickly. Thus, with prior knowledge being uncommon concerning what mechanisms are primarily responsible for a population's dynamics, PVA provides little guidance about how to minimize extinction probability.
2. In general, the results of PVA modeling are extremely sensitive to model structure and parameter values. Models with seemingly slightly different structure may make radically different predictions (Sarkar 2004), and different parameter values for one model may produce markedly different predictions, a problem exacerbated by the difficulties discussed under (1).
3. Currently, PVA models have been developed almost exclusively for single species (occasionally two) and rarely consider more than a few factors affecting population decline. Therefore, only in narrow conservation contexts focused on individual species would they potentially play an important role. Successful models that consider the numerous factors affecting the viability of *multiple*-species assemblages, which are and should be the primary target of actual conservation planning, are unlikely to be developed in the near future (Fieberg and Ellner 2000).

For these reasons and others, PVA has not thus far uncovered nontrivial generalities relevant to CAN design. Consequently, attention has turned to more pragmatic principles, such as designing CANs to minimize distance to anthropogenically transformed areas. This is not to abandon the important goals of PVA or the need for sound theory about biodiversity persistence, but to recognize the weaknesses of existing PVA models and the imperative to act now given significant threats to biodiversity.

Multiple-Criterion Synchronization

The implementation and maintenance of CANs inevitably take place within a context of competing demands upon the allocation and use of land. Consequently, successful implementation strategies should ideally be built upon a wide consensus among agents with different priorities with respect to that usage. Thus, practical CAN implementation involves the attempt to optimize the value of a CAN amongst several different criteria simultaneously. Because different criteria typically conflict, however, the term "synchronization" rather than "optimization" is more accurate. The multiple-constraint synchronization problem involves developing and evaluating procedures designed to support such decision-making processes.

One approach to this task is to "reduce" these various criteria to a single scale, such as monetary cost. For example, cost-benefit analysis has attempted to do this by assessing the amount an agent is willing to pay to improve the conservation value of an area. In practice, however, such estimates are difficult to carry out and are rarely attempted (Norton 1987). Another method, based on multiple-objective decision-making models, does not attempt to reduce the plurality of criteria to a single scale; rather it seeks merely to eliminate those feasible CANs that are suboptimal when evaluated according to all relevant criteria. This method, of course, will typically not result in the determination of a uniquely best solution, but it may be able to reduce the number of potential CANs to one that is small enough so that decision-making bodies can bring other implicit criteria to bear on their ultimate decision (see Rothley 1999 and Sarkar 2004 for applications to conservation planning).

JUSTIN GARSON
JAMES JUSTUS

References

Angermeier, P. L., and J. R. Karr (1994), "Biological Integrity versus Biological Diversity as Policy Directives," *BioScience* 44: 690–697.

Beisinger, S. R., and D. R. McCullough (eds.) (2002), *Population Viability Analysis*. Chicago: University of Chicago Press.

Boyce, M. (1992), "Population Viability Analysis," *Annual Review of Ecology and Systematics* 23: 481–506.

Brower, L. P., and S. B. Malcolm (1991), "Animal Migrations: Endangered Phenomena," *American Zoologist* 31: 265–276.

Caughley, G. (1994), "Directions in Conservation Biology," *Journal of Animal Ecology* 63: 215–244.

Diamond, J. (1975), "The Island Dilemma: Lessons of Modern Biogeographic Studies for the Design of Natural Reserves," *Biological Conservation* 7: 129–146.

Faith, D. P. (2003), "Biodiversity," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/sum2003/entries/biodiversity

Ferrier, S., and G. Watson (1997), *An Evaluation of the Effectiveness of Environmental Surrogates and Modelling Techniques in Predicting the Distribution of Biological*

Diversity: Consultancy Report to the Biodiversity Convention and Strategy Section of the Biodiversity Group, Environment Australia. Arimidale, New South Wales: Environment Australia.

Fieberg, J., and S. P. Ellner (2000), "When Is It Meaningful to Estimate an Extinction Probability?" *Ecology* 8: 2040–2047.

Gilbert, F. S. (1980), "The Equilibrium Theory of Island Biogeography: Fact or Fiction?" *Journal of Biogeography* 7: 209–235.

Gómez-Pompa, A., C. Vázquez-Yanes, and S. Guevera (1972), "The Tropical Rain Forest: A Nonrenewable Resource," *Science* 177: 762–765.

Justus, J., and S. Sarkar (2002), "The Principle of Complementarity in the Design of Reserve Networks to Conserve Biodiversity: A Preliminary History," *Journal of Biosciences* 27: 421–435.

Landres, P. B., J. Verner, and J. W. Thomas (1988), "Ecological Uses of Vertebrate Indicator Species: A Critique," *Conservation Biology* 2: 316–328.

MacArthur, R., and E. O. Wilson (1967), *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press.

Margules, C. R., and R. L. Pressey (2000), "Systematic Conservation Planning," *Nature* 405: 242–253.

Margules, C., A. J. Higgs, and R. W. Rafe (1982), "Modern Biogeographic Theory: Are There Lessons for Nature Reserve Design?" *Biological Conservation* 24: 115–128.

Meffe, G. K., and C. R. Carroll (1994), *Principles of Conservation Biology*. Sunderland, MA: Sinauer Associates.

Norton, B. G. (1987), *Why Preserve Natural Variety?* Princeton, NJ: Princeton University Press.

Norton, B. G. (1994), "On What We Should Save: The Role of Cultures in Determining Conservation Targets," in P. L. Forey, C. J. Humphries, and R. I. Vane-Wright (eds.), *Systematics and Conservation Evaluation*. Oxford: Clarendon Press, 23–29.

Rothley, K. D. (1999), "Designing Bioreserve Networks to Satisfy Multiple, Conflicting Demands," *Ecological Applications* 9: 741–750.

Sarkar, S. (2004), "Conservation Biology," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://stanford.edu/entries/conservation-biology

——— (2005), *Biodiversity and Environmental Philosophy*. New York: Cambridge University Press.

Sarkar, S., and C. R. Margules (2002), "Operationalizing Biodiversity for Conservation Planning," *Journal of Biosciences* 27: 299–308.

Soulé, M. E. (1985), "What Is Conservation Biology?" *BioScience* 35: 727–734.

Takacs, D. (1996), *The Idea of Biodiversity: Philosophies of Paradise*. Baltimore: Johns Hopkins University Press.

CONSTRUCTIVE EMPIRICISM

See **Instrumentalism; Realism**

CONVENTIONALISM

Conventionalism is a philosophical position according to which the truth of certain propositions, such as those of ethics, aesthetics, physics, mathematics, or logic, is in some sense best explained by appeal to intentional human actions, such as linguistic stipulations. In this article, the focus will be on conventionalism concerning physics, mathematics, and logic. Conventionalism concerning empirical science and mathematics appears to have emerged as a distinctive philosophical position only in the latter half of the nineteenth century.

The philosophical motivations for conventionalism are manifold. Early conventionalists such as Pierre Duhem and Henri Poincaré, and more recently Adolf Grünbaum, have seen conventionalism about physical or geometrical principles as justified in part by what they regard as the underdetermination of a physical or geometrical theory by empirical observation (see Duhem Thesis; Poincaré, Henri; Underdetermination of Theories). An (empirically) arbitrary choice of a system from among empirically equivalent theories seems required.

A second motivation, also suggested by Duhem and Poincaré, and explicitly developed later by Rudolf Carnap, stems from their view that any description of the empirical world presupposes a suitable descriptive apparatus, such as a geometry, a metric, or a mathematics. In some cases there appears to be a choice as to which descriptive apparatus to employ, and this choice involves an arbitrary convention (see Carnap, Rudolf).

A third motivation for conventionalism, emphasized by philosophers such as Moritz Schlick, Hans Hahn, and Alfred Ayer, is that appeal to conventions provides a straightforward explanation of a priori propositional knowledge, such as knowledge of mathematical truths (see Ayer, Alfred Jules; Hahn, Hans; Schlick, Moritz). Such truths, it was thought, are known a priori in virtue of the fact that they are stipulated rather than discovered.

It is important to note that conventionalists do not regard the selection of a set of conventions to be wholly arbitrary, with the exception of the radical French conventionalist Edouard LeRoy, who did (see Giedymin 1982, 118–128). Conventionalists acknowledge that pragmatic or instrumental considerations involving human capacities or purposes might be relevant to the adoption of a set of conventions (see Instrumentalism). However, they insist that the empirical evidence does not compel one choice rather than another.

This article focuses on several major figures and issues, but there are a number of other important philosophers with broadly conventionalist leanings, including Duhem, Kazimierz Ajdukiewicz, and Ludwig Wittgenstein, that for the sake of brevity will not here receive the attention they deserve.

Poincaré and Geometric Conventionalism

The development of non-Euclidean geometries and subsequent research into the foundations of geometry provided both the inspiration and the model for much of the conventionalism of the early twentieth century (see Space-Time). The central figure in early conventionalism was the French mathematician and philosopher Poincaré (see Poincaré, Henri). Although a Kantian in his philosophy of mathematics, Poincaré shared with many late-nineteenth-century mathematicians and philosophers the growing conviction that the discovery of non-Euclidean geometries, such as the geometries of Lobatschevsky and Riemann, conjoined with other developments in the foundations of geometry, rendered untenable the Kantian treatment of geometrical axioms as a form of synthetic a priori intuition.

The perceived failings of Kant's analysis of geometry did not, however, lead Poincaré to an empiricist treatment of geometry. If geometry were an experimental science, he reasoned, it would be open to continual revision or falsified outright (the perfectly invariable solids of geometry are never empirically discovered, for instance) (Poincaré [1905] 1952, 49–50). Rather, geometrical axioms are conventions. As such, the axioms (postulates) of a systematic geometry constitute *implicit definitions* of such primitive terms of the system as "point" and "line" (50). This notion of implicit definition originated with J. D. Gergonne (1818), who saw an analogy between a set of sentences with *n* undefined terms and a set of equations with *n* unknowns. The two are analogous in that the roots that satisfy the equations are akin to interpretations of the undefined terms in the set of sentences under which the sentences are true. Of course, not every set of equations determines a set of values, and so too not every set of sentences (system of axioms) constitutes an implicit definition of its primitive terms. Poincaré accommodated this fact by recognizing two constraints on the admissibility of a set of axioms. First, the set must be consistent, and second, it must uniquely determine the objects defined (1952, 150–153).

Although the axioms were in his view conventional, Poincaré thought that experience nonetheless plays a role within geometry. The genesis of geometrical systems is closely tied to experience, in that the systems of geometry that are constructed are selected on the basis both of prior idealized empirical generalizations and of the simplicity and convenience that particular systems (such as Euclidean geometry) may afford their users ([1905] 1952, 70–71; 1952, 114–115). But these empirical considerations do not provide a test of empirically applied geometries. Once elevated to the status of a convention, a geometrical system is not an empirical theory, but is rather akin to a language used, in part, to frame subsequent empirical assertions. As a system of implicit definitions, it is senseless to speak of one geometry being "more true" than another ([1905] 1952, 50).

Poincaré therefore rejected the supposition that an empirical experiment could compel the selection of one geometry over another, provided that both satisfied certain constraints. He was inspired by Sophus Lie's (1959) group-theoretic approach to geometrical transformations, according to which, given an *n*-dimensional space and the possible transformations of figures within it, only a finite group of geometries with differing coordinate systems are possible for that space. Among these geometries there is no principled justification for selecting one (such as Plücker line geometry) over another (such as sphere geometry) (Poincaré [1905] 1952, 46–7). Indeed, one geometry within the group may be transformed into another given a certain method of translation, which Lie derived from Gergonne's theory of reciprocity. Poincaré conjoined the intertransformability of certain geometrical systems with the recognition that alternative geometries yield different conventions governing the notions of "distance" or "congruence" to argue that any attempt to decide by experiment between alternative geometries would be futile, since interpretations of the data presupposed geometric conventions. In a much-discussed passage he wrote:

> If Lobatchevsky's geometry is true, the parallax of a very distant star will be finite. If Riemann's is true, it will be negative. These are the results which seem within the reach of experiment . . . . But what we call a straight line in astronomy is simply the path of a ray of light. If, therefore, we were to discover negative parallaxes, or to prove that all parallaxes are higher than a certain limit, we should have a choice between two conclusions: we could give up Euclidean geometry, or modify the laws of optics, and suppose that light is not rigorously propagated in a straight line. ([1905] 1952, 72–73)

On one reading, advanced by Grünbaum, this passage affirms that there is no fact of the matter as to which is the "correct" metric, and hence supports the *conventionality of spatial metrics*. This reading is further discussed below. Another reading of this passage sees in it an argument closely akin to Duhem's argument leading to the denial of the possibility of a "crucial experiment" that could force the acceptance or elimination of a geometrical theory (see Duhem [1906] 1954, 180–90). Interestingly, Poincaré uses the argument to support a distinction between "conventional" truths and empirical truths, whereas Quine later adduces similar Duhemian considerations on behalf of his claim that there is no such distinction to be drawn (see Duhem Thesis; Quine, Willard Van).

However he may have intended his parallax example, Poincaré would probably have accepted both the conventionality of the spatial metric and the absence of any experimental test capable of deciding between two alternative geometries. As a further illustration of these issues, he proposed a well-known thought experiment ([1905] 1952, 65–67). Imagine a world consisting of a large sphere subject to the following temperature law: At the center of the sphere the temperature is greatest, and at the periphery it is absolute zero. The temperature decreases uniformly as one moves toward the circumference in proportion to the formula $R^2 - r^2$, where R is the radius of the sphere and r the distance from the center. Assume further that all bodies in the sphere are in perfect thermal equilibrium with their environment, that all bodies share a coefficient of dilation proportional to their temperature, and that light in this sphere is transmitted through a medium whose index of refraction is $1/(R^2 - r^2)$. Finally, suppose that there are inhabitants of this sphere and that they adopt a convention that allows them to measure lengths with rigid rods. If the inhabitants assume that these rods are invariant in length under transport, and if they further triangulate the positions in their world with light rays, they might well come to the conclusion that their world has a Lobatchevskian geometry. But they could also infer that their world is Euclidean, by postulating the universal physical forces just described. Again, no empirical experiment seems adequate to establish one geometry over the other, since both are compatible with all known possible observations given appropriate auxiliary hypotheses. Rather, a conventional choice of a geometry seems called for. The choice will be motivated by pragmatic factors, perhaps, but not "imposed" by the experimental facts (Poincaré [1905] 1952, 70–71).

Grünbaum's Conventionalism: The Metric and Simultaneity

Poincaré's parallax and sphere-world examples motivated his view that the choice of a spatial metric is conventional. Adolph Grünbaum has defended this conventionalist conclusion at length (Grünbaum 1968, 1973). Grünbaum claims that there is no unique metric "intrinsic" to a spatial manifold, since between any two points on a real number line, there is an uncountably infinite continuum of points. Hence, if one wishes to specify a metric for a manifold, one must employ "extrinsic" devices such as measuring rods, and must further stipulate the rods' behavior under transport.

Grünbaum defends what he takes to be Poincaré's metric conventionalism against a variety of objections. Perhaps the most serious objection is the claim that Poincaré's result illustrates only a *trivial* point about the conventionality of referring expressions, an objection first directed against Poincaré by Arthur Eddington and subsequently developed by Hilary Putnam and Paul Feyerabend. Eddington objected that Poincaré's examples illustrated not that metrical *relations* (such as the relations of equality or of the congruence of two rods) were conventional, but rather—and only—that the meanings of *words* such as "equal" and "congruent" were conventional. This, Eddington objected, was a trivial point about *semantical* conventionality, and the fact that the selection of different metrics could yield different but apparently equipollent geometries illustrated only that "the meaning assigned to length and distance has to go along with the meaning assigned to space" (Eddington 1953, 9–10). Eddington suggested, and Feyerabend later developed, a simple illustration of this point within the theory of gases. Upon the discovery that Boyle's law:

$$pv = RT$$

holds only approximately of real gases, one could *either* revise Boyle's law in favor of van der Waal's law:

$$(p + a/v^2)(v - b) = RT$$

or one could preserve Boyle's law by redefining pressure as follows:

$$\text{Pressure} =_{\text{Def}} (p + a/v^2)(1 - b/v).$$

(Feyerabend, quoted in Grünbaum 1973, 34)

The possibility of such a redefinition, the objection proceeds, is physically and philosophically uninteresting. Provided that a similar move is available in Poincaré's own examples with expressions like 'congruent,' Poincaré's examples appear to illustrate only platitudes about semantic conventionality.

Grünbaum tries to show that Poincaré's conventionalism is not merely an example of what Grünbaum calls "trivial semantic conventionality" by showing that the conventionality of the metric is the result of a certain absence of "intrinsic" structure within the space-time manifold, which structure can (or must) then be imposed "extrinsically":

> And the metric amorphousness of these continua then serves to *explain* that even *after* the word "congruent" has been pre-empted semantically as a spatial or temporal *equality* predicate by the axioms of congruence, congruence remains ambiguous in the sense that these axioms still allow an infinitude of mutually exclusive congruence classes of intervals. (Grünbaum 1973, 27)

Grünbaum developed a related argument to the effect that the simultaneity relation in special relativity (SR) involves a conventional element. A sympathetic reconstruction of the argument will be attempted here, omitting a number of details in order to present the gist of the argument as briefly and intuitively as possible. Within Newtonian mechanics (NM), there is no finite limit to the speed at which causal signals can be sent. Further, one might take it to be a defining feature of causal relations that effects cannot precede their causes. This asymmetry allows for a distinction between events that are temporally before or after a given event on a world line. There is then only one simultaneity relation within the space-time of NM meeting the constraints mentioned. Consider events (or space-time points, if one prefers) *a* and *b* on two distinct parallel world lines. Event *a* is simultaneous with *b* just in case *a* is not causally connectible to any event on the future part of *b*'s world line but is causally connectible to any event on the past of *b*'s world line. Within SR, however, there is an upper limit to the speed of causal signals. Thus the constraint that effects cannot precede their causes allows any one of a continuum of points along a parallel world line to be a candidate for simultaneity with a given event on the world line of an inertial observer. Grünbaum calls such points "topologically simultaneous" with a point *o* on the observer's world line. In claiming that simultaneity is (to some extent) conventional within the SR picture, Grünbaum is in good company, as Albert Einstein makes a claim to this effect in his original 1905 paper. (It is important to distinguish the issue of the *conventionality* of simultaneity within SR, which has been controversial, from the (observer or frame) relativity of simultaneity within SR, which is unchallenged; see Space-Time.)

Grünbaum claims that the existence of a continuum of possible "planes of simultaneity" through a given point is explained by, or reflects, the fact that the simultaneity relation is not "intrinsic" to the manifold of events within special relativity. One objection that a number of Grünbaum's critics such as Putnam (in Putnam 1975b) have had to his treatment of conventionalism (concerning both the metric and the simultaneity relation) is that the notion of an intrinsic feature, which plays a crucial role for Grünbaum, is never adequately clarified. David Malament (1977) provides a natural reading of 'intrinsic': An intrinsic property or relation of a system is "definable" in terms of a set of basic features. The intuitive idea is that if a relation is definable from relations that are uncontroversially intrinsic, then these ought to count as intrinsic as well. Malament goes on to show that if one takes certain fairly minimal relations to be given as intrinsic, then a unique simultaneity relation, in fact the standard one, is definable from them in a fairly well defined sense of 'definable.'

Malament's result has been taken by many to have definitively settled the issue of whether simultaneity is conventional within SR (see e.g., Norton 1992). In this view, the standard simultaneity relation is nonconventional, since it is definable, or "logically constructible," from uncontroversially intrinsic features of Minkowski space-time, the space-time of SR. Furthermore, standard simultaneity is the only such nonconventional simultaneity relation.

However, there are dissenters, including Grünbaum. A number of authors have attacked one or another of Malament's premises, including for instance the claim that simultaneity must be a transitive relation, that is, the requirement that for arbitrary events x, y, and z, if x sim y and y sim z, then x sim z. Sarkar and Stachel have shown that even if it is allowed that simultaneity must be an equivalence relation, other simultaneity relations (the backward and the forward light cones of events on the given world line) are definable from relations that Malament takes as basic or intrinsic (see Sarkar and Stachel 1999). Peter Spirtes (1981) shows that Malament's result is highly sensitive to the choice of basic or intrinsic relations, which fact might be taken to undercut the significance of Malament's result as well.

A conventionalist might raise a different sort of objection to Malament's results, questioning their relevance to a reasonable, although perhaps not the only reasonable, construal of the question as to whether simultaneity is conventional within SR. Suppose for simplicity that any relation that one is willing to call a simultaneity relation is an equivalence relation, that causes must always precede their effects, and that there is an upper bound to the speed of causal influences within SR. One may now ask whether more than one relation can be defined (extrinsically or otherwise) on the manifold of events (or "space-time points"), meeting these criteria within SR. That is, one might ask whether, given any model M of T (roughly, Minkowski space-time, the "standard" space-time of SR), there is a unique expansion of the model to a model M' for T', where T' adds to T sentences containing a symbol 'sim', which sentences in effect require that 'sim' be an equivalence relation and that causes are not 'sim' with their effects. It is a straightforward matter to see, as Grünbaum shows (the details are omitted here), that there is no such unique expansion of M. The constraints mentioned leave room for infinitely many possible interpretations of 'sim'. In this sense, the simultaneity relation (interpretation of 'sim') might naturally be said to be conventional within SR. In contrast, given a model for the causal structure of NM, the meaning constraints on the concept of simultaneity that are assumed here yield a unique expansion (a unique extension for 'sim').

It should be noted that the conventionality of simultaneity within SR and its nonconventionality within NM in the sense just described reflect structural differences between the two theories (or their standard models). Thus this form of conventionalism appears to escape the charge that the only sense in which simultaneity is conventional is the trivial sense in which the word 'simultaneous' can be used to denote different relations. As Grünbaum frequently emphasizes, the interesting cases of conventionality arise only when one begins with an already meaningful term or concept, whose meaning is constrained in some nontrivial way. It also allows one to interpret Einstein as making a substantive, yet fairly obvious, point when he claims that the standard simultaneity relation within SR is a conventional, or stipulated, choice.

How does this version of conventionalism relate to Malament's arguments? It will be helpful to note a puzzle before proceeding. Wesley Salmon (1977) argues at length that the one-way speed of light is not empirically identifiable within SR. (The basic problem is that the one-way speed seems measurable only by using distant synchronized clocks, but the synchronization of distant clocks would appear to involve light signals and presuppositions about their one-way speeds.) This fact appears to yield as an immediate consequence the conventionality (in the sense of empirical admissibility described above) of simultaneity within SR, given the various apparently admissible ways of synchronizing distant clocks within a reference frame. The puzzle is this: On one hand, the standard view has it that Malament's result effectively proves the falsehood of conventionalism (about simultaneity within SR); on the other hand, Salmon's arguments appear to entail the truth of conventionalism. Yet, his arguments for the empirical inaccessibility of the one-way speed of light seem sound. No one has convincingly shown what is wrong with them. All that the nonconventionalist seems able to provide is the indirect argument that Malament is right, so Salmon (and Grünbaum) must be wrong.

In the present analysis, this puzzle is dissolved by distinguishing two proposals, each of which may justifiably be called versions of conventionalism.

One version claims that a *relation* (thought of as a set or extension) is conventional within a model for a theory if it is not intrinsic within M, where 'intrinsic' might then be interpreted as definability (in the sense of logical constructibility) from uncontroversially intrinsic relations within M. The other version claims that the extension of a *concept* is conventional within a model (or theory) if the meaning of the concept does not entail a unique extension for the concept within the model (or within all models of the theory). Call these versions of conventionalism C_1 and C_2, respectively. Malament shows that simultaneity is not C_1 conventional within SR (leaving aside other possible objections to his result). Grünbaum and Salmon show that simultaneity is C_2 conventional within SR. Malament further considers relations definable from the causal connectibility relation (together with the world line of an inertial observer) within SR, and notes that only one is a nontrivial candidate simultaneity relation (relative to the corresponding inertial frame). He shows there is only one, and that therefore there is a unique intrinsic (and hence non–C_1 conventional) simultaneity relation within SR, the standard one. The C_2 conventionalist may respond that although Malament has pointed out an interesting feature of a particular candidate interpretation of the already meaningful term 'simultaneous,' he has not thereby ruled out all other candidate interpretations. One should not, the C_2 defender will argue, confuse the metalinguistic notion 'definable within M (or T)' with a meaning constraint on the concept of simultaneity, i.e., a constraint on potential interpretations of the already meaningful term 'simultaneous.' It would be absurd to claim that what was meant all along by 'simultaneous' required not only that any candidate be an equivalence relation and that it must "fit with" causal relations in that effects may not precede their causes, but also that for any two events, they are simultaneous if and only if they are in a relation that is definable from the causal connectibility relation.

A possible interpretation of the debate concerning the conventionality of simultaneity within SR, on the present construal, is that Grünbaum, perhaps in order to avoid the charge of simply rediscovering a form of trivial semantic conventionality, appealed to his notion of an intrinsic relation. Failure to express or denote an intrinsic feature was then said to characterize the interesting (nontrivial) cases of conventionality. This notion led to a number of difficulties for Grünbaum, culminating in Malament's apparent refutation of the conventionality of simultaneity within SR. However, it has been argued above that the appeal to intrinsicness turns out to be unnecessary in order to preserve a nontrivial form of conventionalism, C_2.

Grünbaum's arguments concerning the conventionality of the space-time metric for continuous manifolds involve more complex issues than those in the simultaneity case. But some key elements are common to both disputes. Grünbaum claims that continuous manifolds do not have intrinsic metrics, and critics complain that the notion of an intrinsic feature is unclear (see e.g., Stein 1977). Some of the claims that Grünbaum makes lend support to the idea that he is concerned with C_1 conventionalism. He claims, for example, that for a discrete manifold (such that between any two points there are only finitely many points), there is an intrinsic and hence nonconventional metric, where the "distance" between any two points is the number of points between them, plus one. This claim makes sense to the extent that one is concerned with C_1, that is, with the question of whether a relation is definable or logically constructible from a basic set. Grünbaum's arguments do not seem well suited for showing that such a metric would be C_2 nonconventional, since nothing about the meaning of 'distance' constrains the adoption of this as opposed to infinitely other metrics.

Another worry that might be raised concerns Grünbaum's insistence that distance functions defined in terms of physical bodies and their behaviors under transport are extrinsic to a manifold. If the manifold is a purely mathematical object, then such a view seems more plausible. But if one is concerned with a physical manifold of space-time points, it is less obvious that physical bodies should be treated as extrinsic. More importantly, even if one grants Grünbaum that physical bodies and their behaviors are extrinsic to the physical space-time manifold, one might want to grant a scientist the right to specify which features of the structures *that are being posited* are to count as basic or intrinsic. For example, Grünbaum proposes that Sir Isaac Newton in effect made a conceptual error in claiming that whether the temporal interval between one pair of instants is the same as that between another pair of instants is a factual matter, determined by the structure of time itself. Since instants form a temporal continuum in Newton's picture, there is, according to Grünbaum, no intrinsic temporal metric, contrary to Newton's claims. However, one can imagine a Newtonian responding with bewilderment. Does Newton not get to say what relations are intrinsic to the structures that he is positing?

It is difficult to see how to make room for a notion of intrinsicness that can do all of the philosophical work that Grünbaum requires of it, as Stein (1977) and others argue at length. On the other hand, Grünbaum's conventionalism about simultaneity within SR remains defensible (in particular, it escapes the trivialization charge as well as Malament's attempted refutation) if his conclusions are interpreted in the sense of C_2 conventionalism described above.

Mathematical and Logical Conventionalism

Up to now this discussion has focused on conventionalist claims concerning principles of an empirical science, physics. Other propositions that have attracted conventionalist treatment are those from the nonempirical sciences of mathematics and logic. Conventionalism seems especially attractive here, particularly to empiricists and others who find the notion of special faculties of mathematical intuition dubious but nevertheless wish to treat known mathematical or logical propositions as nonempirical.

The axiomatic methods that had motivated Poincaré's geometric conventionalism discussed above received further support with the publication of David Hilbert's *Foundations of Geometry* ([1921] 1962). Hilbert disregarded the intuitive or ordinary meanings of such constituent terms of Euclidean geometry as 'point,' 'line,' and 'plane,' and instead proposed regarding the axioms as purely formal posits (see Hilbert, David). In other words, the axioms function as generalizations about whatever set of things happens to satisfy them. In addition to allowing Hilbert to demonstrate a number of significant results in pure geometry, this method of abstracting from particular applications had immediate philosophical consequences. For instance, in response to Gottlob Frege's objection that without fixing a reference for primitive expressions like 'between,' Hilbert's axiomatic geometry would be equivocal, Hilbert wrote:

> But surely it is self-evident that every theory is merely a framework or schema of concepts together with their necessary relations to one another, and that the basic elements can be construed as one pleases. If I think of my points as some system or other of things, e.g., the system of love, of law, or of chimney sweeps ... and then conceive of all my axioms as relations between these things, then my theorems, e.g., the Pythagorean one, will hold of these things as well. (Hilbert 1971, 13)

Schlick was quick to find in Hilbert's conclusions the possibility of generalizing conventionalism beyond geometry (see Schlick, Moritz).

If the reference of the primitive concepts of an axiomatic system could be left undetermined, as Hilbert had apparently demonstrated, then the axioms would be empty of empirical content, and so knowledge of them would appear to be a priori. Like Poincaré, Schlick rejected a Kantian explanation of how such knowledge is possible, and he saw in Hilbert's approach a demonstration of how an implicit definition of concepts could be obtained through axioms whose validity had been guaranteed (Schlick [1925] 1985, 33). In his *General Theory of Knowledge* Schlick distinguished explicitly between "ordinary concepts," which are defined by ostension, and implicit definitions. Of the latter, he wrote that

> [a] system of truths created with the aid of implicit definitions does not at any point rest on the ground of reality. On the contrary, it floats freely, so to speak, and like the solar system bears within itself the guarantee of its own stability. (37)

The guaranteed stability was to be provided by a consistency proof for a given system of axioms, which Schlick, like Poincaré and Hilbert before him, claimed was a necessary condition in a system of symbolism. Schlick followed Poincaré in treating the axioms as conventions, and hence as known a priori through stipulation (Schlick [1925] 1985, 71). But he diverged markedly from Poincaré in extending this account to mathematics and logic as well. His basis for doing so was Hilbert's formalized theory of arithmetic, which Schlick hoped (wrongly, as it turned out) would eventuate in a consistency proof for arithmetic ([1925] 1985, 357).

The conventionalism that emerged was thus one in which true propositions known a priori were regarded as components of autonomous symbol games that, while perhaps constructed with an eye to an application, are not themselves answerable to an independent reality (Schlick [1925] 1985, 37–38). Schlick thought that the laws of logic, such as the principles of identity, noncontradiction, and the excluded middle, "say nothing at all about the *behavior of reality*. They simply regulate how we *designate the real*" ([1925] 1985, 337). Schlick acknowledged that the negation of logical principles like noncontradiction was unthinkable, but he suggested that this fact was *itself* a convention of symbolism (concerning the notion 'unthinkable'), claiming that "anything which contradicts the principle is termed *unthinkable*" ([1925] 1985, 337).

Although Schlick thought that the structure of a symbol system, including inference systems such as logic, was autonomous and established by a set of implicit definitions, he also recognized the possibility that some of its primitive terms could be coordinated by ordinary definitions with actual objects and properties. In this way, a conventionally established symbol system could be used to describe the empirical world, and an object designated by a primitive term might be empirically discovered to have previously unknown features. Nonetheless, some of the properties and relations had by a primitive term would continue to be governed by the system conventions through which the term is implicitly defined (Schlick [1925] 1985, 48ff). These properties and relations would thus be knowable a priori.

The conventionalism of Schlick's *General Theory of Knowledge* anticipated many of the conventionalist elements of the position advanced by members of the Vienna Circle (see Vienna Circle). Philosophers such as Hahn, Ayer, and Carnap saw in conventionalism the possibility of acknowledging the existence of necessary truths known a priori while simultaneously denying such propositions any metaphysical significance. Inspired by Wittgenstein's analysis of necessary truths in the *Tractatus*, Vienna Circle members identified the truths of mathematics and logic with *tautologies*—statements void of content in virtue of the fact that they hold no matter what the facts of the world may be (Hahn 1980; Ayer 1952). Tautologies, in turn, were equated with analytic statements, and Circle members regarded all analytic statements as either vacuous conventions of symbolism known a priori through stipulation or as derivable consequences of such conventions knowable a priori through proof (see Analyticity).

An especially noteworthy outgrowth of the conventionalism of the Vienna Circle was Carnap's *The Logical Syntax of Language*. In this book Carnap advanced a conventionalist treatment of the truths of mathematics and logic through his espousal of the principle of tolerance, which states:

> *In logic, there are no morals.* Everyone is at liberty to build up his own logic, i.e., his own form of language, as he wishes. All that is required of him is that ... he must state his methods clearly, and give syntactical rules instead of philosophical arguments. (Carnap 1937, 52)

Carnap regarded a language as a linguistic framework that specified logical relations of consequence among propositions. These logical relations are a precondition of description and investigation. In this view, there can be no question of justifying the selection of an ideal or even unique language with reference to facts, since any such justification (including a specification of facts) would presuppose a language. Mathematics is no exception to this. As a system of "framework truths," mathematical truths do not describe any convention-independent fact but are consequences of the decision to adopt one linguistic framework over another, in Carnap's account.

Conventionalism about mathematics and logic has faced a number of important objections. A significant early objection was given by Poincaré (1952, 166–172), who rejected Hilbert's idea that the principle of mathematical induction might merely be an implicit definition of the natural numbers (see Hilbert [1935] 1965, 193). Mathematical induction should not be regarded in this way, Poincaré argued, for it is presupposed by any demonstration of the consistency of the definition of number, since consistency proofs require a mathematical induction on the length of formulas. Treating the induction principle as itself conventional while using it to prove the consistency of the conventions of which it is a part would thus involve a *petitio.*

Carnap's *Logical Syntax of Language* sidestepped such problems by removing the demand that a system of conventions be consistent; the principle of tolerance made no such demand, and Carnap regarded the absence of consistency proofs as of limited significance (1937, 134). It appears that he would have regarded a contradictory language system to be pragmatically useless but not impossible.

Kurt Gödel, however, raised an important objection to this strategy. Gödel claimed that if mathematics is to be "merely" a system of syntactical conventions, then the conventions must be *known* not to entail the truth or falsity of any sentence involving a matter of extralinguistic empirical fact; if it does have such entailments, in Gödel's view, the truth of such sentences is not merely a syntactical, conventional matter (Gödel 1995, 339). If the system contains a contradiction, then it will imply every empirical sentence. But according to Gödel's second incompleteness theorem, a consistency proof required to assure the independence of the system cannot emerge from within the system itself if that system is consistent (1995, 346).

Recent commentators (Ricketts 1994; Goldfarb 1995) have argued that a response to this objection is available from within Carnap's conventionalist framework. Gödel's argument requires acceptance of an extralinguistic domain of empirical facts, which a stipulated language system may or may not imply. But it is doubtful that Carnap would have accepted such a domain, for Carnap considered linguistic conventions, including the conventions constitutive of mathematics, as antecedent to any characterization of the facts, including facts of the empirical world. The issues involved in Gödel's objection are complex and interesting, and a thorough and satisfactory conventionalist response remains to be formulated.

Another objection to conventionalism about logic was advanced by Willard Van Quine (see Quine, Willard Van). Quine objected against Carnap that treating logical laws and inference rules as conventional truths implicitly presupposed the logic that such conventions were intended to establish. Consider for instance a proposed logical convention *MP* of the form "Let all results of putting a statement for *p* and a statement for *q* in the expression 'If *p*, then *q* and *p*, then *q*' be true." In order to apply this convention to particular statements *A* and *B*, it seems that one must reason as follows: *MP* and if *MP*, then (*A* and if *A*, then *B* imply *B*); therefore, *A* and if *A*, then *B* imply *B*). But this requires that one use *modus ponens* in applying the very convention that stipulates the soundness of this inference. Quine concluded that "if logic is to proceed *mediately* from conventions, logic is needed for inferring logic from the conventions" (Quine 1966, 97). Quine at one point suggested that similar reasoning might undermine the intelligibility of the notion of a linguistic convention in general (98–99). However, after David Lewis' analysis of tacit conventions (Lewis 1969), Quine acknowledged the possibility of at least some such conventions (Quine, in Lewis 1969, xii).

Quine's original objection suggests that the conventionalist about logical truth must have an account of how something recognizable as a convention can intelligibly be established "prior to logic." As with Gödel's objection, the issues are difficult and have not yet been given a fully satisfying conventionalist account. However, there are a variety of strategies that a conventionalist might explore. One will be mentioned here. Consider an analogous case, that of rules of grammar. On one hand, one cannot specify rules of grammar without employing a language (which has its grammatical rules). Thus one cannot acquire one's first language by, say, reading its rules in a book. On the other hand, this fact does not seem to rule out the possibility that any given rules of grammar are to some extent arbitrary, and conventionally adopted. Whether the conventionalist can extend this line of thought to a defensible form of conventionalism about logic remains to be conclusively demonstrated.

Cory Juhl
Eric Loomis

The authors acknowledge the helpful input of Samet Bagce, Hans-Johan Glock, Sahotra Sarkar, and David Sosa.

References

Ayer, Alfred J. (1952), *Language, Truth and Logic*. New York: Dover.

Carnap, Rudolph (1937), *The Logical Syntax of Language*. London: Routledge and Kegan Paul.

Duhem, Pierre ([1906] 1954), *The Aim and Structure of Physical Theory*. Translated by Philip P. Wiener. Princeton, NJ: Princeton University Press.

Eddington, A. S (1953), *Space, Time and Gravitation*. Cambridge: Cambridge University Press.

Gergonne, Joseph Diaz (1818), "Essai sur la théorie de la définition," Annales de mathématiques pures et appliquées 9:1–35.

Giedymin, Jerzy (1982), *Science and Convention: Essays on Henri Poincaré's Philosophy of Science and the Conventionalist Tradition*. Oxford: Pergamon.

Gödel, Kurt (1995), "Is Mathematics Syntax of Language?" in Solomon Feferman (ed.), *Kurt Gödel: Collected Works*, vol. 3. Oxford: Oxford University Press, 334–362.

Goldfarb, Warren (1995), "Introductory Note to *1953/9," in Solomon Feferman (ed), *Kurt Gödel: Collected Works*, vol. 3. Oxford: Oxford University Press, 324–334.

Grünbaum, Adolph (1968), *Geometry and Chronometery in Philosophical Perspective*. Minneapolis: University of Minnesota Press.

——— (1973), *Philosophical Problems of Space and Time*, 2nd ed. Dordrecht, Netherlands: Reidel.

Hahn, Hans (1980), *Empiricism, Logic, and Mathematics*. Translated by Brian McGuinness. Dordrecht, Netherlands: Reidel.

Hilbert, David ([1935] 1965), "Die Grundlegung der elementaren Zahlenlehre," in *Gesammelte Abhandlungen*, vol. 3. New York: Chelsea Publishing, 192–195.

——— ([1921] 1962), *Grundlagen der Geometrie*. Stuttgart: B. G. Teubner.

——— (1971), "Hilbert's Reply to Frege," in E. H. W. Kluge (trans., ed.), *On the Foundations of Geometry and Formal Theories of Arithmetic*. London: Yale University Press, 10–14.

Lewis, David (1969), *Convention: A Philosophical Study*. Oxford: Blackwell.

Lie, Sophus (1959), "On a Class of Geometric Transformations," in D. E. Smith (ed.), *A Source Book in Mathematics*. New York: Dover.

Malament, David (1977), "Causal Theories of Time and the Conventionality of Simultaneity," *Nous* 11: 293–300.

Norton, John (1992), "Philosophy of Space and Time," in M. H. Salmon et. al. (eds), *Introduction to the Philosophy of Science*. Indianapolis: Hackett.

Poincaré, Henri ([1905] 1952), *Science and Hypothesis*. Translated by W. J. Greestreet. New York: Dover Publications.

——— (1952), *Science and Method*. Translated by Francis Maitland. New York: Dover Publications.

Putnam, Hilary (1975a), "An examination of Grunbaum's Philosophy of Geometry," in *Mathematics, Matter and Method Philosophical Papers*, vol. 1. New York: Cambridge University Press.

——— (1975b), "Reply to Gerald Massey," in *Mind, Language and Reality, Philosophical Papers*, vol. 2. New York: Cambridge UP.

Quine, Willard Van (1966), "Truth By Convention," in *The Ways of Paradox and Other Essays*. New York: Random House, 70–99.

Ricketts, Thomas (1994), "Carnap's Principle of Tolerance, Empricism, and Conventionalism," in Peter Clark and

Bob Hale (eds.), *Reading Putnam*. Cambridge, MA: Blackwell, 176–200.
Salmon, Wesley (1977), "The Philosophical Significance of the One-Way Speed of Light," *Noûs* 11: 253–292.
Sarkar, Sahotra, and John Stachel (1999), "Did Malament Prove the Non-Conventionality of Simultaneity in the Special Theory of Relativity?" *Philosophy of Science* 66: 208–220.
Schlick, Moritz ([1925] 1985), *General Theory of Knowledge*. Translated by Albert E. Blumberg. La Salle, IL: Open Court.
Spirtes, Peter L. (1981), "Conventionalism and the Philosophy of Henri Poincaré," Ph. D. Dissertation, Department of the History and Philosophy of Science, University of Pittsburgh.
Stein, Howard (1977), "On Space-Time and Ontology: Extract from a Letter to Adolf Grunbaum," in *Foundations of Space-time Theories*, Minnesota Studies in the Philosophy of Science, vol. 8. Minneapolis: University of Minnesota Press.

See also **Analyticity; Carnap, Rudolf; Duhem Thesis; Hilbert, David; Instrumentalism; Logical Empiricism; Poincaré, Henri; Quine, Willard Van; Schlick, Moritz; Space-Time; Vienna Circle**

CORROBORATION

Karl R. Popper (1959) observed that insofar as natural laws have the force of prohibitions, they cannot be adequately formalized as unrestrictedly general material conditionals but incorporate a modal element of natural necessity (Laws of Nature). To distinguish between possible laws and mere correlations, empirical scientists must therefore subject them to severe tests by serious attempts to refute them, where the only evidence that can count in favor of the existence of a law arises from unsuccessful attempts to refute it. He therefore insisted upon a distinction between 'confirmation' and 'corroboration' (see Popper, Karl Raimund).

To appreciate Popper's position, it is essential to consider the nature of natural laws as the objects of inquiry. When laws of nature are taken to have the logical form of unrestrictedly general material conditionals, such as $(x)(Rx \rightarrow Bx)$ for "All ravens are black," using the obvious predicate letters Rx and Bx and $\rightarrow$, as the material conditional, then they have many logically equivalent formulations, such as $(x)(\neg Bx \rightarrow \neg Rx)$, using the same notation, which stands for "Every nonblack thing is a nonraven." As Carl G. Hempel (1965) observed, if it is assumed that confirming a hypothesis requires satisfying its antecedent and then ascertaining whether or not its consequent is satisfied, then if logically equivalent hypotheses are confirmed or disconfirmed by the same evidence, a white shoe as an instance of $\neg Bx$ that is also $\neg Rx$ confirms the hypothesis "All ravens are black" (see Confirmation Theory; Induction, Problem of).

Popper argued that there are no "paradoxes of falsification" parallel to the "paradoxes of confirmation." And, indeed, the only way in which even a material conditional can be falsified is by things satisfying the antecedent but not satisfying the consequent. Emphasis on falsification therefore implies that serious tests of hypotheses *presuppose satisfying their antecedents*, thereby suggesting a methodological maxim of deliberately searching for examples that should be most likely to reveal the falsity of a hypothesis if it is false, such as altering the diet or the habitat of ravens to ascertain whether that would have any effect on their color. But Popper's conception of laws as prohibition was an even more far reaching insight relative to his falsificationist methodology.

Popper's work on *natural necessity* distinguishes it from logical necessity, where the notion of projectible predicates as a pragmatic condition is displaced by the notion of dispositional predicates as a semantic condition. It reflects a conception of laws as relations that cannot be violated or changed and require no enforcement. Fetzer (1981) has pursued this approach, where lawlike sentences take the form of subjunctive conditionals, such as $(x)(Rx \Rightarrow Bx)$, where $\Rightarrow$ stands for the subjunctive conditional. This asserts of everything, "If it were a raven, then it would be black." The truth of this claim, which is logically contingent, depends on a difference between permanent and transient properties. It does not imply the counterpart, "If anything were nonblack, then it would be a nonraven."

Among Popper's most important contributions were his demonstration of how his falsificationist methodology could be extended to encompass statistical laws and his propensity interpretation of probability (see Probability). Distributions of outcomes that have very low probabilities in hypotheses are regarded, by convention, as methodologically falsifying those hypotheses, tentatively and fallibilistically. Popper entertained the prospect of probabilistic measures of corroboration, but likelihood measures provide a far better fit, since universal hypotheses as infinite conjunctions have zero probability. Indeed, Popper holds that the appropriate hypothesis for scientific acceptance is the one that has the greatest content and has withstood the best attempts at its falsification, which turns out to be the least probable among the unfalsified alternatives.

The likelihood of hypothesis H, given evidence E, is simply the probability of evidence E, if hypothesis H is true. While probabilistic measures have to satisfy axioms of summation and multiplication—for example, where mutually exclusive and jointly exhaustive hypotheses must have probabilities that sum to 1—likelihood measures are consistent with arbitrarily many hypotheses of high value. This approach can incorporate the laws of likelihood advanced by Ian Hacking (1965) and a distinction between *preferability* for hypotheses with a higher likelihood on the available evidence and *acceptability* for those that are preferable when sufficient evidence becomes available. Acceptability is partially determined by relative likelihoods, even when likelihoods are low. Popper (1968, 360–91) proposed that where E describes the outcome of a new test and B our background knowledge prior to that test, *the severity of* E *as a test of* H, *relative to* B, might be measured by

$$P(E \mid H \wedge B) - P(E \mid B), \tag{1}$$

that is, as the probability of E, given H and B, minus the probability of E, given B alone. The intent of equation 1 may be more suitably captured by a formulation that employs a symmetrical—and therefore absolute—measure reflecting differences in expectations with respect to outcome distributions over sets of relevant trials, such as

$$\mid P(E \mid H) - P(E \mid B) \mid, \tag{2}$$

which ascribe degrees of nomic expectability to relative frequency data, for example. When B entails E, then $P(E|B) = 1$, and if $P(E|H) = 1$ as well, the severity of any such test is minimal, that is, 0. When H entails E, while B entails $\neg E$, the severity of such a test is maximal, i.e., 1. The acceptance of H may require the revision of B to preserve consistency.

A plausible measure of *the degree of corroboration* C *of* H, *given* E, *relative to* B, would be

$$c(H \mid E \wedge B) = L(H \mid E)[\mid P(E \mid H) - P(E \mid B) \mid], \tag{3}$$

that is, as the product of the likelihood of H, given E, times the severity of E as a test of H, relative to B. This is a Popperian measure, but not necessarily Popper's. Popper suggests (as one possibility)

$$\begin{aligned} &C(H \mid E \wedge B) = \\ &\quad P(E \mid H) - P(E \mid B)/P(E \mid H) + P(E \mid B), \end{aligned} \tag{4}$$

which even he does not find to be entirely satisfactory and which Imre Lakatos severely criticizes (Lakatos 1968, especially 408–16). When alternative hypotheses are available, the appropriate comparative measures appear to be corroboration ratios

$$\frac{C(H_2 \mid E)}{C(H_1 \mid E)} = \frac{L(H_2 \mid E)[\mid P(E \mid H_2) - P(E \mid H_1) \mid]}{L(H_1 \mid E)[\mid P(E \mid H_1) - P(E \mid H_2) \mid]} \tag{5}$$

that reduce to the corresponding likelihood ratios of $L(H_2|E)$ divided by $L(H_1|E)$ and assume increasing significance as a function of the severity of those tests, as Fetzer (1981, 222–230) explains.

Popper also proposed the propensity interpretation of physical probabilities as probabilistic dispositions (Popper 1957 and 1959). In a revised formulation, the single-case propensity interpretation supports a theory of lawlike sentences, logically contingent subjunctive conditionals ascribing permanent dispositional properties (of varying strength) to everything possessing a reference property (Fetzer 1981 and 1993). Propensity hypotheses are testable on the basis of the frequencies they generate across sequences of trials. Long runs are infinite and short runs are finite sequences of single trials. Propensities predict frequencies but also explain them. Frequencies are evidence for the strength of propensities.

Popper (1968) promoted the conception of science as a process of conjectures and (attempted) refutations. While he rejected the conception of science as a process of inductive confirmation exemplified by the work of Hans Reichenbach (1949) and Wesley C. Salmon (1967), his commitments to deductive procedures tended to obscure the role of ampliative reasoning in his own position. Popper rejected a narrow conception of induction, according to which the basic rule of inference is "If *m*/*n* observed *A*s are *B*s, then infer that *m*/*n* *A*s are *B*s, provided a suitable number of *A*s are tested under a wide variety of conditions." And, indeed, the rule he rejects restricts scientific hypotheses to those couched in observational language and cannot separate *bona fide* laws from correlations.

Although it was not always clear, Popper was not thereby rejecting induction in the broad sense of ampliative reasoning. He sometimes tried to formalize his conception of corroboration using the notion of absolute (or prior) probability of the evidence E, $P(E)$, which is typically supposed to be a subjective probability. Grover Maxwell (1974) even develops Popper's approach using Bayes's theorem. However, appeals to priors are inessential to formalizations of Popper's measures (Fetzer 1981), and it would be a mistake to suppose that Popper's account of severe tests, which is a pragmatic conception, could be completely formalizable.

Indeed, Popper's notion of accepting hypotheses on the basis of severe tests, no matter how tentatively and fallibilistically, implies ampliative reasoning. Its implementation for probabilistic hypotheses thereby requires large numbers of trials over a wide variety of conditions, which parallels the narrow inductivist conception. These results, however, must be subjected to severe tests to make sure the frequencies generated are robust and stable under variable conditions. When Volkswagens were first imported into the United States, for example, they were all gray. The narrow inductivist rule of inference justified inferring that all Volkswagens were gray, a conclusion that could not withstand severe tests.

James H. Fetzer

References

Fetzer, James H. (1981), *Scientific Knowledge: Causation, Explanation, and Corroboration*. Dordrecht, Netherlands: D. Reidel.

——— (1993), *Philosophy of Science*. New York: Paragon House.

Hacking, Ian (1965), *Logic of Statistical Inference*. Cambridge: Cambridge University Press.

Hempel, Carl G. (1965), *Aspects of Scientific Explanation*. New York: Free Press.

Lakatos, Imre (1968), "Changes in the Problem of Inductive Logic," in *The Problem of Inductive Logic*. Amsterdam: North-Holland, 315–417.

Maxwell, Grover (1974), "Corroboration without Demarcation," in Paul A. Schilpp (ed.), *The Philosophy of Karl R. Popper*, part 1. LaSalle, IL: Open Court, 292–321.

Popper, Karl R. (1959), *The Logic of Scientific Discovery*. New York: Harper & Row.

——— (1957), "The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory," in Stephan Korner (ed.), *Observation and Interpretation in the Philosophy of Physics*. New York: Dover Publications, 65–70.

——— (1959), "The Propensity Interpretation of Probability," *British Journal for the Philosophy of Science* 10: 25–42.

——— (1968), *Conjectures and Refutations*. New York: Harper & Row.

Reichenbach, Hans (1949), *The Theory of Probability*. Berkeley and Los Angeles: University of California.

Salmon, Wesley C. (1967), *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

See also **Confirmation Theory; Hempel, Carl Gustav; Induction, Problem of; Popper, Karl Raimund; Verisimilitude**

COSMOLOGY

See **Anthropic Principle**

COUNTERFACTUALS

See **Causality**

D

DECISION THEORY

Decision theory seeks to provide a normative account of rational decision making and to determine the extent to which human agents succeed in living up to the rational ideal. Though many decision theories have been proposed, the version of *expected utility theory* developed in L. J. Savage's ([1954] 1972) classic *Foundations of Statistics* remains the best developed and most influential. It will serve as the principal focus of this article. Savage established a general framework for thinking about decision problems. He codified core tenets of the theory of rational preference and argued cogently for them. Most important, he proved a *representation theorem* that helps to legitimize subjective Bayesian approaches to epistemology and to justify *subjective expected utility maximization* as the foundation of rational decision making (see Bayesianism). Indeed, Savage's contributions are so seminal that the best way to approach the topic of decision theory is by treating his theory as a kind of "gold standard" and discussing other views as reactions or additions to it. This is the approach taken here. This article has three sections. The first discusses the general notion of a decision problem. The second introduces the expected utility hypothesis and explains Savage's representation theorem. The third presents the standard theory of rational preference and discusses objections to it.

Decision Problems

Savage's model assumes a rational decision maker, hereafter the *agent*, who uses beliefs about possible *states of the world* to choose *actions* that can be expected to produce desirable *consequences*. The states describe all relevant contingencies that lie beyond the agent's direct control. Any uncertainty that figures into the agent's choice is portrayed as ignorance about which state obtains. *Events* are disjunctions of states that provide less specific descriptions of the circumstances under which choices are made than states do. Consequences serve as objects of *noninstrumental* desire. Each specifies a possible course of events that is sufficiently detailed to settle every matter about which the agent intrinsically cares. Acts are objects of *instrumental* desire; the agent values them only insofar as they provide a means to the end of securing desirable consequences. When there are only finitely many acts and states, the person's choice can be described using a *decision matrix:*

	S_1	S_2	$S_{3\ldots}$	S_n
A_1	$C_{1,1}$	$C_{1,2}$	$C_{1,3\ldots}$	$C_{1,n}$
A_2	$C_{2,1}$	$C_{2,2}$	$C_{2,3\ldots}$	$C_{2,n}$
A_3	$C_{3,1}$	$C_{3,2}$	$C_{3,3\ldots}$	$C_{3,n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_m	$C_{m,1}$	$C_{m,2}$	$C_{m,3\ldots}$	$C_{m,n}$

where S_j = states, A_i = acts, and $C_{i,j}$ = the outcome that A_i will produce when S_j obtains.

This model of decision problems applies to "one-choice" decisions made at a specific time. Though early decision theorists, like Savage, believed that sequences of decisions could be represented by one-shot choices among contingency plans, or *strategies*, this view now has few adherents. The topic of dynamic decision making lies beyond the scope of this article. (For relevant discussions, see Hammond 1988; McClennen 1990; and Levi 1991.)

A decision problem is counted as rational only if the following conditions hold:

- The value of each consequence C is independent of the act and state that bring it about.
- Each act/state pair (A, S) determines a unique consequence $C_{A,S}$.
- The agent cannot causally influence that which the state obtains.

Many misguided objections to expected utility theory involve decision problems that violate these requirements. For example, the following is a central tenet of the theory:

- *Comparative Probability (CP)*: For any event E and consequences C and C^* with C preferred to C^*, the agent prefers an act that produces C when E and C^* when $\neg E$ to an act that produces C^* when E and C when $\neg E$ if and only if the agent is more confident in E than in $\neg E$.

This seems susceptible to counterexample. Imagine a person who is convinced that the average annual inflation rate over the next decade will be either 10% or 1% but that 10% is far more likely:

	E = The rate of inflation will be high over the next decade.	The rate of inflation will not be high over the next decade.
A	Be paid $1,000 in ten years.	Be paid $0 in ten years.
A^*	Be paid $0 in ten years.	Be paid $1,000 in ten years.

Despite confidence in E, such a person may prefer A^* to A on the grounds that $1,000 will be worth more if inflation is low than if it is high. This preference does not refute CP, which applies only to preferences in well-formed decision problems, and this problem is *ill-formed*, since the consequence "Be paid $1,000 in ten years" is worth less when it appears in the upper left than when it appears in the lower right. (Proponents of *state-dependent utility theory* relax this requirement by allowing utilities of outcomes to vary with states. See Karni 1985.) To fix the problem, one needs to rewrite outcomes as follows:

	The rate of inflation will be high over the next decade.	The rate of inflation will not be high over the next decade.
A	Be paid $1,000 after ten years of high inflation.	Be paid $0 after ten years of low inflation.
A^*	Be paid $0 after ten years of low inflation.	Be paid $1,000 after ten, years of high inflation.

When the problem is redescribed this way, the preference for A over A^* does not violate **CP**.

Similar problems arise when the decision maker can influence states of the world. Another core tenet of expected utility theory is:

- *Dominance*. If the agent prefers A's consequences to A^*'s consequences in every possible state of the world, then the agent prefers A to A^*.

Dominance sometimes seems to make absurd recommendations:

	One will contract influenza this winter.	One will not contract influenza this winter.
Get a flu shot.	Get the flu, and suffer the minor pain of a shot.	Avoid the flu, but suffer the minor pain of a shot.
Do not get a flu shot.	Get the flu, but avoid the minor pain of a shot.	Avoid the flu, and avoid the minor pain of a shot.

Here it seems as if Dominance requires one to forgo the shot to avoid the pain, which is terrible advice given that the chances of getting the flu are markedly less with the shot than without it. Again, this is not an objection to expected utility theory, but an ill-posed decision problem. To properly reformulate the problem, one must use states like these:

- One will contract the flu whether or not one gets the shot.
- One will not contract the flu whether or not one gets the shot.
- One will contract the flu if one gets the shot, but not otherwise.
- One will not contract the flu if one gets the shot, but will otherwise.

Dominance reasoning always holds good for states that are independent of the agent's acts, as these states are.

The debate between *causal* and *evidential* decision theorists has to do with the sort of independence that is required here. Evidentialists believe that states must be *evidentially* independent of acts, so that no act provides a sign or signal of the occurrence of any state. Causal decision theorists adopt the stronger requirement that states must be *causally* independent of acts, so that nothing the agent can do will change the probabilities of states (for details, see Jeffrey 1983; Gibbard and Harper 1978; Skyrms 1980; Joyce 1999).

Expected Utility Representations of Preference

Following Savage, it is standard in decision theory to assume that after due deliberation, a rational agent will be able to order acts with respect to their effectiveness as instruments for producing desirable outcomes. This generates a *weak preference ranking* $A \geq A^*$ that holds between acts A and A^* just when, all things considered, the agent strictly prefers A to A^* or is indifferent between them. It is important to understand that this preference ranking among acts is *not* what the agent *starts out* with when making her decision. It is the *end result* of the deliberative process.

Decision theorists have historically understood preferences behavioristically, so that $A > A^*$ *means* that the agent would choose A over A^* if given the chance. Though some social scientists still adhere to this interpretation, it has been widely and effectively criticized. The alternative is to take preferences as representing one's *all-things-considered judgments* about which acts will best serve one's interests. In this reading, saying that one prefers A to A^* means that the balance of one's reasons favors realizing A rather than A^*; whether one can or will choose A is another matter.

An act A is *choiceworthy* when it is weakly preferred to every alternative. According to the expected utility hypothesis, rationally choiceworthy acts maximize the decision maker's *subjective expected utility*. In Savage's framework, an act's expected utility is defined as

$$Exp_{\mathrm{P,U}}(A) = P(S) \times U(C_{A,S})$$

where P is a *probability function* defined over events, and U is a real-valued *utility function* defined over consequences. To show that rationally choiceworthy acts maximize expected utility, Savage imposed a system of axiomatic rationality constraints on preference rankings and then proved that any ranking satisfying his axioms would be consistent with the hypothesis that acts are ranked according to expected utility. (The first result of this type is found in Ramsey 1931.) One can think of Savage as seeking to establish the following two claims:

1. *Theory of Practical Rationality*. Any practically rational agent will have a preference ranking among acts that obeys Savage's axioms.
2. *Existence of Subjective Expected Utility Representations*. For any preference ranking that obeys Savage's axioms, there will be at least one probability/utility pair (P, U) such that:
   - **P** *represents* the agent's beliefs: One event E is taken to be at least as likely as another E^* only if $P(E) \geq P(E^*)$.
   - **U** represents the agent's (intrinsic) desires for consequences: C is weakly preferred to C^* only if $U(C) \geq U(C^*)$.
   - $Exp_{\mathrm{P,U}}$, accurately represents the agent's (instrumental) desires for actions: A is weakly preferred to A^* only if $Exp_{\mathrm{P,U}}(A) \geq Exp_{\mathrm{P,U}}(A^*)$.

It follows directly that one whose preferences satisfy Savage's axioms will always behave *as if* one were choosing acts based on their expected utility (though it is in no way required that one actually use this method in making the decision).

Savage also proves that this representation is unique once a unit and zero point for measuring utility have been fixed. To establish uniqueness, Savage was forced to assume that the agent has determinate preferences over an extremely rich set of options. Many decision theorists reject these "richness assumptions" and so believe that an agent's beliefs and desires should be represented by *sets* of probability/utility pairs, as in Joyce (1999, 102–105).

The Theory of Rational Preference

Since there is no question about the validity of Savage's representation theorem, his case for expected utility maximization rests on the plausibility of his axioms as requirements of practical rationality. Rather than trying to formulate things as Savage does, it will be better to discuss informal versions of those of his axioms and auxiliary assumptions that are components of every expected utility theory.

Frame Invariance and Value Independence

It will serve to begin by considering two principles that are left implicit in most formulations of expected utility theory:

- *Frame Invariance (INV).* A rational agent's preferences among acts should depend only on the consequences the acts produce in various states of the world, and not on the ways in which these consequences, or the acts themselves, happen to be described.
- *Value Independence (VALUE).* A rational agent endows each act with a value that is independent of the decisions in which it figures.

While **INV**'s credentials as a requirement of rationality have never been seriously questioned, a great deal of empirical research suggests that people's preferences do often depend on the way in which decisions are "framed." For example, when presented with the following two decisions (see Tversky and Kahneman 1986),

1. One will be paid \$300 to choose between (a) getting another \$100 for sure or (b) getting another \$200 with probability $\frac{1}{2}$ and \$0 with probability $\frac{1}{2}$, and
2. One will be paid \$500 to make a choice between either (a*) paying back \$100 for sure or (b*) paying back \$0 with probability $\frac{1}{2}$ and paying back \$200 with probability $\frac{1}{2}$,

a surprising number of people prefer (a) in the first choice and (b*) in the second, thus violating **INV**. The different descriptions of the same options lead people to view their choices from different perspectives. In the first case, they see themselves as having \$300 and try to *improve* their lot by choosing between (a) and (b); while in the second, they (wrongly) see themselves having \$500 and try to *preserve* their fortune by choosing between (a*) and (b*). This generates problems when conjoined with the following facts about human behavior (see Kahneman and Tversky 1979; Shafir and Tversky 1995):

- *Divergence from the Status Quo.* People are more concerned with gains and losses, seen as additions or subtractions from the status quo, than with total well-being or overall happiness.
- *Asymmetrical Risk Aversion.* People tend to be risk averse when pursuing gains, but risk seeking when avoiding losses.

People who choose between (a) and (a^*) see \$300 as the status quo, and thus prefer the less risky (a), since they are risk averse when pursuing gains. When choosing between (b) and (b^*), they see \$500 as the status quo and prefer the more risky (b^*) because they are risk seeking when aiming to avoid losses.

VALUE has a number of important implications. First, it entails that one should be able to experimentally "elicit" a person's preference between A and A^* using any of the following methods:

- *Fair Prices.* Have an agent put a "fair price" on each action, and conclude that the higher-priced act is preferred.
- *Choice.* Have an agent choose between A and A^*, and conclude that the chosen act is not dispreferred.
- *Rejection.* Have an agent reject A or A^*, and conclude that the rejected act is not preferred.
- *Exchange.* Award an agent A^*, offer to trade A for A^* plus a small fee, and conclude that A is preferred if the agent makes the trade.

Surprisingly, these procedures can all yield different results, a fact that creates havoc for behaviorist analyses of preference. In cases of *preference reversal*, an agent who sets a higher price on A also selects A^* in a straight choice. Shafir (1993) gives examples in which subjects choose A over A^* and yet reject A for A^*. This happens because they focus more on comparisons among "positive" features of options when choosing, but more on negative features when rejecting. When A has both more pronounced positive features and more pronounced negative features, it can be both chosen and rejected. There are even cases in which an agent will refuse to trade A for A^* and refuse to trade A^* for A. The mere fact that the person 'owns' a prospect seems to make it more valuable to her. (This is referred to as *loss aversion*.)

VALUE also says that an agent's preferences among options should not depend on what other options happen to be available. This entails the following two principles (Sen 1971) (note that these do *not* apply when the addition or deletion of A^{**} provides relevant information about the desirability of A or A^*):

- Principle-α: If the agent will choose A over A^* and A^{**}, then the agent will choose A over A^* even when A^{**} is not available.

- Principle-β: If the agent will choose A in a straight choice against A^*, then the agent will also choose A in a choice between A, A^*, and some third option that is inferior to A^*.

Actual choosers often violate these principles. As salespeople have long known, one can more easily convince a person to buy a product by offering an inferior product at a higher price. Likewise, offering too many good options can lead a person to refrain from buying products he would have purchased had the list of options been smaller. A disconcerting violation of both principles is the finding of Redelmeier and Shafir (1995) that physicians are *less* likely to prescribe pain medication to patients when they can choose between ibuprofen and the inferior piroxicam than when they can choose only ibuprofen. While none of these empirical results have led decision theorists to question the normative standing of **VALUE**, they clearly show that expected utility theory is not an accurate *description* of human behavior.

Completeness

Along with Dominance and Comparative Probability, expected utility theorists also generally state as axioms:

- *Transitivity*. If the agent (strictly or weakly) prefers A to A^* and A^* to A^{**}, then the agent prefers A to A^{**}.
- *Completeness (COM)*. The agent either strictly prefers A to A^*, strictly prefers A^* to A, or is indifferent between them.
- *The "Sure-Thing" Principle (STP)*. If A and A^* produce the same consequences in every state consistent with an event $\neg E$, then the agent's preference between A and A^* depends exclusively only on their consequences when E obtains.

Though all five axioms are controversial in various ways, COM and STP have been the most contentious.

Completeness can fail in two ways. First, an agent might have no definite preference between two prospects either because the agent's intrinsic desires are vague or indeterminate or because the agent has insufficient information to judge which act will better promote desirable consequences. Being indifferent is not the same as having no preference. One who is indifferent between two options judges that they are equally desirable, but one who lacks a clear preference is unable to judge that one option is better than the other, or even that they are equally good—the person simply has no view about their relative desirabilities. Most decision theorists now admit that there is nothing irrational about having a "gappy" preference ranking, and it is becoming standard to treat COM (and any other axiom that requires the *existence* of preferences) as a *requirement of coherent extendibility* (see Jeffrey 1983; Kaplan 1983; Joyce 1999, 103–5). In this reading, it is irrational to have an incomplete preference ranking only if it cannot be extended to a complete ranking that obeys all the other axioms.

A more serious objection to COM comes from those who hold that rational agents may be unable to compare acts or consequences not because of any vagueness in their beliefs or desires, but because they regard the values of these prospects as genuinely *incommensurable* (Raz 1986; Anderson 1993). In one version of the view, a rational agent might regard distinctive standards of evaluation as appropriate to different sorts of prospects and so see prospects that do not fall under a common standard as incomparable. For example, a person might see it as perfectly appropriate to set a monetary price on a share of stock but also regard it as improper to put a price on spending an afternoon with one's children. If this is so, then a person's preference ranking will not compare these prospects, and no extension of it consistent with that person's values will do so. The incommensurability debate is too involved to pursue here. The heart of the issue has to do with the ability of rational agents to "balance off" reasons for and against an option so as to come to an all-things-considered judgment about its desirability. Utility theorists think that such a balancing of reasons is always possible. Proponents of incommensurability deny this.

The Sure-Thing Principle

The Sure-Thing Principle forces preferences to be *separable across events*, so that a rational agent's preference between A and A^* depends *only* on what happens in states of the world in which these prospects produce *different* outcomes. When there are three states to be considered, STP requires an agent facing the following decision to prefer A over A^* if and only if the agent also prefers B over B^*:

	S_1	S_2	S_3
A	C_1	C_2	C_3
A^*	C_1^*	C_2^*	C_3
B	C_1	C_2	D_3
B^*	C_1^*	C_2^*	D_3

In deciding between *A* and *A** or between *B* and *B**, STP tells the agent to *ignore* what happens when S_3 holds, since the same result occurs under S_3 whichever option is chosen. In effect, the requirement is that the agent be able to form a preference between the following two *act types* whether or not the value of *x* is known:

	S_1	S_2	S_3
X	C_1	C_2	x
X^*	C_1^*	C_2^*	x

STP has generated more controversy than any other tenet of expected utility theory. Much of the discussion concerns two putative counterexamples, the *Allais* and *Ellsberg paradoxes*, which seem to show that an important component of rational preference—the amount of risk or uncertainty involved in an option—is nonseparable in the sense required by STP. In the jargon of economists, a prospect involves *risk* when the agent knows the objective probability of each state of the world. It involves *uncertainty* when the agent does not have sufficient information to assign objective probabilities to states. STP entails that an agent's attitudes toward risk and uncertainty can be fully captured by the combination of the agent utility function for outcomes and the probabilistic averaging involved in the computation of expected utilities. The Allais and Ellsberg paradoxes appear to show that the theory is wrong about this.

In *Allais' paradox* (Allais 1990) agents choose between *A* and *A** and then between *B* and *B** (where known probabilities of states are listed):

	0.33	0.01	0.66
A	$2,500	$0	$2,400
*A**	$2,400	$2,400	$2,400
B	$2,500	$0	$0
*B**	$2,400	$2,400	$0

Most people violate STP by preferring *A** to *A* and *B* to *B**, and these preferences remain stable upon reflection. The thinking seems to be that in the first choice one should play it safe and take the sure $2,400, since a 0.33 chance at an extra $100 does not compensate for a 0.01 risk of ending up with nothing. On the other hand, since one will probably end up with nothing in the second choice, the chance of getting an extra $100 makes the risk worth taking. Thus, Allais choosers think (a) that there is more risk involved in choosing in *A* over *A** than in choosing *B* over *B**, and (b) that this added risk justifies their nonseparable preferences.

In *Ellsberg's paradox* (Ellsberg 1961), a ball is drawn at random from an urn that is known to contain 30 red balls and 60 balls that are either white or blue but in unknown proportion. The agent is asked to choose between *A* and *A** and then between *B* and *B**.

	Red	White	Blue
A	$100	$0	$0
*A**	$0	$100	$0
B	$100	$0	$100
*B**	$0	$100	$100

Most people prefer *A* to *A** and *B** to *B*. People tend to prefer risk to equivalent levels of uncertainty when they have something to gain and to prefer uncertainty to risk when they have something to lose. Thus, *A* is preferred to *A** because it has $100 riding on a prospect of known risk 0.33, while *A** has that same sum riding on an uncertainty (ranging between risk 0 and risk 0.66). Likewise, *B** is preferred to *B* because it offers a definite 0.66 risk of $100 where *B* offers only an uncertainty (ranging between risk 0.33 and risk 1.0).

Some decision theorists take the Allais and Ellsberg paradoxes to show that expected utility theory is incapable of capturing rational attitudes toward risk. The problem with STP, they say, is that a rational agent need not be able to form any definite preference between the act types *X* and *X** because information about *x*'s value might provide information about the relative *risk* of options, and this information can be relevant to the agent's preferences.

Many expected utility theorists respond to this objection by arguing that the Allais and Ellsberg paradoxes are *underdescribed* (Broome 1991, 95–115). One can render the usual Allais preferences consistent with STP by rewriting outcomes as follows:

	0.33	0.01	0.66
A	$2,500	$0 instead of a sure $2,400 with A*	$2,400
*A**	$2,400	$2,400	$2,400
B	$2,500	$0 instead of a probable $0 with B*	$0
*B**	$2,400	$2,400	$0

If the agent prefers the second outcome in *A* to the second outcome in *B*, then there is no violation of STP. Moreover, there is a plausible psychological explanation for this preference. A person who ends up with $0 instead of a sure $2,400 might experience pangs of regret that would not be felt if that person thought that ending up with $0 was likely anyhow. Thus, the person's decision really looks like this:

	0.33	0.01	0.66
A	$2,500	$2,400 and pangs of regret	$0
*A**	$2,400.	$2,400	$2,400
B	$2,500	$0 with little regret	$0
*B**	$2,400	$0	$2,400

The Ellsberg paradox can be handled similarly. If the agent feels a special sort of discomfort when gains ride on uncertain prospects (or losses ride on risky prospects), then the correct description of this problem might really be:

	Red	White	Blue
A	$100	$0	$0
*A**	$0	$100 and discomfort	$100
B	$100 and discomfort	$0	$100 and discomfort
*B**	$0	$100	$100

Again, there is no violation of STP here.

This way of eliminating counterexamples to STP worries many people, since it looks like an expected utility theorist can *always* use it. The fact that one can *always* explain away any seeming counterexamples to expected utility theory by redescribing outcomes and postulating the necessary beliefs and desires seems to show that the theory is contentless. This objection is especially effective against behaviorist interpretations of preference. Since behaviorists can appeal to only overt choices to isolate preferences, they have no principled way of distinguishing legitimate from ad hoc redescriptions of decision problems. Nothing in an Allais chooser's behavior, for example, indicates whether the agent is seeking to avoid some (unobservable) feeling of regret or is acting on the basis of nonseparable attitudes toward risk.

The objection is less effective when preferences are understood as all-things-considered judgments, for it is then possible to argue that certain redescriptions are correct because they *best explain* the totality of the person's behavior. If the hypothesis that people experience regret explains a great deal of human behavior, aside from violations of STP, then it is legitimate to use it to explain the common Allais preferences. Consider an analogy: It could be claimed that Newtonian mechanics is empty because (as is true) any pattern of observable motions can be made consistent with Newton's laws by positing the right constellation of forces. What makes this objection unconvincing is the fact that Sir Isaac Newton was able to account for a vast array of distinct motions using the single force of gravity. The same might be true in decision theory. If it can be shown that a small number of relatively simple psychological mechanisms, including feelings of regret or discomfort in situations of risk, explain a great deal of human behavior, then the *best explanation* for the Allais and Ellsberg choices might be among those proposed by the expected utility theorists. Of course, this places a burden on these theorists to show that by standard canons of scientific reasoning, their explanation is indeed the best available.

An alternative response is to take the description of the Allais and Ellsberg paradoxes at face value and to argue that the common preferences are irrational. To see how the argument might go for the former, note that the common rationale for the Allais choices assumes that the difference in risk between *A* and *A** exceeds the difference in risk between *B* and *B**. Proponents of expected utility theory will argue that this is mistaken. The best way to determine how much two options differ in risk, they will claim, is to ask how one might ensure against the increased chances of loss that one assumes in exchanging the less risky option for the more risky one. Someone who switches from *A** to *A* in the Allais paradox can ensure against the risk of loss by buying an insurance policy that pays out $2,400 contingent on the 0.01 probability event. Moreover, the person can ensure against the incurred risk of switching from *B** to *B* by purchasing *the same policy*. Since a single policy does both jobs, the actual change in risk must be the same in each case. Allais choosers, who perceive a greater risk in the first switch, are committed to paying more for the policy when using it as insurance against the *A**-to-*A* risk than when using it as insurance against the *B**-to-*B* risk. This difference in "risk premiums" shows that the Allais choosers' perceptions of risk do not track the actual risks of prospects. Similar things can be said about Ellsberg choosers.

Opponents of expected utility theory may deny that it is appropriate to measure risk by the costs of ensuring against it. In the end, the issue will be settled by the development of a convincing method for measuring the *actual* risks involved in prospects. (Ideally, this theory would be augmented by a plausible psychological account of *perceived* risks that explains the common Allais choices.) While there is a well-developed model of risk *aversion* within expected utility theory, this model does not seek to measure risk itself, only an agent's *attitudes* toward risk. While some progress has been made in the measurement of risk, a great deal remains to be done. It is known that no simple measure (standard deviation, mean absolute deviation, entropy) will do the job. Building on the classic paper by Michael Rothschild and Joseph Stiglitz (1970), economists have made great strides toward providing a definition of the "riskier than" relation. This work strongly suggests that risk is indeed a separable quantity, and thus that the Allais and Ellsberg choosers are irrational. Still, there is no universally accepted way of measuring the amount of risk that prospects involve. Until such a measure is found, the proper interpretation of the Allais and Ellsberg paradoxes is likely to remain controversial, as will expected utility theory itself.

James M. Joyce

References

Allais, Maurice (1990), "Allais Paradox," in J. Eatwell, M. Millgate, P. Newman (eds.), *The New Palgrave: Utility and Probability*. New York: Norton, 3–9.

Anderson, Elizabeth (1993), *Value in Ethics and in Economics*. Cambridge, MA: Harvard University Press.

Broome, John (1991), *Weighing Goods*. Oxford: Blackwell Publishers.

Ellsberg, Daniel (1961), "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics* 75: 643–669.

Gibbard, Allan, and William Harper (1978), "Counterfactuals and Two Kinds of Expected Utility," in C. Hooker, J. Leach, and E. McClennen (eds.), *Foundations and Applications of Decision Theory*. Dordrecht, Netherlands: Reidel, 125–162.

Hammond, Peter (1988), "Consequentialist Foundations for Expected Utility Theory," *Theory and Decision* 25: 25–78.

Jeffrey, Richard (1983), *The Logic of Decision*, 2nd rev. ed. Chicago: University of Chicago Press.

Joyce, James M. (1999), *The Foundations of Causal Decision Theory*. New York: Cambridge University Press.

Kahneman, Daniel, and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrika* 47: 263–291.

Kaplan, Mark (1983), "Decision Theory as Philosophy," *Philosophy of Science* 50: 549–577.

Karni, Edi (1985), *Decision Making Under Uncertainty: The Case of State-Dependent Preferences*. Boston: Harvard University Press.

Levi, Isaac (1991), "Consequentialism and Sequential Choice," in M. Bacharach and S. Hurley (eds.), *Foundations of Decision Theory*. Oxford: Blackwell, 92–112.

McClennen, Edward (1990), *Rationality and Dynamic Choice: Foundational Explorations*. Cambridge: Cambridge University Press.

Ramsey, Frank (1931), "Truth and Probability," in R. Braithwaite (ed.), *The Foundations of Mathematics and Other Logical Essays*. London: Kegan Paul, 156–198.

Raz, Joseph (1986), *The Morality of Freedom*. Oxford: Clarendon Press.

Redelmeier, Donald, and Eldar Shafir (1995), "Medical Decision Making in Situations That Offer Multiple Alternatives," *Journal of the American Medical Association* 273: 302–305.

Rothschild, Michael, and Joseph Stiglitz (1970), "Increasing Risk: I. A Definition," *Journal of Economic Theory* 2: 225–243.

Savage, L. J. ([1954] 1972), *The Foundations of Statistics*, 2nd rev. ed. New York: Dover Press. Originally published by John Wiley and Sons.

Sen, Amartya K. (1971), "Choice Functions and Revealed Preference," *Review of Economic Studies* 38: 307–317.

Shafir, Eldar (1993), "Choosing Versus Rejecting: Why Some Options Are Both Better and Worse Than Others," *Memory and Cognition* 21: 546–556.

Shafir, Eldar, and Amos Tversky (1995), "Decision Making," in E. Smith and D. Osherson (eds.), *An Invitation to Cognitive Science*, vol. 3: *Thinking*, 2nd ed. Cambridge, MA: MIT Press, 77–100.

Skyrms, Brian (1980), *Causal Necessity*. New Haven, CT: Yale University Press.

Tversky, Amos, and Daniel Kahneman (1986), "Rational Choice and the Framing of Decisions," *Journal of Business* 59: S251–S278.

See also **Bayesianism; Game Theory**

PROBLEM OF DEMARCATION

"The problem of demarcation" is Karl Popper's label for the task of discriminating science from nonscience (see Popper, Karl Raimund). Nonscience includes pseudoscience and metaphysics but also logic and pure mathematics, philosophy (including value theory), religion, and politics (Popper 1959, 34; 1963, Ch. 1). Pseudoscience and metaphysics in turn include meaningless language, speculative theory, ad hoc conjectures, and some of what is today called junk science. Given this wide range of targets, the great diversity of legitimate sciences and philosophies of science, and human gullibility, it is not surprising that there is no agreement on whether there is an adequate decision procedure or criterion of demarcation and, if so, what it is. Meanwhile, the logical empiricists formulated their own criterion of demarcation in terms of their empiricist philosophy of language, an approach that Popper rejected (see below).

Traditional solutions to the problem of demarcation have been attempts to answer such questions as, What is science? What is special about science? What is (empirical) knowledge? and, by implication, Why is science important? There is much at stake for a society in the answers to such questions insofar as science enjoys cultural authority in that society. Ironically, in recent decades, the problem of demarcation has lost visibility in philosophical circles even as science and technology have gained unparalleled authority and even though creationists and various postmodernist groups now increasingly challenge that authority, not to mention the legal and political difficulties in identifying "sound science" (see Social Constructionism).

The distinction between science and nonscience is not always invidious. In his *Tractatus Logico-Philosophicus* of 1919, Wittgenstein drew the distinction in part to protect ethics from the incursions of science. The logical empiricists used it to distinguish philosophy from empirical science (see Cognitive Significance; Logical Empiricism).

Problems of demarcation arise at two different levels. One is the public level. Given the centrality of science and technology to modern societies, those societies do in fact demarcate allegedly good science from bad science and from nonscience in various ways. The question is, How *ought* they to accomplish such demarcation? At the other level, the same question arises within specialist scientific communities, although here the basis for discrimination will normally be more technical.

Historical Background

From the ancient Greeks on, Western methodologists have attempted to solve the problem by specifying a *criterion* of demarcation in the form of necessary and sufficient conditions for *epistēmē*, *scientia*, or good science. Historically prominent criteria of demarcation draw upon virtually all the main areas of philosophy. Criteria have been couched in terms of the ontological status of the objects of knowledge (e.g., Platonic Forms, Aristotelian essences), the semantic status of the products of research (science as a body of true or at least meaningful claims about the universe), the epistemological status of the products of research (science as a body of certain or necessary or reliable or appropriately warranted claims), the logical form of those claims (universal or particular, derivability of predictions from them), and value theory (the normative method that produces the claims, e.g., inductive or hypothetico-deductive, or comparison of a field with a model discipline such as physics).

For Aristotle, a claim is scientific if it is

- general or universal,
- absolutely certain, and
- causal-explanatory.

Here "general" means that the claim is of the form "All *A*s are essentially *B*s," and "causal-explanatory" means that the argument "All *A*s are (essentially) *B*s, and such-and-such is *A*; therefore, it is *B*" explains why it is *B* by attributing *A* as the cause. The possessor of genuine scientific knowledge has a demonstrative understanding of the first causes or essences of all things of a given kind. The logic or methodology of science and the investigative process itself are distinct from science proper. Aristotle stated his demarcation criteria in terms of the qualities of the products, not the process of producing them.

Two thousand years later, Galileo, Descartes, Newton, and other seventeenth-century natural philosophers still required virtual certainty for a claim in order to include it in the corpus of scientific knowledge. These early scientists also required causal-explanatory power of a sort; witness Newton's goal of finding true causes (*verae causae*) in his first rule of reasoning in the *Principia.* However, many of the natural philosophers abandoned as impossible Aristotle's demand for first causes and real essences. Newtonian mechanics could demonstrate the motion of the planets in terms of the laws of motion and gravitation, but it failed to find either the cause or the essence of gravity itself. Thus it could not provide a demonstrative chain of reasoning back to first causes (McMullin 2001).

In the wake of the English Civil War, the Royal Society of London expressly excluded religion and politics from its discussions and insisted that scientific discourse be conducted in plain (nonmetaphorical) language. Although Descartes had previously rejected rhetoric and the other humanities as bases for science, the secular saint of the Royal Society was Francis Bacon, who was usually interpreted as a simple inductivist. In this view, to be scientific, a claim must be induced from a body of previously gathered experimental or observational facts. Nature must be allowed to speak first as well as last.

The thinkers of the scientific Enlightenment shaped the modern concern with demarcation. If science is to be the supreme expression of human reason and the broom that sweeps away the cobwebs of tradition and folk wisdom, then it is crucial to distinguish science from pretenders (Amsterdamski 1975, 29). The Enlightenment legacy is that science and representative government are the two sacred institutions of modern society and that the special status of both must be preserved. Somewhat ironically, then, demarcation became a conservative exercise in exclusion. In its strongest versions, the demarcation project is associated with foundationist epistemologies. In particular, strong empiricists have regarded any claim with a suspicion proportional to its distance from experimental observation. (They have legitimized mathematics in a different way.)

In the eighteenth century, Hume insisted that natural science must be thoroughly empirical, since pure reason cannot say anything about the natural world (see Empiricism). Moreover, any meaningful expression must be traceable back to an origin in experience. Kant also insisted that it is philosophy's job to demarcate science from nonscience—but on *a priori* grounds—and also to adjudicate disputes among the sciences. His philosophy became especially influential because it was incorporated within the newly reformed German university system.

In the nineteenth century there was widespread agreement that "Baconian" induction was an overly restrictive method, and that the hypothetico-deductive method was not only legitimate but also superior, given that certainty is unattainable in science. The hypothetico-deductive method cannot achieve certainty, because of the fallacy of affirming the consequent, but neither can the inductive method that it largely supplanted. Some nineteenth-century and virtually all twentieth-century methodologists responded to the clarified logical situation by becoming fallibilists and by adopting self-correcting or successive-approximation methodologies of science in place of the old foundationist ones. Since these methodologists could no longer appeal to fail-safe epistemic status as the mark of substantive scientific claims, some retreated to the *process* that produced them: A claim is scientific if and only if it is produced by a proper application of the scientific *method*, and a discipline is scientific if and only if it is governed by the scientific method. This view enjoys considerable currency today, especially among textbook writers and school administrators. Of course, process or method had been part of the Baconian criterion all along, but the new dispensation considerably broadened what counted as a legitimate process as well as dropped the near-certainty of the final product.

One difficulty with this retreat from substance to method is that it becomes harder to defend the view that science, unlike other institutions, cumulatively produces objectively correct results. Another difficulty of the liberalized conception of process is that there was and is no agreement about whether there is a special scientific method at all, and if so, what that method is, what justifies its use, and whether it has changed historically (see Feyerabend, Paul). After all, how can anyone prove that a particular candidate for "the" scientific method is bound to produce results epistemically superior to those of any other method? Indeed, how can one know in advance whether or not a given method will be fruitful in this world or in any given domain of it? Ironically, prior to the Darwinian era, many methodologists would have said that the ultimate justification of method is theological. Third, John Herschel, William Whewell, Auguste Comte, W. S. Jevons, and others minimized the process of discovery in favor of the testability of the products of that process. Reversing the Bacon-Hume emphasis on antecedents, they asserted that it is observable consequences—predictions—that count, and that novel predictions count most (see Prediction). John Stuart Mill denied that novel predictions have special weight and remained a more conservative inductive-empiricist than Whewell and the hypotheticalists, who increasingly stressed the importance of conceptual and theoretical innovation as well as novel prediction. Popper and the logical empiricists would later reconstruct this divorce of antecedent exploration from logical consequences as involving a distinction of the psychological "context of discovery" from the logical "context of justification or corroboration," thereby reducing scientific method (the "logic" of science) to a logical minimum. As a criterion of demarcation, testability drastically weakens Aristotle's standard, for now a scientific statement need not even be supported by evidence, let alone be established as true (Laudan 1981, Chs. 9–11; Nickles 1987a).

Twentieth-Century Developments

The problem of demarcation was a central feature of the dominant philosophies of science—logical empiricism and Popperianism—at the time when philosophy of science emerged as a professional specialty area within philosophy, namely the period 1925–1970 (see Logical Empiricism; Popper, Karl Raimund; Vienna Circle). In *Tractatus*, §4.11, Wittgenstein had written, "The totality of true propositions is the whole of natural science (or the whole corpus of the natural sciences)." Inspired in part by Wittgenstein, some leading logical empiricists adopted not actual truth but empirical verifiability as their criterion of demarcation of science from metaphysics and pseudoscience. The verifiability criterion served as a criterion of both empirical meaningfulness and meaning itself. Given that a statement is meaningful (verifiable), what exactly does it mean? Roughly, the meaning is given by the method of verification. However, the logical empiricists soon realized that the unrestricted principle is untenable. It excludes mathematics and logic from science, and abstract theoretical and lawlike claims as well. Besides, its own status is unclear, since it itself is not verifiable (see Verifiability).

Pierre Duhem ([1914] 1954) had already shown, contrary to the logical empiricists and to Popper, that theories such as classical mechanics, Maxwellian electromagnetic theory, and relativity theory are not falsifiable in isolation (see Duhem Thesis). Only the larger complexes that include numerous and diverse auxiliary assumptions yield predictions, a point that Willard Van Quine expanded into a controversial, full-blown holism concerning the relation of science to experience, which in turn encouraged Thomas Kuhn (1962) to emphasize the underdetermination of theory by the facts plus logic (see Kuhn, Thomas; Quine, Willard Van). Paul Feyerabend (1981, Chs. 6 and 8) contended that some of the empirical content of a deep theory could be discovered only from the vantage point of another deep theory (see Feyerabend, Paul).

Meanwhile, the operationalism of the physicist Percy Bridgman and of several behavioral scientists required that the *smallest* linguistic units—namely, individual terms—receive operational definitions (see Behaviorism; Bridgman, Percy). Roughly, the meaning of a term is given by the operations that determine whether or not it applies. For example, the scratch test for minerals indicates the meaning of "Mineral *X is harder than* mineral *Y*." Moreover, in a reversal of the consequentialist tendency of methodological thinking, all the terms were to be defined in advance of any theorizing and were to provide a permanent conceptual basis for future science (see Scientific Revolutions).

Carl Hempel's (1965, Chs. 4–5) influential review of the literature summarized the failures of the various criteria of meaning and demarcation proposed by the logical empiricists and operationalists: They are both too restrictive and too permissive (see Cognitive Significance; Hempel, Carl Gustav). In agreement with Quine's attack on the analytic-synthetic distinction, Hempel (1965) concluded: "Theory formation and concept formation go hand in hand; neither can be carried on successfully in isolation from the other" (113). Thereafter, these programs faded in importance (see Analyticity).

The problem of demarcation is most closely associated with Popper, for whom it and the related problem of the growth of knowledge were the two central problems of philosophy (Popper 1959, 34; 1963, Ch. 1). Although both parties regarded empirical testability as the mark of a scientific statement, Popper disagreed with the leading logical empiricists on two main points. First, falsifiability alone counted as testability for Popper (a view that he had to soften in order to allow statistical-probabilistic statements). Second, refusing to take the linguistic turn, Popper rejected the logical empiricists' attempt to derive a demarcation criterion from a theory of meaning. Rudolf Carnap, for example, characterized philosophy of science as *Wissenschaftslogik*, the logic of (the language of) science, and required that all scientific statements meet the above-mentioned empiricist criterion of cognitive significance (see Carnap, Rudolf). All nonempirical statements are "metaphysical" and cognitively meaningless. Popper agreed that metaphysical statements are not scientific, but he insisted that they may be meaningful. He observed that the deepest scientific problems have their roots in metaphysical problems, which have served as a heuristic for modern science.

Popper also rejected inductive criteria of demarcation. A good hypothesis does not have to be winnowed from the empirical facts; nor can it be. Noting the logical asymmetry between verification and falsification (a single counterinstance can refute a universal claim, but no number of positive instances can verify it), Popper proposed falsifiability (empirical refutability) as the criterion of demarcation. For him a statement is scientific if and only if it is falsifiable in principle, that is, if it can fail an empirical test. This is equivalent to saying that there must be some *possible* observation statement (true or false) that logically contradicts the claim in question. Thus Newton's and Einstein's bold theories are scientific because they make risky empirical claims that can fail; but Marxist and Freudian theories are not scientific, despite their claims to wide explanatory power, because their advocates allow nothing to count as a refutation.

Far from making strong claims about reality, these theories actually exclude nothing.

Popper criticized the age-old effort of philosophy to justify scientific claims positively. According to Popper, Hume had shown that scientific claims can never be justified, even to a probabilistic degree. For this reason he spoke of successful tests as "corroborating" rather than confirming hypotheses (see Corroboration). Scientific claims can never escape conjectural status. By contrast, Carnap and other logical empiricists responded to Hume with *probabilism,* the view that the facts can confer probability but not certainty upon universal empirical claims (see Confirmation, Theories of; Inductive Logic).

In another respect, Popper retained the view that science is a special institution, the one that best exemplifies rational and empirical inquiry into new knowledge, or what Popper called "the critical approach." The scientific community is an open society and a model for the others. But since almost any discipline can pursue a critical approach, why does demarcation remain the central problem of his epistemology? Popper's answer was that his criterion of demarcation is the key to solving the problem of the growth of knowledge, that is, the problem of how people learn from experience. The solution to this problem is that they learn from their mistakes by identifying and eliminating error, not the traditionally offered solution that they learn by induction from experience. It is falsification that enables them to learn from experience without induction—the only possible way of learning, he thought. Thus falsifiability is the crucial feature that makes learning and hence science possible. Disciplines that do not promote their own rational and empirical criticism not only are intellectually dishonest but also obstruct the advance of knowledge.

Popper's intuitively appealing criterion is widely cited in public fora. However, it too runs afoul of the objections reviewed by Hempel. For example, Popper's criterion fails utterly for singular statements such as "There are black holes," for unrestricted existential statements are not falsifiable. Moreover, it is not always clear whether Popper is criticizing Marxists and Freudians themselves or the theories they hold. Whether or not Marxists are personally responsive to reasons and evidence is one question; the testability of Marxist theory is another. A theory cannot be dishonest.

Imre Lakatos (1970) provided a thorough critique of the entire spectrum of falsificationist positions and, on this basis, arrived at his own "methodology of scientific research programs," which, in effect, makes demarcation of good from bad science and nonscience a matter of degree (see Lakatos, Imre; Research Programs). Because of Duhem's problem and other complications, what are appraised, Lakatos said, are not theories in isolation but entire series of theories generated appropriately by ongoing research programs (see Duhem Thesis). Competing programs fight long battles of attrition. A research program progresses insofar as

- it makes novel theoretical predictions in heuristically motivated (non–ad hoc) ways;
- some of these predictions are confirmed; and
- successor theories in the program can explain why their predecessors worked as well as they did.

A program degenerates insofar as it lags in these respects. Lakatos did not apply his approach with Popperian ruthlessness, for Lakatos held that it is not necessarily irrational to retain allegiance to a degenerating program for an indefinite period of time. His account implies that the predicates "is scientific" and "is good science" are relative to historical context. Phlogiston, caloric, and ether theories may have been the best available in their day, but anyone defending them today is surely unscientific. Like Popper, Lakatos and successors such as Worrall and Zahar attempt to purify science of ad hoc statements, roughly, theory modifications that are heuristically unmotivated and that lead to no new predictions. But they disagree in detail on what counts as ad hoc and why ad hoc science is bad science (Nickles 1987b).

For Kuhn (1962 and 1970) the mark of a mature science is that it supports a routine problem-solving tradition with a disciplined determination of which problems are fruitful to pursue (see Kuhn, Thomas). In his terms, this will be normal science under a paradigm. Kuhn contested Popper's treatment of falsification and falsifiability. Astrology was (and is) a pseudoscience not because it was unfalsifiable but because it could not sustain a normal-scientific puzzle-solving tradition. Kuhn's view of creation science was similar. Interestingly, Kuhnian normal science fails to be scientific by Popper's criterion, for it is uncritical and convergent rather than divergent. It does not seek major novelty. Kuhnian paradigms are not falsifiable, for their constitutive principles and practices retain a virtual synthetic a priori status for the corresponding scientific communities. Popperian falsifications become Kuhnian anomalies. Its convergent nature enables normal science to build on itself more directly than Popper allowed, yet, in the longer run, given the inevitability of scientific revolutions, not as cumulatively as even Popper wanted. And to Popper's prohibition of ad hoc adjustments in the face of threatened falsification, Kuhn replied that it is often by just such adjustments that scientific knowledge grows. While Popper minimized the importance of scientists' *believing* in scientific propositions, Kuhn emphasized commitment, but to a tradition of problem-solving practices more than to a set of specific beliefs and methodological rules (Rouse 2003).

It is largely on the basis of their implicit, practical knowledge that scientists within a specialist community agree so readily on what is good and bad science. However, this harmony of mutual comprehension goes out the window during a scientific revolution. Insofar as Kuhnian scientific revolutions actually occur in the history of science, they make solving the problem of demarcation more difficult; for by their nature, they radically undermine not only entrenched theories but also the goals, standards, and methodological practices that characterize the previous periods of normal science under a paradigm. They represent discontinuities of scientific development.

Demarcation Difficulties and Reasons for the Demise of the Philosophical Problem

Aside from the particular difficulties faced by each specific proposal, the demarcation enterprise as a whole faces some obstacles. First, the general epistemological problem of demarcating knowledge from nonknowledge is essentially an application of *the problem of the criterion,* formulated in ancient times by Sextus Empiricus. As such, the problem generates a well-known destructive dilemma. How does one decide whether a proposed criterion is correct?—for, either the candidate is self-certifying or else it is justified by a deeper criterion. The first option leads to vicious circularity and the second to vicious regress. (Simply to identify the demarcation and criterion problems would be to classify all genuine knowledge as scientific knowledge, a version of scientism.) Moreover, if the criterion of demarcation itself is an empirical claim, then how can it also be normative, and how can it be justified by appeal to empirical science without begging the question? Both the logical empiricists and the Popperians attempted to evade these difficulties by assigning the criterion of demarcation to methodology or philosophy of science rather than to empirical science itself. With some exceptions (notably Otto Neurath), they held that *philosophy* of science is an a priori or conventional and normative discipline (see Neurath Otto). But if the criterion is merely conventional, a matter of social agreement, then how can it escape being historically situated or historically relativized to competing research programs, and why is everyone obligated to accept the convention?

For Popper, the criterion of demarcation, as one of the "rules of the game of science," regulates science as a whole. Kuhn denied that there are any such rules. Larry Laudan (1981) observed that methodologists have typically used demarcation criteria and other methodological principles as *machines de guerre* in specific historical battles for control of some science or program. Moreover, since technical judgments as to what constitutes good science in a particular subspecialty are usually highly field- and problem-specific, they are not credible unless made by respected members of the research community. Philosophers and general methodologists rarely occupy this position.

The new history of science brought a historical sensitivity to philosophy of science that problematizes the entire project of demarcation. (The logical empiricists had discontinued Ernest Mach's and Duhem's practice of studying the history of science with care.) If a criterion of demarcation is supposed to answer the question, What is science? by delineating what is common to all sciences at all times, an essentialist answer is almost inevitable. Yet the diversity of past science is already so great that any criterion that encompasses all of it is bound to be too weak to be interesting or useful. Ironically, the problem of demarcation for modern science became urgent only when the sciences began to diversify sufficiently that no simple criterion was likely to succeed. These observations raise a further difficulty. It must first be decided which enterprises to include among the sciences to be compared, and this already begs the demarcation question (Amsterdamski 1975).

One reason why candidate criteria of demarcation are too narrow is that they are often backward-looking, attempting to capture what all successful sciences up to now have in common, while pretending to be suprahistorical. Yet science continues to evolve, to ramify, to diversify, to redefine itself; and this is true of methodology and goals as well as content. Insofar as scientific change is occasionally revolutionary, the difficulty may be even worse. On what basis could a criterion of demarcation formulated today presume to legislate for all future science? A criterion with any bite is likely to harm science more than help it. In giving up the search for ultimate essences, both Ptolemaic astronomy and Galilean mechanics failed to be science by Aristotle's lights (Laudan 1983).

Many philosophers of science (as well as most science studies experts) have rejected or minimized the problem of demarcation for one reason or another. Pragmatists such as Quine tend to blur dichotomous distinctions, and this one is no exception (see Quine, Willard Van). For Quine, philosophy is continuous with science, which is in turn continuous with common sense. There is no sharp distinction between purely analytical activity and empirical investigation. And today, in the so-called postmodern era, there is a premium on discriminating the *differences* among the wide variety of seemingly legitimate scientific pursuits rather than upon identifying characteristics that they possess in common. Many science studies experts and culture theorists contest the cultural authority of science and the traditional claims that the scientific enterprise possesses a unique epistemic status (see Social Constructionism). Yet one could say that the very success of the Enlightenment project has made demarcation more difficult and less necessary today, since all major institutions strive to be more rational and scientific and less arbitrary in their practices, including some approaches to theology (Murphy 1990).

Reflecting on the steady weakening of proposed criteria of demarcation, Laudan (1983) concludes that demarcation is no longer an important philosophical problem. Popper's falsifiability criterion, he says, weakens the criterion almost beyond recognition. No longer does the criterion of demarcation mark out a body of belief-worthy claims about the world, let alone demonstrably true claims, let alone claims about ultimate causal essences—for, in Popper's criterion every empirically *false* statement is automatically scientific! Popper completely abandons the traditional attempt to characterize science in terms of either the epistemic or the ontological status of its products.

Laudan's view is that it is wrong to make invidious, holistic distinctions in advance about whether or not something is scientific. Scientists typically proceed piecemeal, he says, willing to consider anything and everything on its merits. They dismiss as bad, marginal, or fringe science much of what they encounter and keep the rest. There is no need for a separate category of pseudoscience. It is enough to reject something as bad science.

This pragmatic move deliberately blurs the distinction between the form and the content of science, i.e., between the logic or method of science and empirical claims themselves. The move rejects the traditional demarcation problem only to raise another, at least equally difficult one: How can philosophers of science (and other members of society) reliably discriminate good science from bad science? Laudan (and Kuhn) would answer that philosophers do not need to. That is a job for contemporary practicing scientists who have demonstrated their expertise. Sometimes the answers will be obvious, but often enough they will be both piecemeal and highly technical. This response may be correct, but how does it play in the sociopolitical arena? And of course it raises yet another question: How should society determine who is an expert?

Demarcation as a Social Problem

Laudan ([1982] 1996, Ch. 12) applies his position to the Arkansas trial of 1981–1982 (*McLean v. Arkansas*) over the teaching of creationism in public school biology classes. He agrees with the decision that creationism should not be taught as biology, but he is severely critical of every point of Judge Overton's philosophical justification of his decision. For example, Overton appeals to Popper's falsifiability criterion to show that creationism is not science. Laudan replies that creationist doctrine itself *is* science by that criterion, since it has been empirically falsified, however unscientific may be the behavior of some of its advocates. The reason it should not be taught is simply that it is *bad* science. Michael Ruse, who had invoked Popper's criterion in court testimony, responded that given the complexities of the legal and social situation, Judge Overton's reasoning was correct, for that is the only practical way to stop the teaching of "creation science" as a serious alternative to biological evolution (Ruse 1982).

One complication is that terms such as "bad science" and "pseudoscience" cover a variety of different sins, including incompetent but honest work, potentially good work that is difficult to test or that has utterly failed to find empirical support, and deliberately dishonest scientific pretensions. There are any number of ways in which science can be bad and many labels for bad or pretended science. "Pseudoscience" is an old term for claims that are (or are treated as) untestable. The chemist Henry Bauer (2001) prefers "anomalistics" to "pseudoscience," since all the standard criteria for the latter have failed and pseudoscience occasionally develops into real science, as Popper acknowledged. "Fringe science" includes sometimes-testable claims widely ignored by the scientific community because they violate the best naturalistic understanding of the cosmos. "Pathological science" is the name that Nobel chemist Irving Langmuir gave to those cases in which the scientists are honestly deceiving themselves, as he claimed was the case with J. B. Rhine's work on extrasensory perception (Park 2000, 40ff). "Junk science" involves deliberately exploiting scientific uncertainty to confuse and mislead judges, juries, and politicians, usually by substituting mere possibility for known probability (Huber 1991). It falls just short of "fraudulent science," in which scientists fudge their results or expert witnesses lie about the current state of knowledge. The physicist Robert Park (2000) lumps all these cases together as "voodoo science." He is especially concerned about claims with public currency that escape full scientific scrutiny because of official secrecy, political intervention, the legal adversary system, and the de facto adversary system employed by the media. The latter results in what Toumey (1996, 76) calls "the pseudosymmetry of scientific authority." That is, "unbiased reporting," like the use of expert witnesses in a courtroom, sometimes pretends that for every expert there is an equal and opposite expert.

This leads to another complication—that philosophers and scientists must make their cases to lay audiences. There is little opportunity to present esoteric detail in a court of law or the popular media. In the case of creationism, given the political, religious, and legal situation in the United States, Ruse can argue rather convincingly that labeling creationism religion rather than science is the only way to keep it from being taught as science in public school classrooms. It is doubtful whether Laudan's more nuanced treatment of the issue would have the same practical effect. Should Judge Overton have ruled that creationism cannot be taught because it is bad science, or that it can *only* be taught as an example of bad science? Surely it would be a bad precedent for sitting judges to rule on what is good or bad science. And yet in a lesser sense they must, for the U.S. Supreme Court's 1993 decision in *Daubert v. Merrill-Dow Pharmaceuticals* makes judges the gatekeepers for keeping unsound science out of the courtroom. *Daubert* requires judges to consider, in addition to their error rate, whether the alleged scientific claims have been tested, whether the claims have been subjected to peer review, and whether the relevant scientific community accepts the claims (full consensus is not required).

A related complication is that legal (and political and public policy) reasoning differs in important ways from scientific reasoning, so one should not expect full convergence between scientific and legal modes of thought and action. For example, scientific conclusions are typically guarded and open to future revision in a way that legal decisions are not. Legal judgments are final (except for appeal) and must be made within a short time span on the basis of the evidence and arguments adduced within that time, whether or not sufficient scientific knowledge is available (Foster, Bernstein, and Huber 1993). The value of a scientific claim or technique often resides in its heuristic potential, not in its known truth or correctness, whereas the judicial system wants the truth now. Scientists seek general understanding of phenomena, whereas judges and attorneys must achieve rapid closure of particular disputes. Scientific conclusions are often statistical (with margins of error given) and not explicitly causal, whereas legal decisions are typically causal and normative (assigning blame), individual, and nonstatistical, although cases involving smoking and cancer have recently broadened the law's conception of scientific reasoning. In the United States and elsewhere, many legal proceedings, both criminal and civil, are explicitly adversarial, whereas scientific competition is adversarial only in a *de facto* way. Scientists rely most heavily on evidential reasons, whereas law courts require all evidence to be introduced via testimony and accepted (or not) on that authority. Consideration of the evidence itself is beyond the purview of the court. The rules of evidence also differ. Judges must decide, in binary fashion, whether or not a given piece of evidence is admissible at all and whether a given witness is admissible as a scientific expert. When there is a jury, the judge instructs it as to what it may and may not take into consideration. In some respects, legal reasoning is more conservative than "pure" scientific reasoning, since lives may be immediately at stake; whereas in science, as Popper says, "our theories die in our stead (Popper 1985, 83).

The current situation is further complicated by the shifting use of the terms "junk science" and "sound science." In the highly litigious context of the United States, "junk science" originally meant dubious claims defended by hired expert witnesses in liability lawsuits, especially against wealthy corporations. While the increasing number of scientifically frivolous lawsuits does indeed threaten the financial stability and the innovative risk-taking of corporations, corporate executives and powerful politicians have corrupted the terminology by labeling as junk science any scientific claim or methodology that challenges their position, and as sound science any claim that favors it (Rampton and Stauber 2001).

Moreover, writing on boundary formation and maintenance in science, the sociologist Thomas Gieryn (1999) contends that the epistemic authority of science derives not from the application of philosophical criteria of demarcation, nor from empirical testing, good practices at the laboratory bench, or competent experimental design. Rather, it is generated downstream where science meets the rest of society: in schools, the media, law courts, etc. It is here where the cultural maps are drawn, he says, with sets of boundaries that confer authority, power, and prestige (or their absence) upon science and other cultural institutions. Of course, when one looks at the detailed ways in which this is accomplished, one finds such things as philosophers' criteria of demarcation being used as weapons, and so, to a degree, the issue comes full circle. From this sociological perspective, Laudan is correct to characterize criteria of demarcation as local *machines de guerre* rather than timeless principles, but the demarcation problem remains important for all that; for if Gieryn is correct, it is precisely the constellation of such maneuvers that establishes cultural boundaries. That is one reason why clashes over science in the courtroom and clashes between the epistemic standards of science and law are worrisome to those who wish to preserve the epistemic autonomy of science (see Lynch and Jasanoff 1998.)

Since the general public frequently confuses science and technology, demarcation issues carry over to technological debates. Fallibilism with respect to technological development presents more problems than fallibilism with respect to basic science. At the turn of the twenty-first century, one of the sharpest disagreements among policymakers concerns the so-called precautionary principle, the idea that scientists and technologists should proceed with caution in areas in which they are ignorant. In its strong form, the principle states that no technology shall be introduced until its safety can be assured—a measure that would curtail innovation. Defenders of the strong form apparently assume that "science" consists of fail-safe knowledge. A weaker form of the principle requires caution when the stakes (utilities) are sufficiently high, even when the probabilities are relatively small or uncertain (e.g., greenhouse gases and global warming). Equally clearly, forging ahead until there is proof of danger is a foolish policy. The problem is where and how to strike the balance, given that these are usually decisions made under large uncertainty.

Conclusion

There is no one simple distinction that marks off science (and its potential technological applications) from pseudoscience, or good science from bad. This is the conclusion of the last two generations of philosophers of science, reinforced by modern science studies. The problem is still more complex than many writers have realized, for these are no context-free distinctions. What sort of demarcation is appropriate within science depends upon subtleties of historical and technical context, and what sort of demarcation is appropriate in public policy contexts will likewise depend upon contextual details, including the particular interests at stake. So what began as a metaphysical or logical issue ends up being a concern modulated by pragmatic reasons (Resnik 2000). While there is some truth to the reported demise of the traditional demarcation problem, context-specific demarcation issues abound and are more important than ever, both within science and in the public arena, a domain that urgently needs more philosophical participation. Before emigrating to the United States and England, the logical empiricists, Popper, and Lakatos were deeply engaged in sociopolitical issues.

Clearly, the demarcation problem cannot be solved by simply identifying science with the body of currently accepted "truths," nor is it possible simply to retreat from substance to method if this implies a commitment to "the" scientific method as a set of rules. Current emphasis on future promise rather than past results, and on scientific practices rather than belief systems and universal logical criteria, may offer a more feasible approach to the problem. Rouse (2003, 119) notes the irony that despite Kuhn's and science studies' challenge to "textbook science," the leading philosophical models of science remain representational and hence lend encouragement to the creationists' fideistic conception of science and of science education. Education continues to emphasize correct beliefs (scientific facts) over productive practices and future-oriented attitudes. Half a century after the decline of logical empiricism, philosophers still tend to view science in terms of a retrospective theory of justification rather than problem-solving productivity and future promise. Although it is not absolute, the belief/practice distinction can also eliminate a persistent confusion over who is an expert. At the frontier of knowledge, where no one knows what lies beyond, there are no experts in the sense of those who know *that* such-and-such is true, but there clearly are experts in the sense of those who know *how* to proceed with frontier research and are able to furnish comparative heuristic appraisals of the competing proposals.

Thomas Nickles

References

Amsterdamski, Stefan (1975), *Between Experience and Metaphysics*. Dordrecht, Netherlands: Reidel.

Bauer, Henry (2001), *Science or Pseudoscience?* Urbana: University of Illinois Press.

Duhem, Pierre ([1914] 1954), *The Aim and Structure of Physical Theory*. Princeton, NJ: Princeton University Press.

Feyerabend, Paul (1981), *Realism, Rationalism and Scientific Method: Philosophical Papers* (Vol. 1). Cambridge: Cambridge University Press, 239–245.

Foster, Kenneth, David Bernstein, and Peter Huber (eds.) (1993), *Phantom Risk: Scientific Inference and the Law*. Cambridge, MA: MIT Press.

Gieryn, Thomas (1999), *Cultural Boundaries of Science: Credibility on the Line*. Chicago: University of Chicago Press.

Hempel, Carl G. (1965), *Aspects of Scientific Explanation*. New York: Free Press.

Huber, Peter (1991), *Galileo's Revenge: Junk Science in the Courtroom*. New York: Basic Books.

Kuhn, Thomas (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

——— (1970), "Logic of Discovery or Psychology of Research?" in I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 1–23.

Lakatos, Imre (1970), "Falsification and the Methodology of Scientific Research Programmes." in Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 91–195.

Laudan, Larry (1981), *Science and Hypothesis*. Dordrecht, Netherlands: Reidel.

——— (1983), "The Demise of the Demarcation Problem," in R. S. Cohen and L. Laudan (eds.), *Essays in Honor of Adolf Grünbaum*. Dordrecht, Netherlands: Reidel, 111–127.

——— ([1982] 1996) *Beyond Positivism and Relativism*. Boulder, CO: Westview Press, 1996.

Lynch, Michael, and Sheila Jasanoff (eds.) (1998), *Social Studies of Science* 28: 5–6. Special issue on "Contested Identities: Science, Law and Forensic Practice."

McMullin, Ernan (2001), "The Impact of Newton's *Principia* on the Philosophy of Science," *Philosophy of Science* 68: 279–310.

Murphy, Nancey (1990), *Theology in the Age of Scientific Reasoning*. Ithaca, NY: Cornell University Press.

Nickles, Thomas (1987a), "From Natural Philosophy to Metaphilosophy of Science," in Robert Kargon and Peter Achinstein (eds.), *Kelvin's Baltimore Lectures and Modern Theoretical Physics: Historical and Philosophical Perspectives*. Cambridge, MA: MIT Press, 507–541.

——— (1987b), "Lakatosian Heuristics and Epistemic Support," *British Journal for the Philosophy of Science* 38: 181–205.

Park, Robert (2000), *Voodoo Science: The Road from Foolishness to Fraud*. Oxford: Oxford University Press.

Popper, Karl (1959), *The Logic of Scientific Discovery*. London: Hutchinson. This is a translation and expansion of Popper's *Logik der Forschung* of 1934.

——— (1963), *Conjectures and Refutations*. New York: Basic Books.

——— (1985), *Popper Selections*. Edited by David Miller. Princeton, NJ: Princeton University Press.

Rampton, Sheldon, and John Stauber (2001), *Trust Us, We're Experts!* New York: Tarcher/Putnam.

Resnik, David (2000), "A Pragmatic Approach to the Demarcation Problem," *Studies in History and Philosophy of Science* 31: 249–267.

Rouse, Joseph (2003), "Kuhn's Philosophy of Scientific Practice," in T. Nickles (ed.), *Thomas Kuhn*. Cambridge: Cambridge University Press, 101–121.

Ruse, Michael (1982), "Pro Judice," *Science, Technology, and Human Values* 7: 19–23.

Toumey, C. (1996), *Conjuring Science*. New Brunswick: Rutgers University Press.

See also **Carnap, Rudolf; Cognitive Significance; Duhem Thesis; Empiricism; Feyerabend, Paul; Lakatos, Imre; Logical Empiricism; Kuhn, Thomas; Popper, Karl Raimund; Prediction; Quine, Willard Van; Research Programs; Social Constructionism**

DETERMINISM

Determinism is a topic of broad interest in philosophy, with important connections to issues in metaphysics, epistemology, ethics, and philosophy of action (see e.g., O'Connor 1995; Belnap 2001). Within the philosophy of biology, there has been some discussion of determinism in evolutionary theory (e.g., Brandon and Carson 1996; Graves, Horan, and Rosenberg 1999) as well as in genetics, where a consensus has emerged that no interesting thesis of determinism can be sustained (Sarkar 1998; Kitcher 2001). Most discussion of determinism in the philosophy of science has focused on the issue as it arises in physics, and this will be the primary focus of this article.

Formulations of Determinism

The thesis of determinism has been defined in numerous ways. The basic idea is that one part of the world's history determines another part of the world's history. In order to extend this intuitive but vague idea into a crisp thesis, it is necessary to make some choices. Which part of the world's history does the determining, and which part gets determined? How should the kind of determination at issue be understood? Should determinism be thought of as a characteristic of a world, a theory, a set of laws, or something else? Is determinism an all-or-nothing affair, or are there useful notions of degrees of determinism, or "partial determinism"? This section will survey a number of different ways in which these questions have been answered.

Perhaps the most famous exposition of the doctrine of determinism in the context of modern science is due to Pierre Laplace:

> We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes. (Laplace [1814] 1951, 282)

Laplace's formulation makes a claim about the nature of the universe, *viz.*, that the total state at one time causally determines the state at later times and is causally determined by the state at past times. It also makes a claim about predictability, namely that an idealized intelligence would be able to predict (or retrodict) the total state of the universe at any time given a specification of the state at any other time. The first claim is (broadly speaking) metaphysical or ontic, while the second seems to be epistemological. Presumably, Laplace thought that each claim followed from the other, but as will be seen, there are many physical contexts in which this is not so. There has been some controversy about whether determinism should be thought of primarily as an epistemological thesis or an ontic one.

Laplace's 'intelligence' is an extremely unrealistic idealization, and in order to understand his characterization of determinism, one must understand how this idealization is supposed to function. Karl Popper in effect takes Laplace's intelligence to be a limiting case of an actual observer and so takes determinism to be a thesis about the kind of prediction available in principle to actual observers. Actual observers, in addition to having more limited powers of calculation than Laplace's intelligence, must gather the information they use for making predictions from empirical observations. Accordingly, Popper defines *scientific determinism* as

> the doctrine that the state of any closed physical system at any given future instant of time can be predicted, even from within the system, with any specified degree of precision, by deducing the prediction from theories, in conjunction with initial conditions whose required degree of precision can always be calculated . . . if the prediction task is given. (Popper 1982, 36)

Popper thus takes seriously the link between determinism and predictability, where prediction must be done by an agent who is part of the total system the agent wishes to predict; the agent makes allowances for the fact that such predictions can never be expected to be perfectly precise (see Popper, Karl Raimund; Prediction).

Popper goes on to argue (41–85) that scientific determinism is false, even if classical mechanics, which has been traditionally regarded as a deterministic theory, is true. Popper's formulation thus makes determinism an epistemological thesis. Many other contemporary philosophers of science who write about determinism, for example, John Earman (1986, 6–10) and Patrick Suppes (1993), insist that determinism is an ontic or physical thesis and should not be analyzed in terms of epistemological concepts such as predictability. In their view, Laplace's characterization is acceptable only if the reference to the idealized intelligence is understood as an aid to the imagination; what is really important is that the laws of nature, together with the state of the universe at a given time, suffice to determine the state of the world at other times, whether it is in principle possible to exploit this fact to make predictions or not. Thus, it is crucially important to distinguish between determinism proper and predictability.

One may of course accommodate both views about determinism by distinguishing between a predictability sense of determinism and an ontic/physical sense of determinism. But advocates of the two views see more than mere terminological disagreement here. Popper (1982, 8) argues that such nonepistemological conceptions of determinism are not falsifiable and are thus metaphysical rather than scientific. (For criticism of this argument, see Earman 1986, 10.) One argument for the importance of distinguishing between predictability and determinism, given by Suppes (1993, 245–246), concerns Turing machines. A Turing machine is, in an intuitive sense, an outstanding paradigm of a deterministic system. But it is known that there is no algorithm for determining whether an arbitrary Turing machine in an arbitrary configuration will ever halt. Suppes argues that this shows that it is possible for a 'good sense' deterministic system to be in an initial state such that there is no method available for any epistemic subject to predict its future behavior.

If one decides to formulate determinism in a way that distinguishes it from predictability, one still has options. A typical logical-empiricist formulation makes determinism a property of theories and defines deterministic theories syntactically: A theory is deterministic if and only if a set of sentences specifying the state of the world at one time, together with lawlike sentences drawn from the theory, deductively entail sentences characterizing the state of the world at any other time (Nagel 1953, 420–423) (see Nagel, Ernest). The problem with this definition of determinism, pointed out by Richard Montague (1974, 303–304) is that if the set of possible states has the cardinality of the continuum, there will not be enough sentences to describe the state of the world at a given time sufficiently precisely for the required deductions to go through. Hence, the syntactic formulation makes determinism exceedingly fragile.

Montague (1974) proposes an alternative, semantic characterization of deterministic theories. A theory is associated with a class of *models*, in the sense used in formal semantics (see Scientific Models; Theories). Montague then defines a number of senses of determinism. The general pattern is that a theory is deterministic in a given sense if and only if all models of the theory that agree on the state of the world at one time also agree at certain other times. Following the lead of Earman (1986, 12–14), it is possible to strip away much of the formal semantic apparatus employed by Montague and characterize these senses of determinism in terms of physically possible worlds. Let W be the set of physically possible worlds allowed by the theory T. Then:

- T is *futuristically deterministic* if and only if: For any w_1, w_2, W, and any time t, if w_1 and w_2 agree on the complete physical state at t, then they agree on the complete physical state at any time $t^* > t$.
- T is *historically deterministic* if and only if: For any w_1, w_2, W, and any time t, if w_1 and w_2 agree on the complete physical state at t, then they agree on the complete physical state at any time $t^* < t$.
- T is *Laplacian deterministic* (or simply *deterministic*) if and only if T is both futuristically and historically deterministic.

These definitions provide a straightforward explication of the intuitive idea that one part of the world's history determines another part of that history: The physical possibilities allowed by the former leave no room for variation in the latter. These definitions establish a general pattern that can be used to generate further varieties of determinism:

- T is *(X, Y) deterministic* if and only if: For any w_1, w_2, W, if w_1 and w_2 agree on the complete physical state in spatiotemporal region X, then they agree on the complete physical state in spatiotemporal region Y (cf. Earman 1986, 17).

This pattern is useful when one turns to relativistic physics, where there is generally no well-defined sense of 'the state of the world at a time t' (see Space-Time).

Further variations on this theme are available. For example, one might hold that determinism is not an all-or-nothing affair. A theory may be *conditionally deterministic* in the sense that any two physically possible worlds allowed by that theory that agree at all times on a certain range of magnitudes or properties, and agree at a given time about everything, also agree at all times about everything. Alternatively, one may hold that some aspects of the world are deterministic (say, the properties in the set P) and others are not. One way to capture this intuitive idea is as follows:

- T is *Laplacian deterministic in the properties in the set* P if and only if: For any w_1, w_2, W, and any time t, if w_1 and w_2 agree on all of the properties in P at t, then they agree on all of the properties in P at all times.

In this way, one can formulate the idea that, for example, the world is deterministic in its physical aspects but not in its mental aspects. But as many authors have pointed out (Popper 1982, 25–26; Earman 1986, 13–14), determinism with respect to P coupled with the failure of determinism for other properties leads to the consequence that the properties in P must be nomologically independent of those outside of P. In the case where P is the set of physical properties, assumed not to include the mental properties, this leads to epiphenomenalism about the mental. (Other senses of less-than-complete determinism are defined by Montague 1974, 321–324; Earman 1986, 13–14.)

The senses of determinism just discussed take the basic notion to be that of a *deterministic theory*, where the property of determinism is defined by quantifying over all the physically possible worlds allowed by the theory. Alternatively, one can define determinism as a property of a set of laws, proceeding as above, but quantifying over all the possible worlds allowed by that set of laws. Determinism can also be defined as a property of a world. One straightforward way of doing this is to define a world as deterministic just in case the laws of nature of that world are deterministic. But Earman (1986, 12–13) provides a more sensitive way of defining a deterministic world, which may be stated as follows:

- Let W be the class of physically possible worlds relative to w. Then w is (futuristically, historically, Laplacian) deterministic if and only if: For any $w^* \in W$ and any time t, if w and w^* agree on the complete physical state at t, then they agree on the complete physical state (at all times $t' > t$, at all times $t' < t$, at all times t).

This definition allows, for example, that a world may be Laplacian deterministic even if the laws of that world are not Laplacian deterministic. (This means that it could be that the laws alone do not guarantee that the present state of the universe determines its state for all times, but given the actual state of the universe, its entire history is settled by the laws.) But it entails that a sufficient condition for the Laplacian determinism of a set of laws is that each physically possible world allowed by those laws is Laplacian deterministic.

In the case of a theory whose laws take the form of differential equations or partial differential equations, these definitions can be reformulated in terms of boundary value problems and their solutions, eliminating reference to physically possible worlds. So, for example, Laplacian determinism can be reformulated:

- T is *Laplacian deterministic* just in case a physically possible specification of the physical magnitudes at a time t, together with the laws of T, define a well-posed boundary value problem: There exists a unique solution for all time satisfying the boundary values provided by the specified physical magnitudes at t.

As will be seen later, space-time theories allow for additional versions of this formulation, with different characterizations of the boundary conditions.

In the following sections, the question of whether various theories of modern physics are deterministic will be examined. Of primary concern will be the ontic versions of determinism favored by Earman (1986) and Suppes (1993), but occasional reference will be made to the predictability conception of determinism favored by Popper (1982).

Determinism in Classical Physics

Classical physics is traditionally viewed as the very paradigm of a deterministic physical theory (see Classical Mechanics). This is probably in large part due to Laplace's influential formulation, which was produced in the context of a discussion of classical mechanics. However, it is now known that classical physics is not deterministic, in either the predictability sense or the ontic sense.

Classical physics does not satisfy any very interesting requirement of predictability. One counterexample is provided by the three-body problem in Newtonian gravitation theory: No closed analytic solution of the general problem exists, and methods of numerical approximation give predictions that are accurate only for limited periods of time (Suppes 1993, 244–245). Further, many classical systems are known in which future evolution depends so sensitively on small differences in initial data that reliable predictions for arbitrary future times are impossible. One of the most famous is Lorenz's model of the weather (Earman 1986; Suppes 1993). More generally, many classical systems exhibit the feature known as *chaos*, which rules out the possibility of predictability, though not that of Laplacian determinism as defined above. Chaotic systems are for this reason an excellent case for illustrating the way in which the two senses of determinism can come apart. Chaos in classical systems will be discussed in a later section.

Though the failure of predictability in classical physics is now widely appreciated, it is still widely but falsely believed that classical physics does satisfy the ontic formulation of Laplacian determinism. A counterexample to Laplacian determinism in classical physics is provided by any complete, consistent specification of the history of a physically possible world at a given time where this specification together with the laws of classical physics do not determine a unique future (or past) for that time. There are several known such counterexamples, and they come in a variety of kinds.

One interesting class of counterexamples is provided by collisions between perfectly elastic bodies. The most straightforward counterexample of this class is simply a collision of three or more such bodies. For a collision of three such bodies, the standard classical laws of elastic impact determine four equations, concerning conservation of each component of the total momentum and conservation of the kinetic energy. A system of three particles has nine degrees of freedom, which may be reducible to six by means of selecting a frame of reference in which the center of mass of the system is at rest. This leaves more variables than equations, so a unique solution representing the evolution of the system after the collision is not determined (cf. Earman 1986, 38–39).

More exotic elastic-collision counterexamples can be constructed if the possibility of an infinite number of particles is allowed. One such counterexample, discovered by Jon Perez Laraudogoitia (1996), goes as follows. Consider a countable infinite series of perfectly elastic balls of unit mass, laid out in a straight line of unit length, with each ball half the diameter of the preceding one, and the distance between balls decreasing by half with each successive ball (see Figure 1a). Now suppose that the first ball is struck by a cue ball of unit mass moving with unit speed. In one unit of time, the motion will have been communicated to each of the infinitely many balls. Since the balls are all perfectly elastic, each will come to rest when it strikes its successor, so that at the end of one unit of time, all of the balls are at rest, with the n-th ball occupying the initial position of the $(n + 1)$-th ball and the cue ball occupying the initial position of the first ball (see Figure 1b). The important thing to notice is that no ball pops out on the right-hand side; every ball has a successor, so every ball gets stopped. The evolution just described is a solution of the classical equations of elastic impact. But those equations are time-reversal invariant, so the time-reversed process, leading from Figure 1b to the time-reversal of Figure 1a in a unit of time, is also a solution. Now consider a physically possible world in which the balls are in the situation depicted in Figure 1b at all times $t < t^*$. The laws leave open the possibility that the balls will remain in this configuration forever, and they also leave open the possibility that the time-reversal of the original process will occur, starting at any time later than t^*. This shows that the classical mechanics of elastic particles is not a deterministic theory (in any of the senses discussed above).

A similar counterexample, which uses elastic balls of the same size but initially distributed over an infinite region of space, is given by Oscar Lanford (1975). A curious feature of such counterexamples is that they involve violations of global conservation of momentum and energy; for example, in the Laraudogoitia example, the total momentum and energy of the system is zero before t^* but nonzero after t^*. But there is no violation of *local* conservation of energy or momentum, for the conservation laws are satisfied by each of the elastic collisions. Whether the violation of the global conservation laws is sufficient to show that these are not genuine counterexamples to determinism in classical physics is a point that will be addressed at the end of this section.

A second class of counterexamples to determinism in classical physics involves systems of massive particles governed by the classical law of gravitation. It is possible for such systems to exhibit *singularities* in which some quantity of motion figuring in a law of classical physics becomes arbitrarily large in a finite amount of time (see Classical

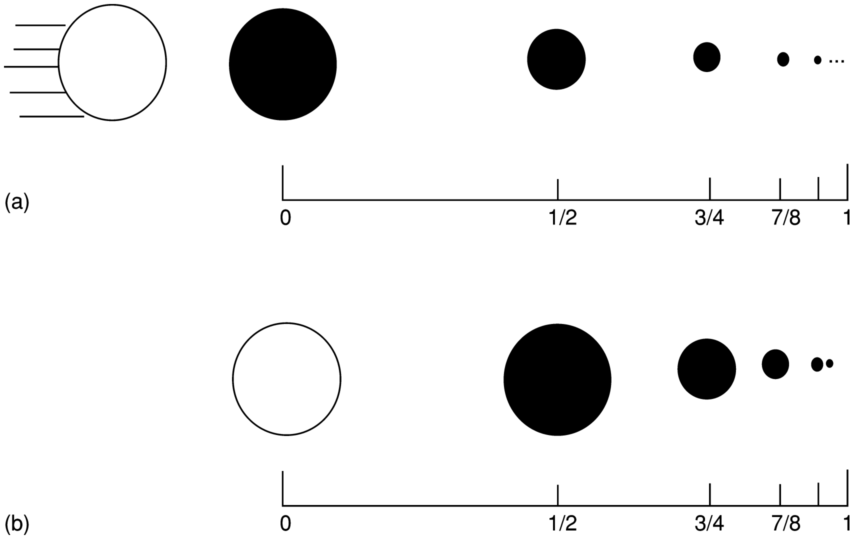


Fig. 1. (a) The Laraudogoitia example: Initial state. (b) The Laraudogoitia example: Final state.

Mechanics). The simplest kind of singularity occurs when two point particles, accelerated by their mutual gravitation, collide, so that the denominator in the inverse-square law blows up. In such cases, there will be a finite time t^* such that no solution of the classical equations up to time t^* is extendable to times later than t^*, which is to say that the laws of classical physics do not determine what happens after t^*. This problem can perhaps be dealt with by saying that, really, point particles are just an idealization, useful only because they approximate the behavior of extended bodies, whose centers of mass will never actually coincide. But this move makes it necessary to provide a theory of what happens during particle collisions, and the problem of determinism for triple-body elastic collisions mentioned above looms large.

Furthermore, it is known that there exist solutions to the classical dynamical equations with noncollision singularities. (For details and references to the physics literature, see Earman 1986, 33–39.) Such noncollision singularities entail the physical possibility of a quite bizarre species of counterexample to determinism. What typically happens in the case of such a noncollision singularity is that one or more bodies will be accelerated to an indefinitely large velocity in a finite period of time; the form of its trajectory is depicted in Figure 2a, in which a particle reaches arbitrarily high speeds before the finite time t^*. Since, again, the laws of classical mechanics are time reversible, the time reverse of this process, depicted in Figure 2b, is physically possible relative to classical mechanics. Figure 2b shows a particle suddenly "coming in from infinity," making it somewhat appropriate to call it a "space invader." Clearly, in Figure 2b, nothing about the state of the world prior to time $-t^*$ can determine whether a space invader will appear at $-t^*$ or not. So here is another striking violation of determinism in classical physics, which, unlike the Laraudogoitia and Lanford examples, does not depend on an improbable initial arrangement of an infinite number of elastic balls.

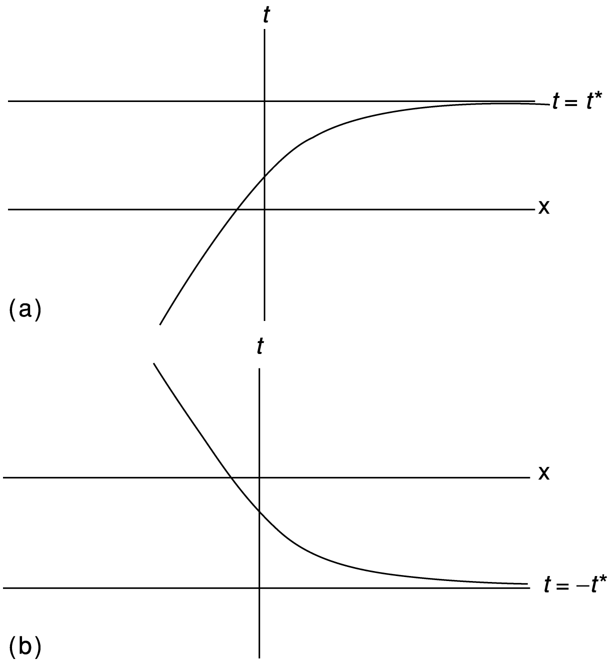


Fig. 2. (a) World-line of a "space fugitive" which may result from a non-collision singularity in classical gravitation theory. (b) A "space invader." (The time-reversal of the process depicted in Figure 2a.)

The Lanford example and the space-invader example depend on the fact that in classical physics, there is no upper bound on the speed with which an influence can propagate. This is a feature shared by classical nonrelativistic theories of fields (rather than particles), and this fact can be exploited to construct violations of determinism for classical field theories. The basic idea is that the field equations permit solutions involving a field-theoretic analog of space invaders: a wave or other disturbance in the field propagating "in from infinity" (Earman 1986, 40–45 and 48–52).

As noted above, the Laraudogoitia and Lanford examples involve violations of global conservation laws, and the space-invader example does as well, since the mass, energy, and momentum of the space invader will be added to that of the whole universe after it appears. It may be thought that this is sufficient to rule out such examples as genuine violations of determinism in classical physics, since the laws of classical physics include the global conservation laws. But there are reasons to be dubious of this move (Earman 1986, 37–39), for none of the examples just mentioned involves any violation of any *local* conservation law. (For example, the space-invader example does not violate the principle that a particle's mass must remain constant along its entire world line, which is a plausible candidate for the local principle of the conservation of mass.) Further, the global conservation laws have at best a dubious status as *fundamental* laws of classical physics. They can be derived from more fundamental laws, given certain assumptions—for example, that the universe as a whole is a closed system (which rules out space invaders) or that there are only finitely many particles (ruling out Laraudogoitia's example). One could rule out the examples of indeterminism in question by defining classical physics in such a way that it includes such assumptions. But what the examples in question show is that the fundamental laws of classical physics do not by themselves guarantee that such assumptions hold, for theses examples provide solutions of the classical equations of motion in which these assumptions are violated. To stipulate that classical physics must be understood as essentially including these assumptions (for example, that the universe as a whole is a closed system) is arguably tantamount to trying to make classical physics deterministic by fiat. Moreover, it would still not address the problem of elastic collisions involving three or more bodies.

Chaos and Unpredictability in Classical Physics

Unfortunately there is no general definition of 'chaos' that is widely agreed upon (Belot and Earman 1997, 150), but there are a variety of conditions that are generally thought to characterize chaotic systems. These conditions come in two categories: those that amount to a kind of unpredictability and those that amount to a kind of highly sensitive dependence on initial conditions. Both kinds of condition seem to be essential to the concept of chaos.

These conditions can be made precise by making use of the concept of an *abstract dynamical system*, which can be defined as an ordered triple $<X, T, \mu>$ where X is a mathematical space (which can be thought of a state space or a phase space), T is an invertible mapping of X onto itself that represents the unit time evolution of the system, and μ is a probability measure on X, which can be thought of as representing either the state of one's information about the system or a statistical measure on an ensemble of systems (Belot and Earman 1997, 151).

The notion of highly sensitive dependence on initial conditions can be made precise by means of Liapunov exponents. Liapunov exponents are quantities defined for trajectories through the state space, and they characterize the rate of exponential divergence of nearby trajectories (see Lichtenberg and Lieberman 1992, 296, for the technical details). So a given trajectory t through the state space X has positive (nonzero) Liapunov exponents just in case trajectories that start out very close to t will diverge from t exponentially. In simpler terms, if the Liapunov exponents are greater than zero, then initial conditions that are very close together will lead to later states that differ greatly, and the amount of this difference increases exponentially with time.

The unpredictability of a dynamical system can be made precise in more than one way. First, one can use the algorithmic concepts of complexity and randomness. The *conditional complexity* of finite sequence S given information I is defined as the length of the shortest program that, when fed to a universal Turing machine together with input I, will yield S as output. For an infinite sequence S, the complexity of S, $K(S)$, is defined as:

$$\lim_{n\to\infty} \frac{1}{n} K(S_n|n)$$

where S_n is the initial sequence of S having length n, and the last n in this formula should be read as the information that the sequence S_n is of length n. An infinite sequence is random if and only if its complexity is greater than zero. Roughly speaking, an infinite sequence is random just in case no matter how long an initial segment one takes, the shortest computer program that will deliver that initial segment as output is comparable in length to the segment itself. This means that there is no way to compress the information contained in the sequence. This concept of randomness can be extended to the trajectories of an abstract dynamical system by partitioning the states in X into cells and then considering the sequence of cells in which a trajectory is found at a given time, at one unit time later, at one unit time later than that, and so on. This can be thought of as a way of coding a trajectory, by using an infinite sequence of cells. A trajectory is random just in case there exists some partition of the state space into cells such that the sequence of cells that codes the trajectory is random in the sense defined above (see Belot and Earman 1997, 152, for a more rigorous exposition).

A second way to characterize unpredictability makes use of a hierarchy of statistical properties of dynamical systems. The weakest property in this hierarchy is ergodicity: The system $<X, T, \mu>$ is *ergodic* if and only if for every function f of the state space X, for every point x in X, the time average of $f(x)$ equals the average of f over the whole of X (weighted by μ). A stronger property is mixing: The system $<X, T, \mu>$ is mixing if and only if it converges to equilibrium, in the sense that for any function f defined over X, for every point x in X, $f(x)$ approaches the average value of f over X (weighted by μ) as the time gets arbitrarily large. An even stronger property is that of being a K-system. K-systems are dynamical systems with positive, nonzero metric entropy (see Lichtenberg and Lieberman 1992, 304, for details). K-systems conform to the 0–1 law, which says that a complete specification of the system's entire past history does not enable one to predict with certainty any future event except those whose probability of occurrence is 1 independently of the past history

(Batterman 1992, 60). *K*-systems thus exhibit a very strong form of unpredictability. An even stronger form is exhibited by Bernoulli systems, in which it is true at each time that any future event that does not have a probability of 1 independently of the past history is completely statistically independent of the entire past history. There are classical mechanical systems exhibiting each of the properties in this hierarchy (Lichtenberg and Lieberman 1992).

These two ways of characterizing the unpredictability of a dynamical system—by using (1) algorithmic concepts of complexity and randomness and (2) a hierarchy of statistical properties—are related by Brudno's theorem, which says that if an abstract dynamical system $<X, T, \mu>$ is ergodic, X is compact, and T is a homeomorphism, then for almost all trajectories, the complexity of the trajectory is equal to the metric entropy of the system. Hence, almost all trajectories are random if and only if the system is a *K*-system (Batterman 1992, 61; Belot and Earman 1997, 157).

Unpredictability is related to the concept of extreme sensitivity to initial conditions by Pesin's theorem, which says that under certain conditions common among classical systems, the metric entropy of an abstract dynamical system $<X, T, \mu>$ is equal to the average value of the sum of all positive Liapunov exponents (where this average is weighted by μ) (Lichtenberg and Lieberman 1992, 304; Belot and Earman 1997, 157). Thus, for systems that satisfy these conditions, having nonzero Liapunov exponents throughout a region of X of nonzero measure is equivalent to being a *K*-system; in other words, roughly speaking, having very many trajectories that diverge exponentially from very similar initial conditions is equivalent to being unpredictable in the sense in which a *K*-system is.

Mixing, being a *K*-system, and having positive Liapunov exponents have all been proposed as necessary conditions, sufficient conditions, or necessary-and-sufficient conditions for chaos (Belot and Earman 1997). Again, there is no universally accepted definition of chaos. It is sometimes assumed that unpredictability and sensitive dependence on initial conditions are so intimately related that chaos can be defined simply in terms of one or the other. For example, Joseph Ford (1989, 350) defines chaos as "a synonym for randomness" in the algorithmic sense. But as Robert Batterman (1992, 62–63) points out, behavior that is random in this sense can be generated by systems that exhibit no exponential divergence of trajectories at all (such as a spinning roulette wheel) and are therefore poor candidates for chaos. What does seem clear is that a very strong form of unpredictability is a consequence of chaos (Batterman 1992, 63). Yet, chaos is perfectly compatible with Laplacian determinism as defined above. (In fact, the preceding discussion of dynamical systems has presumed throughout that the systems in question are deterministic, in that there is a unique, invertible mapping T that represents dynamical evolution.) This shows how important it is to distinguish between the predictability and ontic senses of determinism.

Determinism in Quantum Physics

The standard formulation of quantum mechanics posits two dynamical processes (see Quantum Mechanics). The first of these is evolution of the quantum state according to the Schrödinger equation, which is linear, continuous, and deterministic. In fact, there is a clear sense in which this evolution is more deterministic than is evolution in classical mechanics, for Schrödinger evolution does not exhibit the sensitivity to small changes in initial conditions allowed in classical mechanics (Belot and Earman 1997). Schrödinger evolution is supposed to take place in any system not being observed. When an observation takes place, the second dynamical process, called "state reduction" or "collapse," kicks in. State reduction is discontinuous: The system being observed typically jumps from a state in which it is in a superposition of values of the quantity being measured to one in which it has some definite value, with the probabilities of the various possible outcomes given by the Born rule. The probabilistic nature of state reduction entails that the standard formulation of quantum mechanics is indeterministic in all of the senses discussed above.

It is generally considered problematic that the standard formulation of the theory uses 'observation' as a primitive concept and that the standard dynamics discriminates on the basis of whether an observation is taking place (cf. Albert 1992). There are a variety of contemporary approaches to dealing with this problem (the "measurement problem"), some of which seek to provide a fuller account of state reduction and some of which seek to eliminate state reduction altogether (see Quantum Measurement Problem). It remains an open question which approach is preferable, along with whether the best approach will preserve or eliminate the indeterminism of the standard formulation of the theory.

Among the interesting current options for dealing with the measurement problem are a family of interpretations called *modal interpretations*. According to modal interpretations, all evolution of the quantum state takes place according to the Schrödinger equation without state reduction, and the physical state of a system is not completely characterized by its quantum state. Some, but not all, quantities that characterize a system have determinate values at a given moment in time, and a precise rule is supplied for determining which do and which do not. The quantum state of a system determines at most probabilistic information about the values of the quantities that do have definite values (Bub 1997, 173–80). Modal interpretations are indeterministic in the predictability sense; some modal interpretations, but not all, are indeterministic in the ontic sense (235–236).

Another approach that has received a great deal of attention is David Bohm's (1952) alternative to the standard quantum theory. In Bohm's theory, the world consists of particles with definite positions at all times and wave functions that exert a nonlocal and stochastic influence on the motions of particles. Bohm's theory, too, preserves the indeterminism of the standard theory.

Hugh Everett (1957) proposed an alternative to the standard version of quantum mechanics according to which there is no state reduction and the physical quantities that characterize physical systems do not in general have determinate values, but only values relative to a given state of the rest of the universe. In this account, when an observer measures the value of some physical quantity, the observer's state typically "branches" into a number of different "relative states" such that every physically possible outcome is observed in one of these branches, and all branches are ontologically on a par. This revised version of quantum mechanics is not deterministic in the predictability sense, because it permits an observer to make only probabilistic predictions about the results of future measurements. But it satisfies Laplacian determinism, since the total state of the universe at one time determines the state at any other time. Precisely because of this, many critics have argued that Everett deprived the quantum probabilities of any intelligible meaning (Barrett 1997). What could it mean to say that the probability of getting a certain result is, say, 0.75, when every possible result is bound to be realized on some branch or other? Critics have also taken aim at the very notion of "branching observer states," arguing that it is unintelligible or at least stands in need of interpretation (e.g., Albert and Loewer 1988; Barrett 1997). Attempts to fill in the needed interpretation include assorted versions of the "many-worlds" interpretation (e.g., Albert 1992, 112–116) and the "many-minds" interpretation (e.g., Lockwood 1996).

Determinism in Relativistic Physics

Many of the failures of determinism in classical physics are due to the absence of any limit on the speed with which causal influence can propagate. In relativistic physics (excluding the possibility of tachyons), a speed limit is imposed, suggesting that relativistic physics may be more friendly to determinism than is classical physics. As will emerge, this is the case. But matters are complicated by the fact that in relativistic space-times, absolute time is not in general definable. The definitions of determinism presented in the first section referred to the complete physical state of the world at a given time. So these definitions will have to be modified before they will be applicable in relativistic physics.

To this end, it is useful to introduce some terminology (see Space-Time). A *relativistic space-time* $<M, g>$ is a four-dimensional manifold M and a metric of Lorentz signature g defined everywhere on M. In special relativity, M is topologically equivalent to $\mathbf{R}^4$, and g is the constant Minkowski metric. In general relativity, M can take any of a variety of topological structures, and g can vary from point to point. A model of general relativity is a triple $<M, g, T>$ with $<M, g>$ a space-time and T a stress-energy tensor, where g and T jointly satisfy Albert Einstein's field equations. Henceforth, it will be assumed that space-times have a temporal orientation, allowing one to distinguish light cones and timelike curves as either future directed or past directed. (There are models of general relativity in which this is not the case, such as the Gödel space-time model [Earman 1986]. If such models cannot be excluded, then there seems little hope of formulating any very interesting form of determinism satisfied by general relativity.)

An *achronal hypersurface* in a space-time $<M, g>$ is a three-dimensional surface no two points of which may be joined by a timelike curve. Let $<M, g>$ be a space-time, and let S be any achronal surface in $<M, g>$. Then *the future domain of dependence of* S, $D^+(S)$, is the set of all points p in M such that every future-directed timelike curve in M, with future endpoint p and no past endpoint, intersects S. The *past domain of dependence of* S, $D^-(S)$, is defined analogously. The *domain of dependence of* S, $D(S)$, may be defined as the union of $D^+(S)$ and $D^-(S)$ (Geroch 1977, 83–84). Figure 3 illustrates a domain of dependence in Minkowski space-time.

A family of theorems implies that a complete specification of all physical magnitudes over the S suffices to determine, up to a diffeomorphism, all such magnitudes throughout $D(S)$. (See Geroch

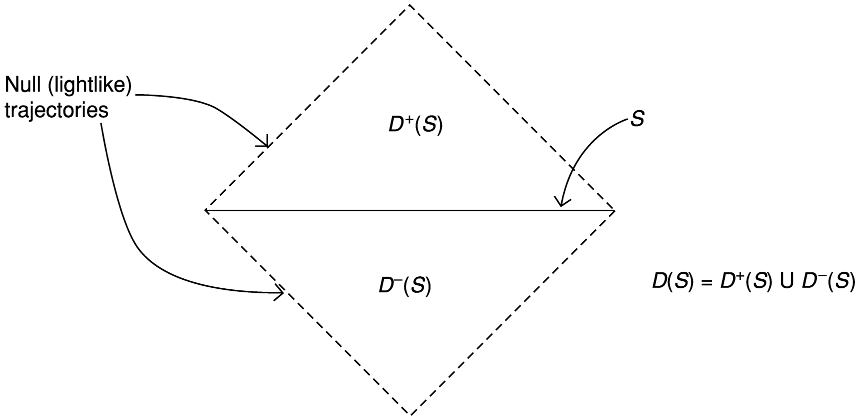


Fig. 3. Domain of dependence of a surface S in Minkowski space-time.

1977, 84; Geroch restricts his discussion to $D^+(S)$, but symmetry considerations allow these results to be extended to $D(S)$.) This shows that every model that agrees with a given model on the physical magnitudes is related to S by a diffeomorphism the restriction of which to S is the identity map. In special relativity, it is assumed that space-time has the fixed structure of Minkowski space-time, and in this setting any such diffeomorphism is just the identity map. So special relativity has the property that a complete specification of the physical magnitudes over a surface S suffices to determine the physical magnitudes over that surface's domain of dependence $D(S)$. This in itself is an interesting, albeit local, form of determinism. But special relativity also has a property that is a natural analog of Laplacian determinism. The natural special-relativistic analog of 'the complete state of the world at a given time' is the specification of all physical magnitudes over a spacelike hyperplane, which can be thought of as all of space at a given time relative to some inertial reference frame. For such a surface S, $D(S)$ is the entire space-time. So a specification of all physical magnitudes over a spacelike hypersurface suffices, in special relativity, to determine the complete physical history of the world (cf. Earman 1986, 58–60).

Things are more complicated in the case of general relativity. Assume for the moment that two models of general relativity that are related by a diffeomorphism represent the same physical situation, or the same physically possible world. Then, general relativity exhibits a weak form of determinism, in that the complete physical state over the surface S suffices to determine the physical state throughout $D(S)$.

But it is not so clear that general relativity satisfies an analog of Laplacian determinism. In general relativity, a maximally extended spacelike hyperplane (i.e., a global time-slice) need not be such that its domain of dependence is the entire space-time, in marked contrast to the situation in special relativity. A spacelike surface S in a space-time $<M, g>$ whose domain of dependence $D(S)$ *does* include all of $<M, g>$ is called a *Cauchy surface* (Earman 1986, 176–177.) It seems clear that no natural analog of Laplacian determinism can be true of a general-relativistic world that does not have a Cauchy surface, for there is no analog of the complete state of the world at a given time that suffices to determine what is going on throughout the space-time. Further, there are many models of general relativity that lack Cauchy surfaces.

One kind of example, shown in Figure 4, can be generated by starting with a space-time $<M, g>$, where M has the topology of $\mathbf{R}^4$, and deleting a compact region from it. For example, the space-time of Figure 4 has no Cauchy surface: The point p_1 is not in the domain of dependence of the surface S_1 because the timelike curve C_1 has no past endpoint and does not intersect S_1; and the point p_2 is not in the domain of dependence of the surface S_2 because of the timelike curve C_2. Clearly, the existence of a Cauchy surface requires the absence of such "holes" in the manifold. But the absence of such holes is not a sufficient condition for the existence of a Cauchy surface; there are less contrived examples of general relativistic space-times lacking Cauchy surfaces. One example is anti–de Sitter space-time; others are space-times containing singularities (though not all space-times with singularities lack Cauchy surfaces). Perhaps the general theory of relativity can be strengthened by adding conditions that guarantee the existence of Cauchy surfaces. But there are important difficulties facing this task (see Earman 1986, 177–183, for a detailed discussion).

Thus far, this section has dealt with nonpredictability senses of determinism. It turns out that relativistic physics is far more hostile to the

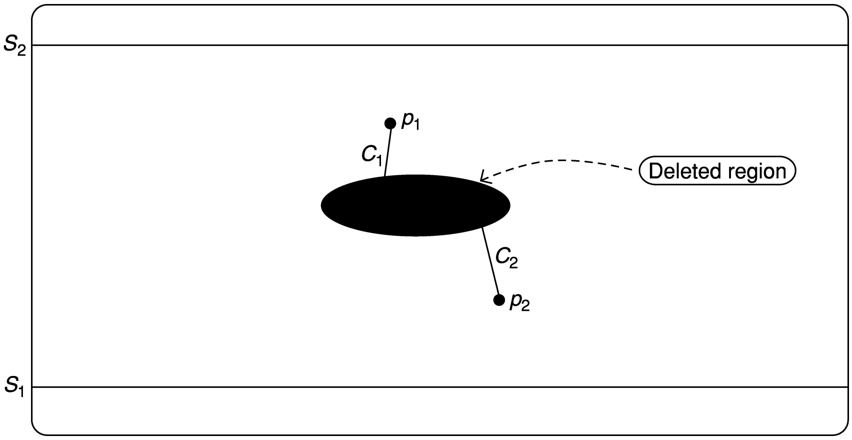


Fig. 4. A "hole-laden" general-relativistic space-time that lacks a Cauchy surface.

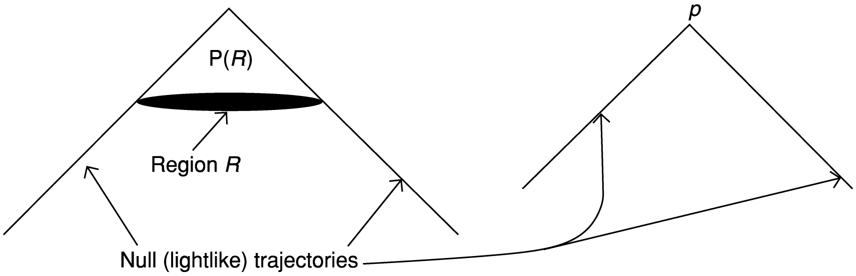


Fig. 5. Domains of prediction in special relativity (Minkowski space-time). P(*R*), the domain of prediction of the region R, is the only region to the future of *R* where the physical state is determined by the physical state in regions in *R*'s causal past; hence it is the only region where the physical state can be predicted on the basis of information in principle available to an observer occupying *R*. For a point *p*, the domain of prediction is null.

predictability conception of determinism. If it is assumed that predictions must be made on the basis of the laws together with data drawn from empirical observations, then what an observer can predict is limited to what is determined by the laws together with the physical state throughout the region of space-time contained in one's own past-directed light cone. In special relativity, this means that a pointlike observer can reliably predict nothing at all, and an extended observer can make reliable predictions concerning only a very limited spatiotemporal region (see Figure 5) (Popper 1982, 57–62; Earman 1986, 63–6). In general relativity, matters are more complicated but not much more hospitable to predictability (see Geroch 1977 and Hogarth 1993 for details).

Is It Possible to Learn Whether the World Is Deterministic on the Basis of Empirical Evidence?

The characterization of determinism as a property of a physical theory suggests that empirical evidence can show whether determinism is true, since it can show which theory or theories have a chance of being true, while analysis of a theory can show whether that theory is deterministic. But an argument of Suppes (1993) suggests that, in fact, it may be impossible to determine whether our world is deterministic.

Suppes (1993, 254) cites a theorem due to Ornstein to the effect that there exist processes that can equally well be analyzed as deterministic classical-mechanical systems or as indeterministic semi-Markov processes. Suppes goes on to argue that it is plausible that this result applies to a great many processes found in the actual world. The conclusion Suppes draws is that the issue of determinism is "transcendental," not capable of being settled by empirical research. Be that as it may, the concept of determinism continues to serve as a useful tool for probing the foundations of a variety of physical theories.

John T. Roberts

References

Albert, David (1992), *Quantum Mechanics and Experience.* Cambridge, MA: Harvard University Press.

Albert, David, and Barry Loewer (1988), "Interpreting the Many-Worlds Interpretation," *Synthese* 77: 195–213.

Barrett, Jeffrey A. (1997), "On Everett's Formulation of Quantum Mechanics," *Monist* 80: 70–96.

Batterman, Robert W. (1992), "Defining Chaos," *Philosophy of Science* 60: 43–66.

Belnap, Nuel D. (2001), *Facing the Future: Agents and Choices in Our Indeterminist World.* Oxford: Oxford University Press.

Belot, Gordon, and John Earman (1997), "Chaos Out of Order: Quantum Mechanics, the Correspondence Principle and Chaos," *Studies in History and Philosophy of Modern Physics* 28: 147–182.

Bohm, David (1952), "A Suggested Interpretation of Quantum Theory in Terms of 'Hidden Variables,'" *Physical Review* 85: 166–193.

Brandon, Robert, and Scott Carson (1996), "The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variables Proof' but No Room for Determinism Either," *Philosophy of Science* 63: 315–337.

Bub, Jeffrey (1997), *Interpreting the Quantum World.* Cambridge: Cambridge University Press.

Earman, John (1986), *A Primer on Determinism.* Dordrecht, Netherlands: Reidel.

Everett, Hugh M. (1957), "'Relative-State' Formulation of Quantum Mechanics," *Reviews of Modern Physics* 29: 454–462.

Ford, Joseph (1989), "What Is Chaos, That We Should Be Mindful of It?." in Paul Davies (ed.), *The New Physics.* Cambridge: Cambridge University Press, 348–371.

Geroch, Robert (1977), "Prediction in General Relativity," in John Earman, Clark Glymour, and John Stachel (eds.), *Foundations of Space-Time Theories: Minnesota Studies in the Philosophy of Science*, vol. 8. Minneapolis: University of Minnesota Press, 81–93.

Graves, Leslie, Barbara L. Horan, and Alexander Rosenberg (1999), "Is Indeterminism the Source of the Statistical Character of Evolutionary Theory?" *Philosophy of Science* 66: 140–157.

Hogarth, Mark (1993), "Predicting the Future in Relativistic Spacetimes," *Studies in History and Philosophy of Science* 24: 721–739.

Kitcher, Philip (2001), "Battling the Undead: How (and How Not) to Resist Genetic Determinism," in Rama S. Singh, Costas B. Krimbas, Diane B. Paul, and John Beatty (eds.), *Thinking About Evolution: Historical, Philosophical, and Political Perspectives.* Cambridge: Cambridge University Press, 396–414.

Lanford, Oscar E. (1975), "Time Evolution of Large Classical Systems," in J. Moser (ed.), *Dynamical Systems, Theory and Applications.* New York: Springer-Verlag, 1–111.

Laplace, Pierre Simon ([1814] 1951), *A Philosophical Essay on Probabilities.* Translated by Frederick Wilson Truscott and Frederick Lincoln Emory. New York: Dover Publications, 1951. Originally published as *Essai Philosophique sur les Probabilités* (Paris: Courcier).

Laraudogoitia, Jon Perez (1996), "A Beautiful Supertask," *Mind* 105: 81–83.

Lichtenberg, A. J., and M. A. Lieberman (1992), *Regular and Chaotic Dynamics*, 2nd ed. New York: Springer-Verlag.

Lockwood, Michael (1996), "'Many Minds' Interpretations of Quantum Mechanics," *British Journal for the Philosophy of Science* 47: 159–188.

Montague, Richard (1974), "Deterministic Theories," in Richmond H. Thomason (ed.), *Formal Philosophy.* New Haven, CT: Yale University Press, 303–359.

Nagel, Ernest (1953), "The Causal Character of Modern Physical Theory," in Herbert Feigl and May Brodbeck (eds.), *Readings in the Philosophy of Science.* New York: Appleton-Century-Crofts, 419–437.

O'Connor, Timothy (ed.) (1995), *Agents, Causes, and Events: Essays on Indeterminism and Free Will.* Oxford: Oxford University Press.

Popper, Karl R. (1982), *The Open Universe: An Argument for Indeterminism.* Totowa, NJ: Rowman and Littlefield.

Sarkar, Sahotra (1998), *Genetics and Reductionism.* Cambridge: Cambridge University Press.

Suppes, Patrick (1993), "The Transcendental Character of Determinism," in P. A. French, T. E. Uehling, and H. K. Wettstein (eds.), *Midwest Studies in Philosophy Volume 18.* Notre Dame, Indiana: University of Notre Dame Press, 242–257.

See also **Prediction; Quantum Mechanics; Space-Time**

DUHEM THESIS

The Duhem thesis holds that scientific hypotheses are not tested against experimental data in isolation but as part of a larger body of beliefs. This holistic epistemological doctrine was first put forward by Pierre Duhem in his *Aim and Structure of Physical Theory* in 1906 (Duhem [1906] 1954). It runs counter to the view that the rational acceptability of any scientific hypothesis can be unambiguously determined by empirical data. In particular it challenges the possibility that a hypothesis can be conclusively falsified by data. Karl Popper ([1935] 1951) subsequently held such a possibility as distinctive of the demarcation criterion of scientific belief.

Duhem developed his position through a critique of a staple of traditional doctrine of scientific method: the Baconian and Newtonian idea of the crucial experiment, or *experimentum crucis*. In order to derive a prediction with which a hypothesis or theory may be tested, the prediction must be itself testable; but the latter is possible only if we introduce into the derivation assumptions about the functioning of the experimental apparatus or measurement instrument. For example, one way in which the phenomenon of superconductivity was established experimentally was by deriving the mathematical expression of physical effects in terms of values of the magnetic field in and around a superconducting crystal—the Meissner effect. Changes in the distribution of such values could be measured and mapped out with a galvanometer, which was assumed to convert magnetic variations into measurable variations in an electric current according to Ampere's law. Assumptions of this sort are typically known as auxiliary hypotheses.

The holistic thesis led Duhem to a form of conventionalism in scientific methodology that required 'good sense' on the part of the scientist in the choice of hypotheses or theories and an act of faith in the belief in their truth. This kind of conventionalism is different from the one advocated by Henri Poincaré. The latter emphasizes the different possible definitions of geometrical and mechanical systems compatible with our experience of the world. They are neither a priori nor empirical; instead, they are conventions, in this case convenient preconditions of experimental physics, and hence not amenable to empirical testing. Poincaré's conventionalism differs from Duhem's holism. The Duhem thesis entails a form of underdetermination of theory by data that differs from the thesis that indefinitely many alternative hypotheses are compatible and can be deductively connected to a given body of empirical data. It entails that for any successful theoretical hypothesis, all of its rivals—that is, any falsified, or incompatible, hypothesis—can be made to fit the data with a suitable modification in the auxiliary hypotheses.

In the subsequent three decades, the Duhem thesis played an important role in the formulation of views associated with logical empiricism and the members of the Vienna Circle. Thus, Moritz Schlick argued in 1915 that geometry was nonempirical, since the non-Euclidean geometry used in general relativity is justified as part of the simplest total system of natural laws consistent with the empirical data, so that Euclidean geometry remains a genuine possibility as part of a different, more complicated total system of natural laws (Schlick [1915] 1979).

For Otto Neurath, in the same decade, the thesis legitimized the possibility of retaining a hypothesis in the face of conflict with data. It suggested the possibility of imagining an indefinite number of theoretical possibilities; it opened the door for pragmatic considerations, especially in the world of natural and social phenomena; and it motivated the methodological desirability of a unification of the sciences, allowing for exact predictions of the behavior of complex phenomena involving factors studied by different sciences. Neurath's holism was more radical than Duhem had envisioned. In the early 1930s Neurath stressed that within the evolving complexes of beliefs that make up science and culture, the distinction between analytic and synthetic statements, just like the validity of logical principles, would be a contingent historical matter (Neurath 1983). Similarly, Rudolf Carnap (1937) argued, in *The Logical Syntax of Language*, that the Duhem thesis was compatible with a distinction between meaning-constitutive (analytic) and factual (synthetic) statements, and defended further the more general type of holism that allowed for the revisability of both types of statements. This appreciation led to a defense of a spirit of tolerance—Carnap's *principle of tolerance*—and of pragmatic considerations that included logic and mathematics, as well as the choice of linguistic frameworks (Carnap 1937).

In the 1950s, Willard Van Quine pointed to an alleged failure of Carnap's attempts to articulate an account of analyticity and revived Duhem's arguments as part of his rejection of the distinction between analytic and synthetic propositions. Quinean, more radical holism entails equality of status of empirical (physical) and theoretical (logical or mathematical) statements. Moreover, Quine's additional criticism of accounts of analyticity prompted his epistemological naturalism, shifting the emphasis to synthetic empirical statements originating in sensory stimuli at the periphery of our "web of beliefs." In the holistic web of science, Quine distinguished beliefs only by their different degrees of centrality and entrenchment, a naturalistic counterpart to the more traditional reference to degrees of certainty.

This view has been criticized most recently by Friedman (2001), who draws attention instead to different functions or roles of beliefs within the evolving and unpredictable whole of science. One such role is played by constitutive principles, as relativized a priori in the post-Kantian tradition of Reichenbach's and Kuhn's ideas. Such principles made other beliefs within the constituted framework empirically testable. For instance, Sir Isaac Newton's universal law of gravitation gets empirical sense and application from his laws of motion linking forces to mass and acceleration and defining an inertial frame of reference. A similar idea appears in Poincaré's formulation of conventionalism.

Another sort of assumption involved in the application and testing of hypotheses concerns the complete description of the situation in terms of the absence of interfering factors or disturbing influences, known as provisos (following Hempel), *ceteris paribus* ("other things being equal") clauses (following John Stuart Mill), or more radically, *ceteris absentibus* clauses ("other things being absent"). The assumption that *ceteris paribus* clauses cannot be dismissed raised a debate about the contents and conditions of applicability of natural laws. For Cartwright (1989), their presence has important philosophical implications. She argues that the causal nature of the contents of *ceteris paribus* conditions renders the reduction of causal laws to statements of regularities about manifest behavior impossible. They imply also that in cases of composition, laws about components cannot satisfy Hempel's criteria of scientific explanation; instead, they can only literally describe either fictions (i.e., models) or causal capacities, but not manifest behavior. Their descriptive contents have been defended also in terms of counterfactual statements about the isolation of systems, as in the description of dispositions (Suppe 2000) or in terms of additional explanatory commitments subject to empirical testability (Pietroski and Rey).

In the 1980s Hempel (1988) raised the so-called problem of provisos: The acceptable determinate formulation of a law in a theory as well as the derivation of empirical consequences require the explicit statement of an indefinitely large number of relevant provisos, but this requirement is impossible to satisfy. The problem can be solved by assuming that a theory need not ground its isolation conditions. Such grounding can be either a pragmatic and contextual question (Lange 1993) or else a question of independent causal knowledge about experimental settings (Suppe 2000).

JORDI CAT

References

Carnap, R. (1937), *The Logical Syntax of Language*. London: Kegan, Paul, Trechner, Teuben & Co.

Cartwright, N. (1989), *Nature's Capacities and Their Measurement*. Oxford: Oxford University Press.

Duhem, P. ([1906] 1954), *The Aim and Structure of Physical Theory*. Princeton, NJ: Princeton University Press.

Friedman, M. (2001), *The Dynamics of Reason: The 1999 Kant Lectures at Stanford University*. Stanford, CA: CSLI Publications.

Hempel, C. G. (1988), "Provisoes: A Problem Concerning the Inferential Function of Scientific Theories," *Erkenntnis* 28: 147–164.

Lange, M. (1993) "Natural Laws and the Problem of Provisos," *Erkenntnis* 38: 233–248.

Neurath. O. (1983), *Philosophical Papers, 1913–1946*. Dordrecht, Netherlands: Reidel.

Popper, K. R.([1935] 1951), *Logic of Scientific Discovery*. London: Unwin.

Schlick, M. ([1915] 1979), "The Philosophical Significance of the Principle of Relativity," in *Philosophical Papers*, vol. 1: *1909–1922*. Dordrecht, Netherlands: Reidel, 153–189.

Suppe, F. (2000), "Hempel and the Problem of Provisos," in J. Fetzer (ed.), *Science, Explanation and Rationality: The Philosophy of C. G. Hempel*. Oxford: Oxford University Press, 187–213.

See also **Neurath, Otto; Poincaré, Henri; Popper, Karl; Quine, Willard Van; Underdetermination of Theories**

DUTCH BOOK ARGUMENT

The Dutch Book argument was first presented by Frank Ramsey in his 1926 paper "Truth and Probability" (Ramsey 1926) (see Ramsey, Frank Plumpton). The argument purports to show that an agent's degrees of belief, or degrees of confidence, should satisfy the Kolmogorov axioms of probability (termed coherence) (see Probability). It is often cited by Bayesians, who take degrees of belief to be probabilities and endorse a probabilistic approach to theory confirmation (see Bayesianism; Confirmation Theory). If $P(H)$ represents the probability assigned to the statement (or sentence or proposition) H, then the axioms require that:

1. $0 \leq P(H) \leq 1$;
2. if H is a tautology, then $P(H) = 1$; and
3. if H_1 and H_2 are mutually exclusive, then $P(H_1 \vee H_2) = P(H_1) + P(H_2)$.

The Dutch Book argument (DBA) assumes that an agent's degrees of belief are linked to a set of betting quotients, so that if an agent's degrees of belief are incoherent (i.e., fail to satisfy the axioms), then the agent will possess an incoherent set of betting quotients. The argument then appeals to a mathematical theorem, the Dutch Book theorem, which states that if a set of betting quotients fails to satisfy the axioms of probability, there will be a series of bets, each of which is individually fair according to that set of betting quotients but which taken together will produce a net loss (a "Dutch Book"). The fact that incoherent degrees of belief are linked to a sure loss is taken as reflecting a defect in those beliefs. The mechanics of making a Dutch Book against an agent whose betting quotients fail to satisfy the axioms will be considered first, and will be followed by a discussion of what the possibility of constructing such a series of bets reveals about a person's beliefs.

To convey the content of the Dutch Book theorem, it will be shown how a bookie can exploit a bettor whose betting quotients violate the probability axioms. With de Finetti (1937), it is here assumed that a bet on a proposition H is an arrangement that has the following canonical form:

H	Payoff
T	$S - qS$
F	$-qS$

If H is true, the bettor on H collects the amount $S - qS$, but if H is false the bettor loses qS. The quantity S is called the stake and q is the betting quotient. S is the amount won if H is true and qS is the cost of the bet. Here it is assumed that an agent will bet either for or against H, provided that the betting quotient $Q(H)$ equals q. These are presumed to be fair bets by the agent's lights, or as having an expected value of zero. It can now be shown that if the agent's betting quotients fail to satisfy the axioms, a Dutch Book can be made against the agent by a clever bookie, provided that the agent will take either side of a bet for which $Q(H)$ equals q. For simplicity, the stakes for each bet will be set at \$1.

- **Axiom 1:** Suppose that $Q(H) < 0$. In this case, the bookie buys the bet that pays \$1 if H is true and 0 otherwise, for the negative price $Q(H)$. This is equivalent to betting against H for the agent and so the payoff table is as follows:

H	Payoff
T	$-[1 - Q(H)]$
F	$Q(H)$

  Since $Q(H)$ is negative, the agent will suffer a net loss regardless of the truth value of H. Suppose on the other hand that $Q(H) > 1$. In this case the bookie sells the bet that pays \$1 if H is true and 0 otherwise, for $Q(H)$. Of course, this means that the agent will pay more for the bet than it is possible to gain and so end up with a net loss to the bookie.
- **Axiom 2:** Suppose that an agent's betting quotient for a tautology H is not equal to 1. The case where $Q(H) > 1$ was included above, so assume that $Q(H) < 1$. Here the bookie will buy the bet in which the agent pays the bookie \$1 if H is true, and nothing if H is false, for $Q(H)$. The payoff table for the agent will be:

H	Payoff
T	$-[1 - Q(H)]$
F	$Q(H)$

  Notice that the agent is bound to lose the amount $[1 - Q(H)]$, since H is a tautology and hence must be true.
- **Axiom 3** (additivity): Assume that H_1 and H_2 are mutually exclusive and that $Q(H_1 \vee H_2) \neq Q(H_1) + Q(H_2)$. There are two cases,

  a) $Q(H_1 \vee H_2) > Q(H_1) + Q(H_2)$
  and
  b) $Q(H_1 \vee H_2) < Q(H_1) + Q(H_2)$.

  Suppose that $Q(H_1 \vee H_2) < Q(H_1) + Q(H_2)$, then the bookie will offer the agent the bet that pays \$1 if H_1 and 0 otherwise for $Q(H_1)$ and the bet which pays \$1 if H_2 is true and 0 otherwise for $Q(H_2)$. The bookie then buys the bet which will lead to a gain of \$1, if $(H_1 \vee H_2)$ is true and 0 otherwise, for the price of $Q(H_1 \vee H_2)$.

The possible payoffs to the agent are summed up in the following table:

H_1	H_2	Net Payoff
T	F	$[1 - \mathrm{Q}(H_1) - \mathrm{Q}(H_2) + \mathrm{Q}(H_1 \vee H_2) - 1]$
F	T	$[1 - \mathrm{Q}(H_1) - \mathrm{Q}(H_2) + \mathrm{Q}(H_1 \vee H_2) - 1]$
F	F	$[-\mathrm{Q}(H_1) - \mathrm{Q}(H_2) + \mathrm{Q}(H_1 \vee H_2)]$

Since $\mathrm{Q}(H_1 \vee H_2) < \mathrm{Q}(H_1) + \mathrm{Q}(H_2)$, the agent is assured of a loss in each case. If $\mathrm{Q}(H_1 \vee H_2) > \mathrm{Q}(H_1) + \mathrm{Q}(H_2)$, then the bookie simply reverses the direction of the bets.

It has been demonstrated above how an incoherent set of betting quotients can be exploited to produce a sure loss. It can also be shown that the bookie is not guaranteed a net profit if the agent's betting quotients are coherent. It is now time to examine what such betting quotients show about an agent's degrees of belief. Ramsey understood a person's degree of belief in a proposition as reflecting the person's willingness to act on it, and maintained that betting quotients are at least an approximate measure of a person's degrees of belief. He argued from what is in effect the Dutch Book theorem that

> If anyone's mental condition violated (the laws of probability), his choice would depend on the precise form in which the options were offered him, which would be absurd. He could have a book made against him by a cunning bettor and would then stand to lose in any event. (Ramsey 1926, 80)

This has been interpreted as the claim that incoherence is irrational because it leaves the agent vulnerable to bad consequences, and so is a kind of pragmatic defect. If the argument is understood in this way, it is open to serious objections. The main problem is that the link between having incoherent degrees of belief and suffering a loss is a weak one. First an agent's degrees of belief need not correspond to the bets that the agent will consider fair, or be willing to take. A person might be highly confident in a proposition, yet be unwilling to bet at the corresponding odds, because of risk aversion or a view of gambling as inappropriate. Even if the agent's degrees of belief and betting quotients match for individual bets, the agent is not compelled to regard the corresponding bets as jointly fair, as is needed to show that violation of the additive law involves Dutch Book vulnerability. Putting aside objections that an incoherent agent need not regard bets involved in producing a Dutch Book as fair given corresponding degrees of belief, the connection between the existence of such bets and suffering a bad outcome is rather tenuous. Whether an incoherent agent suffers an actual loss depends on whether the necessary bets will actually be made and collected. Furthermore, the agent could simply refuse to bet, and thus avoid any potential loss. Finally, the assumption that it is irrational to put oneself in a situation that could lead to a sure loss seems simply wrong, for putting oneself in such a situation could be the best available (if not a terribly desirable) option.

It has been argued that the DBA is misunderstood if it is thought to force compliance with the probability axioms as a means of avoiding bad outcomes. The suggestion is that it is not the threat of a sure loss that is the problem, but rather that having choice guide beliefs that are tied to a sure loss signals an inconsistency in those beliefs. Indeed, this interpretation fits well with Ramsey's claim that "any definite set of degrees of belief which broke (the probability axioms) would be inconsistent in the sense that it violated the laws of preference between options" (Ramsey 1926, 80). In this reading, having degrees of belief can be reduced to having certain preferences, with incoherence then being inconsistency of preference. Moreover, given appropriate constraints on a set of preferences, a utility function can be defined relative to which the value of bets is additive. It can thus be argued that by appealing to the theory of preference, a crucial assumption of the DBA can be defended. Further, within the theory of preference it is possible to define both utility and probability functions such that an agent's preferences can be represented as maximizing expected utility relative to those utility and probability functions. Such representation theorems yield a direct argument for probabilism by showing that given rational preferences, an agent can be interpreted as having degrees of belief that satisfy the probability axioms. There is controversy over the attempt to justify the DBA by appealing to principles of utility theory. Some view it as at best irrelevant, given the representation theorems. Others raise questions about representation arguments and see the DBA as providing important motivation for the Bayesian constraints on rational belief (see discussion, see Armendt 1993).

Some philosophers have objected to Ramsey's idea that incoherence reduces to inconsistency of preference, as well as to the idea that the force of the DBA derives from fundamental assumptions about preference and decision. Several attempts have been made to "depragmatize" the DBA and to show clearly that incoherence involves a type of inconsistency that is essentially epistemic rather than pragmatic in nature (Howson and Urbach 1993; Christensen 1996). Instead of reducing degrees of belief to preferences, they are reduced to evaluations of the fairness of bets, or are understood as justifying certain bets as fair. In each case, the Dutch Book theorem is invoked to establish that incoherence involves believing bets to be fair that cannot be, or as justifying bets as fair that cannot be fair. Here incoherence bears a clearer resemblance to inconsistency for full belief than on Ramsey's preference interpretation, but it is doubtful that the argument can be made to work without the resources of decision theory.

The attempts to depragmatize the DBA have foundered in providing a noncircular, and genuinely nonpragmatic, account of fairness according to which violation of the probability axioms *always* involves the appropriate sort of unfairness (see Maher 1997). Further, it is just as implausible that degrees of belief reduce to judgments of fairness, or that they alone justify certain beliefs as fair, as is the claim that degrees of belief can be reduced to preferences. Still the underlying idea that incoherence is an epistemic, and not essentially a pragmatic, defect is surely correct. It is the notion of valuation, together with the underlying logic of propositions, that yields the Dutch Book theorem and suggests that coherence is, at least, an epistemic ideal.

Dutch Book arguments have also been devised in support of the principles of conditionalization, Jeffrey conditionalization, and reflection. (For discussion, see Teller 1973; Armendt 1980; and van Fraassen 1995.)

SUSAN VINEBERG

References

Armendt, B. (1980), "Is There a Dutch Book Argument for Probability Kinematics?" *Philosophy of Science* 47: 583–588.

——— (1993), "Dutch Books, Additivity and Utility Theory," *Philosophical Topics* 21: 1–20.

Christensen, D. (1996), "Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers," *Journal of Philosophy* 93: 450–479.

Christensen, D. (2001), "Preference-Based Arguments for Probabilism," *Philosophy of Science* 68: 356–376.

de Finetti, B. (1937), "Foresight: Its Logical Laws, Its Subjective Sources," in H. E. Kyburg and H. E. K. Smokler (eds.), *Studies in Subjective Probability*. Huntington, NY: Robert E. Kreiger Publishing Co.

Howson, C., and P. Urbach (1993), *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.

Kaplan, M. (1996), *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.

Kennedy, R., and C. Chihara (1979), "The Dutch Book Argument: Its Logical Flaws, Its Subjective Sources," *Philosophical Studies* 36: 19–33.

Maher, P. (1993), *Betting on Theories*. Cambridge: Cambridge University Press.

——— (1997), "Depragmatized Dutch Book Arguments," *Philosophy of Science* 64: 291–305.

Ramsey, P. F. (1926), "Truth and Probability," in H. E. Kyburg and H. E. K. Smokler (eds.), *Studies in Subjective Probability*. Huntington, NY: Robert E. Kreiger Publishing Co.

Skyrms, B. (1975), *Choice and Chance*. Belmont, CA: Wadsworth.

——— (1987), in N. Rescher (ed.), Coherence. Scientific Inquiry in Philosophical Perspective. Pittsburgh: University of Pittsburgh Press, 225–242.

Teller, P. (1973), "Conditionalization and Observation," *Synthese* 26: 218–258.

van Fraassen, B. (1995). "Belief and the Problem of Ulysses and the Sirens," *Philosophical Studies* 77: 7–37.

Vineberg, S. (2001), "The Notion of Consistency for Partial Belief," *Philosophical Studies* 102: 281–296.

See also **Bayesianism; Confirmation Theory; Decision Theory; Probability**

E

ECOLOGY

Ecology is composed of a remarkably diverse set of scientific disciplines, and many different subfields can be distinguished, such as physiological, behavioral, evolutionary, population, community, ecosystem, and landscape ecology. Clearly, no summary will do them all justice. However, for the present context, ecology as a science can be divided into three basic areas—population, community, and ecosystem ecology. This article will introduce some of the fundamental philosophical issues raised by these three disciplines.

The first order of business is to ask, What is the science of ecology? and more importantly, What is it *not?* (see Brennan 1988). Sometimes the term "ecology" is treated as synonymous or coextensive with three different concepts or sets of concepts:

- *The science of ecology*: the study of organisms, their groups, and their relation to their environment.
- *Environmentalism*: a set of sociopolitical views about the right relationship between humans and nature.
- *The ecology of an organism, population, or community*: in the case of organisms, roughly the life history of that organism.

This essay will focus only on ecology in the first of these senses—as a set of scientific disciplines. It should be noted, however, that there are important questions about how the science of ecology is related to environmental ethics and public policy (see Ludwig, Mangel, and Haddad 2001).

Metaphysics and Ecological Communities

One of the standard topics of ecology is succession. Succession concerns the structural and compositional changes that occur in communities and ecosystems as populations and species replace each other. Traditionally, succession is broken into three stages. The first is the *pioneer* stage, when the first colonizers arrive in an area; each subsequent stage is called a sere; up to the final, relatively stable stage, called the *climax*. Succession is either *primary* or *secondary*. Primary succession involves the colonization of bare ground where no ecosystem has been present. Examples of areas where primary succession occurs are sand dunes, volcanic flows, mud flats, and glacial tills. Secondary succession involves the replacement of communities after some disturbance that may involve abandoned fields, wind-blown gaps in forests, or wildfires. An example of temperate terrestrial secondary succession is the sequence of annual weeds, perennial weeds, shrubs, young pine forest, and oak forest with a well-defined canopy.

One of the foundational controversies in community ecology arose between Frederic Clements (1916) and Henry Gleason (1917) concerning succession and the nature of communities. Clements argued that communities follow a very specific sequence of stages that can be characterized in terms of nutrient cycling, species diversity, and biomass. He claimed that there is a single climax community that is self-perpetuating and tightly integrated as the result of biotic interactions among species. Clements considered communities to be "superorganisms":

> The developmental study of vegetation necessarily rests upon the assumption that the unit or climax formation is an organic entity. As an organism the formation arises, grows, matures, and dies. . . . The life-history of a formation is a complex but definite process, comparable in its chief features with the life-history of an individual plant. (1916, 16).

Gleason considered Clements' views to be without empirical support, and argued that succession results from individual species' physiological requirements and local climatic conditions:

> [I]t may be said that every species of plant is a law unto itself, the distribution of which in space depends upon its individual peculiarities of migration and environmental requirements. (1917, 26)

Hence, Gleason's views were oriented more to the individual. Likewise, he did not think that there was a final climax community, but rather communities were continually changing and nonequilibrial. These two approaches to succession and communities continue to be of influence in ecology (see Levins and Lewontin 1985; Simberloff 1980).

Nonetheless, one is still left wondering what communities are. Ecologists have conceived of communities in roughly three different ways (Shrader-Frechette and McCoy 1993, 11–31):

- Communities are groups of species at particular places and times and nothing more.
- Communities are functionally interrelated groups of species.
- Communities are groups of species that are organism-like entities.

Biologists grant that an ecological community is minimally a set of species. However, what else, if anything, is required? As Kristin Shrader-Frechette and Earl McCoy (1993, 13) ask: "Envision a group of species occurring in the same place at the same time. Conceptually, what attributes might be used to link these species together, such that they could be distinguished from other similar groups?" Better sense can be made of the three different concepts of communities by considering some metaphysics.

All objects (except possibly the very simplest) are composed of parts. Those parts may or may not be related to each other. Objects can be classified by the relations that exist between their parts, as either aggregates, wholes, or individuals. If an object is an *aggregate*, its parts bear little or no causal relations to one another. Thus, aggregates are not causally integrated at a time or over time. If an object is a *whole*, then certain causal relations exist among its parts. Wholes exist as causally structured entities that are minimally integrated at a time and through time. Finally, an *individual* is an object whose parts bear causal relations to one another such that the object is highly structured and integrated. The differences between aggregates, wholes, and individuals concern the causal relations amongst their parts and the strength of those relations. These objects form a continuum, and the differences between them are of degree. Communities can exist as aggregates, wholes, or individuals.

Now consider the sort of community that Gleason (1926) had in mind:

> Are we not justified in coming to the general conclusion, far removed from the prevailing opinion, that an association [i.e., community] is not an organism, scarcely even a vegetation unit, but merely a *coincidence*? (16)

Communities, according to Gleason, are composed of whatever species coexist in space and time. This is a Gleasonian community: a group of species in a particular area at a particular time. In effect, this type of community consists of aggregates—objects whose parts bear few if any causal relations to one another.

Now consider those communities that exist as wholes. Here there is a set of species that exists as a structured entity—there are causal relations that at least weakly integrate the species at a time and through time. This type of community concept is sometimes associated with George Evelyn Hutchinson. Hutchinson thought of communities as having "feedback loops" that assure their self-regulation and persistence.

What sorts of causal relations or feedback loops might bind species in a community? One possibility is that various interspecific interactions exist amongst organisms and populations. Between any two species, these interactions can be classified as either positive (+), negative (–), or nonexistent (0), depending on how they affect the growth or abundance of the respective species. These relations include competition [–,–], predator-prey [–,+], mutualisms [+,+], amensalisms [–,0], and commensalism [0,+]. Likewise, some interactions take place between more than two species. These *indirect effects* occur when a donor species' influence is transmitted through at least one transmitter species to a receiver species. Finally, pairwise interactions themselves may be *nonadditive* if the interaction between the pair changes as the number of species in the community changes. If there are interspecific interactions between species that integrate the species into something more than an aggregate—a whole—then this community will be called a Hutchinsonian community: a group of species that at least weakly interact with one another and no others. The community exists as a group of species structured by various interspecific causal relations. One can also see why some ecologists are skeptical of the existence of plant communities and animal communities, since they leave out causally salient parts.

Finally, a community may be a tightly integrated group of species that bear various causal relations among their component species. The community forms an individual, as if it were a multicellular organism. This is a Clementsian community: a group of species that strongly interact with one another.

It is an empirical issue whether any of these community concepts apply to any group of species. Nonetheless, some progress has been made in understanding what ecological communities *might* be. Next, several arguments will be considered for why one might think ecological communities do not exist. Here is one such argument. Communities are real only if they have distinct boundaries. However, many purported communities do not have distinct boundaries. Hence, many purported communities are not real (see Simberloff 1980, 16–17; Levins and Lewontin).

There are several general points to be made about this argument. By all accounts, a community consists of a group of species. Moreover, the community exists wherever those species exist. Thus, its boundary consists of its outermost species. So, though it may be difficult, a community's boundary can be determined by its species' boundaries. However, putatively different communities blend continuously into one another unless there is some ecotone—a relatively discrete zone of transition (Figure 1). If they blend continuously, then it is not clear where the communities begin and end.

This may be an epistemological problem for ecologists, but from a metaphysical point of view, it need not be. For example, if two Hutchinsonian or Clementsian communities share a common habitat, they still are distinct in virtue of the causal interactions between their respective species. As Richard Levins and Richard Lewontin write, "The question of boundaries of communities is really secondary to the issues of interaction among species" (1985, 138). Hence, the problem of continuous overlap need not be a particular problem for the Hutchinsonian and Clementsian approaches.

There does seem to be reason for denying the existence of Gleasonian communities. Recall that a Gleasonian community is a set of species in a particular area at a particular time. Suppose there is a group of *n* species at a particular place at a particular time. If the group is a Gleasonian community, then it can properly be asked why some other ($n + 1$)-th species is *not* a member of the community. In one of the other approaches, the answer would be given by the causal interactions of the *n* species. The ($n + 1$)-th species would be excluded from such interactions. However, in the Gleasonian approach, it appears that the membership of the community is not secured by mind-independent

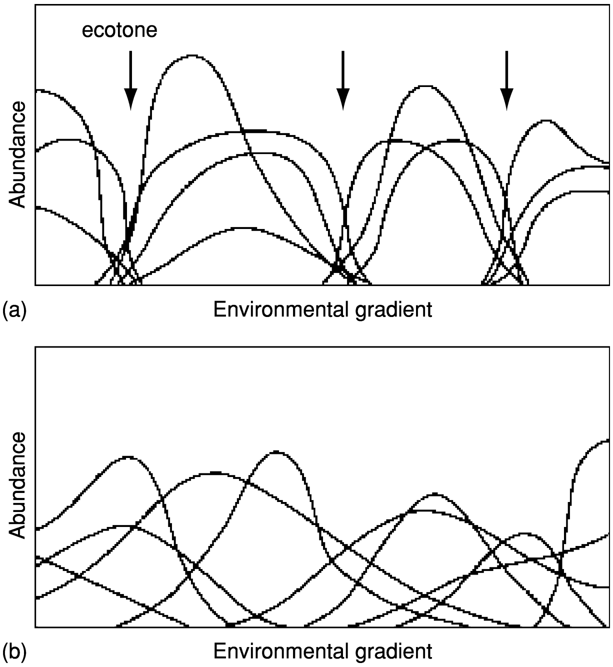


Fig. 1. (a) Ecotones generating discrete boundaries between groups of species.
(b) No ecotones generating discrete boundaries between species.
[*From Ricklefs and Miller 2000, 524*]

causal interactions but rather by the ecologist's choice about spatial and temporal boundaries. However, if Gleasonian communities objectively exist, they must exist independently of mind. The communities depend on ecologists' decisions—arbitrary or not—as to what species to consider members of them. Hence, they do not objectively exist. This in effect was the view of the ecologist Robert MacArthur (1962):

> Irrespective of how other ecologists use the term "community"—and there are almost as many uses as there are ecologists—I use it here to mean any set of organisms currently living near each other and about which it is interesting to talk. (189–190)

Likewise, something is *natural* only if it does not depend on human activities. Hence, even if Gleasonian communities exist, they would be nonnatural in this sense. Hence, they either do not objectively exist or are nonnatural.

This discussion has considered the nature of ecological communities and has only skimmed some of the issues. There are, however, many conceptual and metaphysical problems concerning ecological entities. *Token* ecological communities may exist—however, what about *types* of communities, or biomes, such as temperate grassland, chaparral, savanna, deserts, taiga, and tropical rain forests? Much of early community ecology consisted of classifying communities, and traditional accounts of succession seem to depend on such classifications. Do other ecological entities like ecosystems or guilds exist? If ecosystems exist, do they possess fashionable properties like *health* and *integrity*?

A Balance of Nature?

One might consider it "folk wisdom" that flora and fauna exhibit a "balance of nature" (Egerton 1973; Pimm 1993). Ecologists have often thought that the more diverse or complex a community or ecosystem is, the more stable it would be. This section will consider the diversity/complexity/stability hypothesis conceptually.

As was mentioned in the introductory section, ecologists have debated the meanings of 'community' for some time. Similarly, stability has been construed as the return of species abundances to pre-perturbation equilibrium values, the resistance of invasion by exotics, and the persistence of species composition of the community after a disturbance. At first glance, one might conclude that ecology is in conceptual disarray, since ecologists cannot even agree on what they are theorizing about (Shrader-Frechette and McCoy 1993).

To formulate diversity/stability hypotheses carefully, ecologists have provided precise notions of ecological stability. This hypothesis can be understood as the following claim:

- As the diversity or complexity of a community increases, so does the stability of the community.

However, this is really a schema for a variety of hypotheses depending on how one characterizes stability, diversity, and complexity. In order to understand the concept(s) of stability, it is useful to begin with an examination of the work of Stuart Pimm (1984a, 1984b, and 1993).

Pimm distinguishes among complexity, stability, and variables of interest. The complexity of a community can be defined in terms of species richness, connectance, interaction strength, or evenness. *Species richness* is the number of species in a community. *Connectance* is the number of interspecific interactions divided by those possible. *Interaction strength* is the mean magnitude of interspecific interaction, i.e., the size of the effect of one species' density on the growth rate of another species. Species *evenness* is the variance of the species abundance distribution. The variables of interest are individual species abundances, species taxonomic composition, and trophic level abundance. One important issue to note is that diversity (species richness and evenness) forms a proper part of the complexity concept. Hence, as ecologists have moved from evaluating diversity/stability to evaluating complexity/stability, they have broadened the very nature of their hypotheses.

The stability of a community is characterized in one of the following ways (Pimm 1984b, 322):

- *Stability*: just in case all the variables in a system return to their initial equilibrium values following a perturbation
- *Resilience*: how fast the variables return to their equilibrium following a perturbation
- *Persistence*: how long the value of a variable lasts before it changes to a new value
- *Resistance*: the degree to which a variable is changed following a perturbation
- *Variability*: the degree to which a variable varies over time

Thus, there are four definitions of complexity, five of stability, and three variables of interest. Consequently, there are an extremely large number of contenders for the complexity/stability hypothesis.

Robert May (1973) was one of the first to explore precisely the connections between complexity and stability with what are called *local stability analyses*. May assumes that there is a community of species described by a set of nonlinear first-order differential equations. Let $N_i(t)$ be the density of the species i at time t. To determine the joint equilibrium density $N_i{}^*$ of the species, their growth rates, $dN_i(t)/dt$, are set equal to zero and the equations are solved. One must then determine whether the joint equilibrium density is stable or not. That is, if the species were perturbed in a relatively small way from that joint density, would it return in the limit? If the community would return, then it is asymptotically locally stable and is not locally stable otherwise.

May constructed his model communities with S species by choosing the interaction coefficients a_{ij}, a parameter measuring the effect of species j on species i, at random. Thus, some species interaction coefficients were greater than, less than, or equal to zero, and hence some species pairs interacted as competitors, predator and prey, and mutualists. He defined the connectance C of the community as the proportion of interspecific interactions a_{ij} not equal to zero. The intensity I of the interspecific interaction a_{ij} was a random variable with a mean of 0 and a variance of I^2. May infamously demonstrated that a community is qualitatively stable if and only if

$$I(SC)^{1/2} < 1.$$

Hence, *increases* in the number of species, connectance, or interaction strength all lead to a *decrease* in the stability of a community. May's result has not gone uncriticized. Nonetheless, more realistic models lead to the same general result, that stability decreases with increasing complexity.

Pimm (1984b) investigated larger perturbations of a different kind than the arbitrarily small demographic ones of May's analysis. Pimm's larger perturbation was the deletion of single species from the community. Informally, a community is species-deletion stable if and only if following the removal of a species from the community, all of the remaining species are maintained at a new locally stable equilibrium. Pimm found with qualifications that the number of interactions decreases the community's species-deletion stability.

Empirically oriented ecologists have not always looked favorably on the work of May and his mathematical cohorts. For example, J. S. McNaughton (1977) argued that the truth of the diversity/stability hypothesis depends on empirical tests and that all else are "acts of faith, not science" (516). One study he and his colleagues conducted was on the grasslands in the Serengeti-Mara ecosystem in Tanzania and Kenya. McNaughton examined the effect of the grazing African buffalo *Syncerus caffer* on the grasslands. He found that species diversity in the more diverse stands decreased more than in the less diverse stands because of grazing. Amazingly, though, the more diverse community suffered less of a reduction in primary production (biomass) than the less diverse community. McNaughton concluded from his study that "[t]he weight of evidence resulting from explicit tests of the diversity-stability hypothesis . . . suggests, not that the hypothesis is invalid, but that it is correct" (1977, 522). It thus seemed that species diversity stabilizes ecosystem properties like primary production, and so the diversity/stability hypothesis is true and the recent models incorrect.

In 1983, Anthony King and Stuart Pimm replied to McNaughton's work, attempting to "resolve this apparent contradiction between theory and empiricism" (King and Pimm 1983). King and Pimm devised grazing food web models with n plant species and one herbivore. They examined the models with respect to three types of complexity—species richness, connectance, and species diversity. They found that each type of complexity increased relative to biomass stability, which is the ratio of the total plant biomass without the herbivore to the total plant biomass with the herbivore. They also found that if stability is determined by species composition of the community, then stability decreases with increasing complexity. So King and Pimm's and McNaughton's results generally coincide. King and Pimm (1983) argue that McNaughton was incorrect in supposing that either the field ecologists or the mathematical modelers were right. There are different types of stability, which increasing complexity can increase or decrease independently of each other. The conflict between the work of McNaughton and that of the modelers was only apparent.

Since 1982, David Tilman has continued the work of May, Pimm, and McNaughton by conducting large-scale experiments at Cedar Creek Natural History Area in Minnesota. These experiments have shown that species richness is positively correlated with plant community stability—there is a decreased coefficient of variability of plant community biomass with increasing numbers of species. However, diversity does not seem to have much effect on the variability of the component populations. There is still much controversy over whether increasing diversity *causes* decreasing plant biomass variability.

Lastly, Shrader-Frechette and McCoy (1993) argue that the terms "stability" (and "community") are "ambiguous, imprecise, and inconsistent." They claim that if community ecology is to produce predictive, general theories that are adequate for environmental applications, the foundational concepts of ecology must be clear and precise. Otherwise, there will be conceptual confusion, and different interpretations of those concepts will lead to different conservation strategies (54, 57–8). They conclude that the theories of community ecology are not well equipped for conservation purposes (for a response, see Odenbaugh 2001).

Ecological Theories: Contingency, Predictive Accuracy, and Explanation

This section will consider various methodological problems that have haunted ecological theory. Ecology has not suffered from a lack of theories. However, these models and the practice from which they arise have been heavily criticized (Peters 1991; Shrader-Frechette and McCoy 1993). As Simberloff (1981) writes,

> Ecology is awash in all manner of untested (and often untestable) models, most claiming to be heuristic, many simple elaborations of earlier untested models. Entire journals are devoted to such work, and are as remote from biological reality as are faith-healers. (3)

Critics have been skeptical of the construction of theory for these and other reasons. Whatever its merits, this skepticism does force one to wonder how the success of model building, if it is successful, *could* arise. Put dramatically, how is it distinct from a sort of numerology? This section will consider three questions:

- Can ecologists build *successful* theories and models?
- How should ecologists *evaluate* their theories and models?
- Can ecological theories or models be *explanatory?*

For want of space, the treatment here will consider some characteristic answers and is not intended to be exhaustive.

Can Ecologists Build Successful Models and Theories?

Ecologists have long desired general theories that account for the behavior of populations, communities, and ecosystems. More than any other ecologist, MacArthur has been associated with the building of such theories, often in mathematical form. He argues that ecologists are in the business of finding and explaining general patterns in the distribution and abundance of organisms. They should seek theories that minimize history and emphasize the equilibria so dear to their hearts.

However, if a research program like MacArthur's is to succeed, the biological world must cooperate. Philosophers and ecologists have suggested two problems with such theorizing. First, there is the problem of *contingency* (Sterelny 2001). One can argue that there simply are no general patterns about which ecologists can theorize. For example, historical accidents of distribution involving geographic barriers can play important causal roles in determining which species occur where. Australia, for instance, has bats and marsupials but very few other mammals. As Kim Sterelny (2001) writes:

> The worry posed by extreme versions of the contingency hypothesis is that there are no patterns at all. The thought here is that membership and abundance within a community is sensitive to so many causal factors that we cannot project from one community to another. (158–9)

More generally, ecological systems can be sensitively dependent on their prior states. This means that if the system's state at time t had been otherwise, then the system at $t + \Delta t$ would be significantly different. However, if ecological systems exhibit this "sensitive dependence" or if history matters, then ecologists should provide narratives, not mathematical models. At least in part, ecology would consist in labor-intensive case studies (Shrader-Frechette and McCoy 1993).

The second problem is that of *complexity*. Ecological systems can be exceedingly complex. They have large numbers of parts that usually interact in nonlinear ways. Moreover, ecologists themselves are limited cognitively. First, there are limitations that arise from the inability to manipulate experimentally the systems as directly as is desirable. In the field, there are multifarious factors at work and only some of them are recognized at any given time. Second, there are limitations in the ability to use mathematical representations of the systems of interest. Present capacities for storing and retrieving information, carrying out various inferences, and abstracting from details make it difficult to use certain types of mathematical formalisms. Hence, ecological modeling may be too labor intensive and mathematically intractable to be of any use for prediction (Levins 1966). There may be general ecological patterns that cannot be discerned or explained.

In light of the problems of contingency and complexity, many ecologists have accepted theoretical pluralism (McIntosh 1987). First, metaphysically, ecologists must grant that there is no single biotic or abiotic process that is responsible for ecological patterns. Second, models must be built with differing degrees of realism, generality, and precision. Some models should be more mechanistic and some more phenomenological. Moreover, one may have to trade these desiderata off, as Levins (1966) has long suggested. Finally, methodologically there must be a dynamic division of labor amongst modelers, laboratory experimentalists, and field workers.

How Should Ecologists Evaluate Their Theories and Models?

There are two issues to consider here concerning the role of prediction in modeling:

1. Should ecological models be evaluated on the basis of their predictive accuracy and that alone?
2. Provided that some models make predictions that can be tested, how should those predictions be evaluated?

Critics of ecological modeling offer the following argument. If models are going to be epistemically successful, then these theoretical hypotheses must be empirically testable. However, models are not straightforwardly testable. They make either *no* predictions, no *testable* predictions, or testable *false* predictions. Ecologist R. H. Peters (1991) writes,

> Ecology seeks to predict the abundances, distributions and other characteristics of organisms in nature. . . . This book contends that much of contemporary ecology predicts neither the characteristics of organisms nor much of anything else. Therefore it represents neither ecological nor more general scientific knowledge. (17)

Therefore, theoretical models are not a successful part of ecology.

Different critics recommend different ways of coping with the predictive failure of models. Even without delving into those proposals, serious problems can be seen with the preceding argument. First, some ecological models can accurately represent some empirical systems. Second, the argument assumes that predictive accuracy is the most important function of models. However, models, even empirically inaccurate ones, can be used for a variety of purposes. For example, they allow ecologists to explore possibilities, clarify ecological concepts, and provide conceptual frameworks for experimentation and fieldwork. As theoretical ecologist Hal Caswell (1988) argues, it is false to think that the only important thing to do with theories is to test them, that refuted theories must be abandoned, and that idealizations are a "methodological evil."

Models must be evaluated for performing the tasks for which they are designed. Theoretical ecologists like Caswell have argued that theories are tools: theoretical instruments that allow biologists to further their cognitive goals, which include predictive accuracy, but not exclusively so. As William Wimsatt (1987) has suggested, "False models can lead to truer theories." However, ecologists and philosophers have been slow to explain how the heuristics of model building work and what the standards are.

These pragmatists have also suggested that model building is *inescapable* for ecologists. For example, Charles Elton, without mathematics, suggested that communities that are more complex are more stable. Through the work of May and Pimm, it can be seen that there are many different ways of characterizing stability, complexity, and variables of interest. As Caswell writes:

> None of these distinctions were, or could have been, drawn by Elton. Their importance became apparent only as the original verbal theory was studied using mathematical models. (1988, 35)

The same is true in more applied matters. One trend in applied ecology is population viability analysis (Boyce 1992). Ecologists utilize simple logistic equations, Leslie projection matrices, and probabilistic models of demographic and environmental stochasticity to simulate the expected time to extinction of various species. These tools are needed because the relevant autoecological data are lacking and, if conservation is at stake, it is impossible to perturb experimentally these demographically depressed populations. There is no choice but to predict the fates of endangered species with mathematical models even if they are not especially accurate. Thus, model building is an essential part of theoretical and applied ecology.

Turning to the question of how the predictions of models should be tested, one of the most cantankerous debates in ecology concerns the *null hypotheses* (Gotelli and Graves 1996). In essence, the debate arose over the importance of interspecific competition in structuring properties of organisms such as body size and resource use. This debate also led to more general issues surrounding how ecological theories should be tested and evaluated. In 1975, Jared Diamond published a study on the distribution of

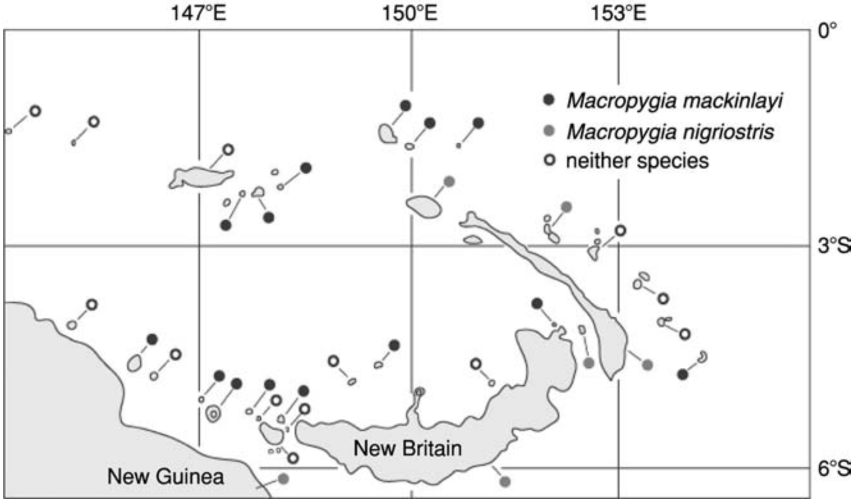


Fig. 2. The distribution of *Macropygia* species in the Bismarck Archipelago.
[*From Ricklefs and Miller 2000, 613*]

bird species among fifty islands in the Bismarck Archipelago near New Guinea (Diamond 1975). Diamond recognized that certain combinations of species were never found together in the archipelago. For example, two species of cuckoo dove, *Macropygia nigriostris* and *M. mackinlayi*, occurred on six and fourteen islands, respectively (Figure 2). However, they never co-occurred on any island.

This "checkerboard pattern," or complementary distribution, suggested that interspecific competition was at work through niche differentiation.

In the late 1970s, Edward Connor and Daniel Simberloff (1979) argued that Diamond's work was seriously flawed. They suggested that the checkerboard distribution could have resulted from random colonization rather than competition. They devised *neutral* or *null models* of communities that retained certain features, such as the number of species per island, the relative abundances of species, and their incidence functions (the probability of a species occurring on an island given the total number of species on that island), but they reassembled the rest at random excluding the effects of competition. If the actual data differ in statistically significant ways from the null hypothesis, then the null is rejected and the interaction is strongly suggested. Connor and Simberloff claimed that null hypotheses were more parsimonious and "logically prior" to competitionist hypotheses. Contrary to the Popperian philosophy adopted by Connor and Simberloff, Diamond looked for confirming evidence as opposed to first trying to refute a null hypothesis.

The work of Simberloff and his colleagues has been heavily criticized. First, in traditional Neyman-Pearson statistical testing, one formulates two mutually exclusive and exhaustive hypotheses, the null and the alternative. However, the null hypotheses articulated by the Florida group were not always logically inconsistent with competitionist hypotheses according to Michael Gilpin and Diamond. Key features of the null models—the species pools, dispersal abilities of species, and "incidence functions" of species—could be affected by competition (Gilpin and Diamond 1983). Hence, the "ghost of competition past" might be built into the null model itself, and thus it might have a "hidden structure." Second, Connor and Simberloff (1979) performed their analyses using sets of species that were not restricted to guilds (groups of species that utilize similar resources in similar ways). Competition is to be expected between two species only if they occupy the same guild. One would thus bury the effects of competition in a morass of irrelevant data (Gilpin and Diamond 1983).

It should be noted that Connor and Simberloff argued that even if one could delineate guild assignments with good evidence and had the checkerboard pattern of Figure 2, one still could not conclude that interspecific competition had been in operation. Likewise, they argued (Strong et al. 1984) that Gilpin and Diamond had not provided independent evidence for their hidden-structure claims.

The null model controversy continued in paper after paper and forced ecologists to address subtle issues concerning how predictions of ecological theory should be evaluated. It has invigorated hypothesis testing in ecology and led to more refined statistical tools for judging theory.

Can Ecological Theories or Models Be Explanatory?

Theories and models in ecology presumably provide scientific understanding of the systems they represent. A common philosophical supposition is that a theory or model explains some event or regularity only if it is true. Generally, though, ecological models are highly idealized—whatever their virtues, truth is not one of them. Hence, they cannot be explanatory. However, it does appear that ecological models do explain some events and regularities. Thus, models in ecology are not explanatory, or truth is not a necessity for successful scientific explanation.

As an example, consider the following *why*-question: Why is omnivory [feeding on more than one trophic level] rare in vertebrate food webs rather than common? Pimm and Lawton (1978), using Lotka-Volterra community models, gave one possible explanation for this. They demonstrated by computer simulations that food webs with omnivores were generally dynamically unstable. Either they were locally unstable or, if locally stable, their return time was excessively long. Hence, vertebrate food webs with omnivores would be unlikely to persist. A possible answer to this *why*-question is that vertebrate food webs with omnivores are dynamically fragile and hence do not persist.

The Lotka-Volterra community model is a caricature of empirical food webs. Some of the idealizations of the model include assuming that there is no migration and no age or genetic structure in the populations and that the density dependence is linear. Nonetheless, Pimm and Lawton argued that dynamical models explain various patterns of food webs, including the infrequency of omnivory. The fact that the model is severely idealized does not render it unexplanatory.

There are several ways to deal with this problem, one of which will be discussed here (see Cartwright 1983; Wimsatt 1987). Philosopher Gregory J. Cooper (2003) has offered a position that countenances the possibility that false ecological theories and models are explanatory. He argues, following Nancy Cartwright, that ecological models represent the *capacities* or *tendencies* of objects, which is how they would behave if there were no interfering forces. The dynamical equations of the models are true only of these dispositions or propensities. So, for example, the Lotka-Volterra model is false of most, if not all, actual food webs, though true of "interference-free" food webs. Cooper's proposal requires that capacities and tendencies exist that may sound implausible to empiricists. However, like Cartwright, he believes that much of science cannot be accounted for without them. Nonetheless, even if the existence of capacities and tendencies is accepted, how idealized models explain ecological dynamics when there are interfering forces still needs to be understood.

Conclusion

Ecology presents philosophy with several conceptual and methodological problems. These issues are not just of an abstract bent, but speak to how one should understand the role of science in society (see Conservation Biology). Many issues of importance are connected to the empirical studies and theoretical analyses that ecologists perform. These include determining the status of invasive species, considering whether a population is threatened or endangered, and estimating the risks in losing many of the communities of plants and animals across the globe. To make sense of the roles these ecologists play in policy formation, their scientific activities must also be considered. These issues are enveloped in political and ethical issues of some complexity—all the more reason to exercise philosophical care. After all, how the science of ecology is understood affects both human lives and the environment.

JAY ODENBAUGH

References

Boyce, Mark (1992), "Population Viability Analysis," *Annual Review of Ecology and Systematics* 23: 481–506.

Brennan, Andrew (1988), *Thinking About Nature: An Investigation of Nature, Value and Ecology*. Athens: University of Georgia Press.

Cartwright, Nancy (1983), *How the Laws of Physics Lie*. Cambridge: Cambridge University Press.

Caswell, Hal (1988), "Theory and Models in Ecology: A Different Perspective," *Ecological Modelling* 43: 33–44.

Clements, Frederic (1916), *Plant Succession*. Washington, DC: Carnegie Institution of Washington, Publication no. 242.

Connor, E. F., and Daniel Simberloff (1979), "The Assembly of Species Communities: Chance or Competition?" *Ecology* 60: 1132–1140.

Cooper, Greg (2003), *The Science of the Struggle for Existence: On the Foundations of Ecology*. Cambridge: Cambridge University Press.

Diamond, Jared (1975), "Assembly of Species Communities," in M. L. Cody and J. M. Diamond (eds.), *Ecology and Evolution of Communities*. Cambridge, MA: Belknap Press of Harvard University Press.

Egerton, Frank (1973), "Changing Concepts of the Balance of Nature," *Quarterly Review of Biology* 48: 322–350.

Gilpin, Michael, and Jared Diamond (1983), "Are Species Co-occurrences on Islands Non-Random, and Are Null Hypotheses Useful in Community Ecology?" in Donald Strong, Daniel Simberloff, L. G. Abele, and A. B. Thistle (eds.), *Ecological Communities: Conceptual Issues and the Evidence*. Princeton, NJ: Princeton University Press.

Gleason, Henry (1926), "The Individualistic Concept of the Plant Association," *Bulletin of the Torrey Botanical Club* 53: 7–26.

——— (1917), "The Structure and Development of the Plant Association," *Bulletin of the Torrey Botanical Club* 44: 463–481.

Gotelli, Nicholas, and Gary Graves (1996), *Null Models in Ecology*. Washington, DC: Smithsonian Institution.

King, Anthony, and Stuart Pimm (1983), "Complexity and Stability: A Reconciliation of Theoretical and Experimental Results," *American Naturalist* 122: 229–239.

Levins, Richard (1966), "The Strategy of Model Building in Population Biology," *American Scientist* 54: 421–431.

Levins, Richard, and Richard Lewontin (1985), "Dialectics and Reductionism in Ecology," in *The Dialectical Biologist*. Cambridge: Harvard University Press.

Ludwig, Donald, Mark Mangel, and Brent Haddad (2001), "Ecology, Conservation, and Public Policy," *Annual Review of Ecology and Systematics* 32: 481–517.

MacArthur, Robert (1962), "Patterns of Terrestrial Bird Communities," in D. Farner, J. King, and K. Parkes (eds.), *Avian Biology*, vol. 1. New York: Academic Press.

May, Robert (1973), *The Stability and Complexity of Model Ecosystems*. Princeton, NJ: Princeton University Press.

McIntosh, Robert (1987), "Pluralism in Ecology," *Annual Review of Ecology and Systematics* 18: 321–341.

McNaughton, John (1977), "Diversity and Stability of Ecological Communities: A Comment on the Role of Empiricism in Ecology," *American Naturalist* 111: 515–525.

Odenbaugh, Jay (2001), "Ecological Stability, Model Building, and Environmental Policy: A Reply to Some of the Pessimism," *Philosophy of Science* 68 (Proceedings), S493–S505.

Peters, Robert (1991), *A Critique for Ecology*. Cambridge: Cambridge University Press.

Pimm, Stuart (1984a), *Food Webs*. London: Chapman and Hall.

——— (1984b), "The Complexity and Stability of Ecosystems," *Nature* 307: 321–326.

——— (1993), *The Balance of Nature?* Chicago: University of Chicago Press.

Pimm, Stuart, and John Lawton (1978), "On Feeding on More Than One Trophic Level," *Nature* 275: 542–544.

Ricklefs, Robert, and Gary Miller (2000), *Ecology*. New York: W. H. Freeman and Co., 2000.

Shrader-Frechette, Kristin, and Earl McCoy (1993), *Method in Ecology*. Cambridge: Cambridge University Press.

Simberloff, Daniel (1980), "A Succession of Paradigms in Ecology: Essentialism to Probabilism to Materialism and Probabilism," *Synthese* 43: 3–39.

Sterelny, Kim (2001), "Darwin's Tangled Bank," in *The Evolution of Agency and Other Essays*. Cambridge: Cambridge University Press.

Strong, Donald, Daniel Simberloff, L. G. Abele, and A. B. Thistle (eds.) (1984), *Ecological Communities: Conceptual Issues and the Evidence*. Princeton, NJ: Princeton University Press, 1984.

Wimsatt, William (1987), "False Models as Means to Truer Theories," in M. Nitecki and A. Hoffman (eds.), *Neutral Models in Biology*. London: Oxford University Press.

PHILOSOPHY OF ECONOMICS

The philosophy of economics concerns itself with conceptual, methodological, and ethical issues that arise within the scientific discipline of economics. The philosophy of economics is now a well-established subdiscipline within philosophy. Significant contributions to the discipline include Allen Buchanan (1985), Daniel Hausman (1984 and 1992), Hausman and Michael McPherson (1996), Daniel Little (1995), Amartya Sen (1987), and Alexander Rosenberg (1992). The primary focus is on issues of methodology and epistemology—the methods, concepts, and theories through which economists attempt to arrive at knowledge about economic processes. Philosophy of economics is also concerned with the ways in which ethical values are involved in economic reasoning—the values of human welfare, social justice, and the trade-offs among priorities that economic choices require. Economic reasoning has implications for justice and human welfare; more importantly, economic reasoning often makes inexplicit but significant ethical assumptions that philosophers of economics have found it worthwhile to scrutinize. Finally, the philosophy of economics is concerned with the concrete social assumptions that are made by economists. Philosophers have given attention to the institutions and structures through which economic activity and change take place. What is a "market"? Are there alternative institutions through which modern economic activity can proceed? What are some of the institutional variants that exist within the general framework of a market economy? What are some of the roles that the state can play within economic development so as to promote efficiency, equity, human well-being, productivity, or growth?

The dimension of the philosophy of economics that falls within the philosophy of science has to do with the status of economic analysis as a body of empirical knowledge. Primary questions include: What is economic knowledge *about*? What kind of knowledge is provided by the discipline of economics? How does it relate to other social sciences and the bodies of knowledge contained in those disciplines? How is economic knowledge justified or evaluated? Does economic theory purport to offer abstract theories of real social processes—their mechanisms, dynamics, and institutions? What is the nature of economic explanation? What is the relationship between abstract mathematical models and theorems, on the one hand, and the empirical reality of economic behavior and institutions, on the other? What is the nature of the concepts and theories in terms of which economic beliefs are formulated? Are there lawlike regularities among economic phenomena? What is the status of predictions in economics?

The Intellectual Role of the Philosophy of Economics

Philosophers are not empirical researchers, and on the whole they are not formal theory builders. So what constructive role does philosophy have to play in economics? There are several. First, philosophers are well prepared to examine the logical and rational features of an empirical discipline. How do theoretical claims in the discipline relate to empirical evidence? How do pragmatic features of theories such as simplicity, ease of computation, and the like play a role in the rational appraisal of a theory? How do presuppositions and traditions of research serve to structure the forward development of the theories and hypotheses of the discipline? Second, philosophers are well equipped to consider topics having to do with the concepts and theories that economists employ—for example, economic rationality, Nash equilibrium, perfect competition, transaction costs, or asymmetric information. Philosophers can offer useful analysis of the strengths and weaknesses of such concepts and theories—thereby helping practicing economists to further refine the theoretical foundations of their discipline. In this role the philosopher serves as a conceptual clarifier for the discipline, working in partnership with the practitioners to bring about more successful economic theories and explanations.

This account describes the position of the philosopher as the "underlaborer" of the economist. But in fact, the line between criticism and theory formation is not a sharp one. Economists such as Sen (1999, 1973, and 1987) and philosophers such as Hausman (1992; Hausman and McPherson 1996) have demonstrated that there is a very constructive crossing of the frontier that is possible between philosophy and economics; philosophical expertise can result in significant substantive progress with regard to important theoretical or empirical problems within the discipline of economics. The cumulative contents of the journal *Economics and Philosophy* provide clear evidence of the productive engagements that are possible when philosophy meets economics.

Important Questions in the Philosophy of Economics

Are There Laws in Economics?

The concept of a "law of nature" has been central to the modern understanding of the natural sciences. The intellectual power of classical physics derived from the fact that it was capable of advancing statements of physical laws that were simple and universal—laws of gravitation and planetary motion, optics, electricity and magnetism, and so on. Is this an essential feature of a successful empirical science? And does economics possess such laws? Several authors are affirmative on both points (Kincaid 1996). However, several points have emerged in recent discussions of the social sciences—including economics—that lead to doubt about the centrality of laws to them. First, there are significant differences between natural and social phenomena that should raise doubts about the availability of strong "laws of nature" describing social phenomena. Second, it is clear that there are regularities within the discipline of empirical economics—consumption usually rises when prices fall, trade increases when transport costs fall, and infant mortality usually falls when states devote more resources to public health. But these are fairly humdrum empirical regularities, exception laden and obvious. Are there strong "economic laws" that have the force of Maxwell's laws of electromagnetic propagation? Nothing in current economic theory provides reason to think that there are such laws. The foundational assumptions of economic theory plainly do not fall in the category of laws of nature. And as will be shown below, the assumption of economic rationality does not constitute a universal generalization about individual behavior. Here, as is the case in other areas of social science, it is more justifiable to seek out causal mechanisms rather than social laws (see Laws of Nature; Explanation).

How Do Economic Laws Relate to Individual Behavior?

Economists generally assume some form of the doctrine of methodological individualism (see Methodological Individualism). This principle maintains that social facts, social entities, and social laws are constituted by facts about individuals and their behavior—defined prior to the social properties that individuals possess (Kincaid 1986; Miller 1978; Watkins 1968). The doctrine of methodological individualism is an expression of the perspective of reductionism, that is, the view within scientific metaphysics that insists that higher-level structures must be explained by reference to the properties of the lower-level entities that make them up. According to the reductionist, there are no "emergent" properties of complex entities or structures. A more satisfactory perspective on this issue has been developed on the basis of the theory of supervenience—the view that differences in higher-level properties depend upon differences in lower-level properties (Kim 1984). Extreme versions of methodological individualism and reductionism make the task of social or economic explanation very difficult, since explaining or describing individual behavior seems to require referring to social entities (rules, structures, institutions).

Are the Assumptions of Economics "Realistic?"

Do economic theories and hypotheses serve to describe unobservable economic mechanisms and structures? Milton Friedman (1953) set the stage for one answer to this question by arguing for an instrumentalist interpretation of economic assumptions. In Friedman's view, the value of a theory is entirely expressed in its ability to predict observable phenomena; the theory is an instrument of prediction (see Instrumentalism). Instrumentalism, however, has generally faced strong criticism from philosophers of science (Leplin 1984). This doctrine makes the empirical success of a theory a source of mystery. The best explanation of a theory's having generally reliable predictions about a range of phenomena is that the mechanisms that it postulates are in fact true. So it is a deficiency in a theory that the mechanisms it postulates are implausible or false. And economic theory would be substantially undermined if its premises were judged to be profoundly inconsistent with the real underlying causal processes that constitute a working economy. Against this instrumentalist framework, Hausman puts forward a realist approach to economic theory (Hausman 1992) (see Realism). Within this approach, the goal for the economist is to arrive at assumptions that are approximately true. (This methodological principle suggests that economists ought to pay greater attention to economic institutions, comparative economic analysis, and economic history; see below.)

Are Economic Theories Testable or Falsifiable?

Before asking whether a theory is testable, it is necessary to have a clear specification of its empirical content. This requires asking the question, What is the theory intended to describe, predict, or explain? A theory has empirical content if it makes assertions about causal processes underlying a domain of phenomena and those assertions have consequences for observable states of the world. Under these circumstances it is possible to engage in a variety of forms of empirical test of the hypothesis. The investigator can

- perform experiments (arrange the world in a certain way, observe the outcome, and compare this with the theory's predicted outcome);
- undertake a protocol of controlled observation (collect "before/after" cases and compare the outcomes with the theory's predictions);
- perform process-tracing observation (examine elements of a process in order to assess whether the postulated causal processes did in fact occur); and so on.

Through these efforts it is possible to bring empirical evidence to bear on the task of assessing the truth of the hypothesis. So the question is this: Does economic theory contain substantive assumptions about the causal workings of the economic world that are intended to have implications for future observable states of the economic world? And is it possible to perform observations of states of the world that confirm or falsify the theory (Hands 1992)? In principle, it is clear that the answer to this question is affirmative. Consider a range of theories of specific economic processes—economic growth, trade, unemployment, wages, or discrimination. Such theories have predictive consequences, and it is not especially difficult to describe the observations that would test these theories. The epistemic difficulty comes later: Most theories of complex phenomena are in fact falsified—without necessarily being far from the mark in their description of the underlying processes. So, how is it possible to distinguish among "falsified" theories to single out the more likely from the less likely (Lakatos 1974)? Are economic theories simply formal mathematical systems, without empirical relevance?

Rosenberg makes a case for the formalist view of economic theory, having concluded that economists have not succeeded in producing empirical theories or explanations of real empirical phenomena (Rosenberg 1992, Ch. 8). Rosenberg likens microeconomics to Euclidean geometry rather than classical physics or evolutionary biology; the "theory" is a set of abstract and nonempirical axioms, and the exercise of "doing economics" is one of deriving theorems from these axioms. However, this is not a satisfactory way of understanding the intellectual program of economics. The intellectual charge for the discipline of economics—not always or successfully achieved—is to provide a social-scientific basis for understanding, explaining, and, perhaps, predicting economic phenomena. Why do interest rates affect investment levels? Why are inflation and unemployment related? Why is economic growth more rapid in the context of one set of institutions than another? What are the causal links that secure connections among economic variables? These are the sorts of questions that economists are charged to answer. And the approach to economic theorizing that stipulates that the discipline is purely formal will not aid in shedding light on these real, though unobservable, economic mechanisms. In this line of thought, the persistent mathematization of economics ought to be construed as a means to an end rather than the end itself. The formal or mathematical machinery of economics is intellectually valuable only insofar as it contributes to a better understanding of real, empirically given economic processes, causes, and systems.

What Is the Status of the Concept of Economic Rationality?

The concept of economic rationality is foundational within economic theory, and especially so within neoclassical economics. Economists are concerned with analyzing individual rational choice in the context of reward, risk, and uncertainty, where the individual's outcome depends on the probabilities and rewards associated with the various options available for choice (see Decision Theory). And they are concerned with rationality in the context of strategic interactions among two or more agents, where the individual's reward depends on the rational choices made by other agents (see Game Theory). What is the nature of the "decisions rules" that constitute rational behavior in these two stylized contexts? A special concern for philosophers of economics has been to provide critical examination of the theory of economic rationality. Taken together, these criticisms have led to a substantial enhancement in our understanding of the concept of rationality. First, philosophers have devoted a great deal of attention to the gap between a theory of utility and a theory of individual preference. Second, they have taken issue with the assumption of egoism or rational self-interest that is presupposed in the pure theory (Sen 1987; Anderson 2000). Third, philosophers and others have pointed out that real psychological actors reason in ways that are at odds with the pure theory of economic rationality (Kahneman, Slovic, and Tversky 1982; Simon 1983). Fourth, philosophers and others have devoted significant attention to the assumptions underlying game theory. Finally, some philosophers have undertaken to study the characteristics of "economic rationality" in real human persons through experiment (Schmidtz 1991). For example, Axelrod (1984) has used experimental settings to examine how real human reasoners deal with Prisoners' Dilemmas; he finds that experimental subjects are frequently able to achieve cooperation rather than defection, contrary to the prediction of two-person game theory. The results of this research suggest that real reasoners behave intelligently—but differently from the axioms of the theory of pure economic rationality.

What Is the Role of Ethical Values in Economics?

Economists often portray their science as "value free"—as a technical analysis of the demands of rationality in the allocation of resources rather than a specific set of value or policy commitments. In this interpretation, the economist wishes to be understood in an analogy with the civil engineer rather than the transportation policymaker: The engineer can prescribe how to build a stable bridge, but not where, when, or why to do so. It is for citizens and policymakers to make the judgments about the public goods that are needed in order to decide whether a given road or bridge is socially desirable; it is for the technical specialist to provide design and estimate of costs. This description of the discipline of economics fails in several important respects, however. Economic theory contains a family of substantive presuppositions about the nature of the good—individual and social—that directly influence the policy recommendations to which economic theory gives rise. For example, the assumption of rational egoism is inconsistent with several of the values of communitarianism; the assumption that equity is subordinate to efficiency is inconsistent with an egalitarian political philosophy; and the assumption that a bundle of commodities constitutes individual "well-being" is inconsistent with a more Aristotelian conception of the good human life (Nussbaum 2000). So the premises and assumptions of economics are substantially intertwined with normative assumptions about the good human life and the good society. This is not a deficiency, but it needs to be recognized in order to observe the workings of the unstated value assumptions. And it certainly invalidates the assumption of value-free social science. In general, it seems fair to say that the ethical assumptions that neoclassical economics presupposes fall together into a family of normative ideals that privilege individualism, inequality, and the minimal exercise of public policy.

Is Distributive Justice a Topic for Economists?

Once it is recognized that economics has ethical content, it becomes apparent that it is necessary to examine the content of these ethical premises in detail and offer a critique when these assumptions are found wanting. In particular, economics is obliged to confront issues of distributive justice much more explicitly than it has to date. A market economy implies some degree of inequality, in terms of outcomes (wealth and income), opportunity, power and influence, and levels of well-being (health, longevity, education). What sorts of inequalities are morally acceptable in a just society? How extensive can inequalities be before they create differences among citizens that interfere with their human dignity and the preconditions of democracy? Throughout the past 30 years philosophers have made substantial contributions to current understanding of these issues of distributive justice and the moral status of inequality (Nozick 1974; Elster 1992; Rawls 2001). There is more to be done.

Is There a Basis for Rational Debate About Economic Institutions?

What sort of social world does economic theory presuppose? In considering this type of question, philosophers begin to move into substantive debates about the nature of the empirical phenomena under study. The discussion falls under the rubric of "criticism," in that it focuses on blind spots that can be discerned within the visual field of economic theorizing. Economists make assumptions about the institutions that constitute the framework of economic transactions, and these assumptions are sometimes inflexible and unrealistic. It is therefore worthwhile for philosophers to devote attention to the shortcomings of the social-institutional assumptions that economists often make. The new institutionalism in the social sciences has focused substantial interest on the specifics of the institutions within which social activity takes place (Powell and DiMaggio 1991; Brinton and Nee 1998). Institutions matter; so a more refined account of the economic institutions of a particular market economy may lead to better understanding of the phenomena. For example, incorporation of transaction costs and asymmetric information between buyer and seller has significantly changed the current understanding of market institutions. One strand of philosophical criticism comes from the level of abstractness of typical economic theories. Greater empirical detail may well change the inferences that can be drawn about the workings of the institution. Market "imperfections" may be the rule rather than the exception—so it is important to incorporate some of these empirical characteristics into accepted theories of economic institutions.

Are There Alternative Economic Institutions That Can Work in a Modern Economy?

Economic activity within a modern society requires institutions that define the use, management, and enjoyment of resources; the deployment and management of labor; and the management of enterprises. Neoclassical economics presupposes private ownership of capital, "free" workers who do not own property, and states that have minimal economic influence. Are there other institutions through which economic activity might be conducted within a modern and productive society (Elster and Moene 1989)? For example, what is the economic logic of workers' cooperatives? How could worker-controlled pension funds be used to enhance democratic equality? Is there more to be learned from the experience of market socialism, state ownership, or workers' control of industrial processes? Are alternative institutions feasible? Are they efficient? Are they equitable?

What Can Be Learned from Comparative Economic Analysis?

Economic development has proceeded in very different ways in different nations and regions since the emergence of modern technologies and economic institutions. Market institutions developed very differently in Britain, France, and the United States during the nineteenth and twentieth centuries. Collectivized economies followed different institutional trajectories in Yugoslavia, the Soviet Union, and China. What can be learned about economic processes and dynamics by studying and comparing national economies in significant detail? For example, what do the parallel yet different experiences of China and India since 1945 teach about alternative pathways of economic development (Drèze and Sen 1989)? Does this sort of comparative economic research provide a "post–Cold War" basis for analyzing the political economy of development? As economists come to confront the intellectual challenge of providing realistic causal accounts of economic systems, they will be able to arrive at significant new insights through comparative economic analysis.

What Is the Intellectual Relevance of the History of Western Industrial Capitalism for Economic Theory?

Reexamination of the history of European capitalism suggests that there were feasible alternative paths of economic development besides mass manufacture and specialized production (Sabel and Zeitlin 1997). Mass manufacture and mass unskilled labor represented one important alternative, but there were others that were historically feasible. As Sabel and Zeitlin demonstrate, another feasible system of industrial production involves highly skilled workers, flexible production, and flexible tools and production processes (Sabel and Zeitlin 1985). Once again, the moral for the discipline of economics is an important one: It is possible to arrive at more empirically satisfactory economic theories in the context of considering the range of institutions through which economic activity and growth have taken place.

Conclusions

The philosophy of economics serves as a source of sympathetic yet rigorous critique of the science of economics, broadly construed. It raises familiar questions about the epistemology of this branch of the social sciences—questions about theory structure, theory confirmation, explanatory adequacy, and the like. It questions the implicit normative assumptions that economics contains. It raises some of the ethical questions that economics is almost forced to confront, but rarely does. And it suggests the value of a broader and more eclectic approach to economic theorizing: making more extensive use of alternative theoretical approaches, incorporating more study of economic institutions, paying more attention to comparative economic trajectories, and giving more rigorous attention to economic history. Economics will be a more successful social science when it embraces more of the role it often played in the nineteenth century as a seminal social science—an area of social inquiry that was equally interested in the concrete social and economic institutions that constituted a "modern" economy, interested in the ethical implications of the social phenomena with which it was concerned and willing to consider a variety of theoretical models in aspiring to the goal of achieving a scientific understanding of economic processes, institutions, and outcomes.

Daniel Little

References

Anderson, Elizabeth (2000), "Beyond Homo Economicus," *Philosophy and Public Affairs* 29: 170–200.

Axelrod, Robert M. (1984), *The Evolution of Cooperation*. New York: Basic Books.

Brinton, Mary C., and Victor Nee (eds.) (1998), *New Institutionalism in Sociology*. New York: Russell Sage Foundation.

Buchanan, Allen E. (1985), *Ethics, Efficiency, and the Market*. Totowa, NJ: Rowman & Allanheld.

Drèze, Jean, and Amartya Kumar Sen (1989), *Hunger and Public Action*. Oxford: Clarendon Press.

Elster, Jon (1992), *Local Justice*. New York: Russell Sage Foundation.

Elster, Jon, and Karl Ove Moene (eds.) (1989), *Alternatives to Capitalism*. Cambridge and New York: Cambridge University Press.

Friedman, Milton (1953), *Essays in Positive Economics*. Chicago: University of Chicago Press.

Hands, D. Wade (1992), *Testing, Rationality, and Progress: Essays on the Popperian Tradition in Economic Methodology, Worldly Philosophy*. Lanham, MD: Rowman & Littlefield Publishers.

Hausman, Daniel M. (ed.) (1984), *The Philosophy of Economics: An Anthology*. Cambridge and New York: Cambridge University Press.

——— (1992), *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.

Hausman, Daniel M., and Michael S. McPherson (1996), *Economic Analysis and Moral Philosophy*. Cambridge and New York: Cambridge University Press.

Kahneman, D., P. Slovic, and A. Tversky (1982), *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kim, Jaegwon (1984), "Supervenience and Supervenient Causation," *Southern Journal of Philosophy* 22(suppl): S45–S56).

Kincaid, Harold (1996), *Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research*. Cambridge and New York: Cambridge University Press.

——— (1986), "Reduction, Explanation, and Individualism," *Philosophy of Science* 54: 492–513.

Lakatos, Imre (1974), "Methodology of Scientific Research Programmes," in Lakatos and Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

Leplin, Jarrett (1984), *Scientific Realism*. Berkeley and LA: University of California Press.

Little, Daniel (1995), *On the Reliability of Economic Models: Essays in the Philosophy of Economics. Recent Economic Thought Series*. Boston: Kluwer Academic Publishers.

Miller, Richard (1978), "Methodological Individualism and Social Explanation," *Philosophy of Science* 45: 387–414.

Nozick, Robert (1974), *Anarchy, State, and Utopia*. New York: Basic Books.

Nussbaum, Martha Craven (2000), *Women and Human Development: The Capabilities Approach*. Cambridge and New York: Cambridge University Press.

Powell, Walter, and Paul J. DiMaggio (eds.) (1991), *The New Institutionalism in Organizational Analysis*. Chicago: University of Chicago Press.

Rawls, John (2001), *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.

Rosenberg, Alexander (1992), *Economics: Mathematical Politics or Science of Diminishing Returns?* Chicago: University of Chicago Press.

Sabel, Charles F., and Jonathan Zeitlin (1985), "Historical Alternatives to Mass Production: Politics, Markets and Technology in Nineteenth Century Industrialization," *Past and Present* 108: 133–176.

——— (eds.) (1997), *Worlds of Possibility: Flexibility and Mass Production in Western Industrialization*. Cambridge and New York: Cambridge University Press.

Schmidtz, David (1991), *The Limits of Government: An Essay on the Public Goods Argument*. Boulder, CO: Westview Press.

Sen, Amartya (1987), *On Ethics and Economics*. New York: Basil Blackwell.

——— (1973). *On Economic Inequality*. Oxford: Oxford University Press.

——— (1999). *Development as Freedom*. New York: Knopf.

Simon, Herbert A. (1983), *Reason in Human Affairs*. Stanford, CA: Stanford University Press.

Watkins, J. W. N. (1968), "Methodological Individualism and Social Tendencies," in May Brodbeck (ed.), *Readings in the Philosophy of the Social Sciences*. New York: Macmillan.

See also **Behaviorism; Game Theory; Methodological Individualism; Social Sciences, Philosophy of**

EMERGENCE

The concept of emergence stems from a family of related doctrines known collectively as emergentism. Regardless of variation in its formulation, emergentists generally hold an ontological premise and an epistemological one. The *ontological premise* is that (i) there are properties (or laws) that obtain of certain complex physical entities that do not obtain of any of the individual parts or lower-level constituents of those entities. The *epistemological premise* is that (ii) the instantiation of those properties cannot be derived from an exhaustive knowledge of the nonrelational properties of the parts, in addition to any laws of composition that obtain among lower-level entities (e.g., additivity, fundamental forces) and statements of definition. (In the following, it will generally be assumed that fundamentally, properties are emergent and that laws are emergent in a derivative sense; thus "properties" will be used instead of "properties and laws.") If one allows a broad reading to include nonmaterial parts (e.g., *élan vital*, mind substance), then it follows from (i) and (ii) that emergentism rejects the use of any sort of substance dualism for the purpose of explaining the appearance of a higher-level property. Furthermore, if a "reductionist" explanation is understood as one that explains a property of an entity in terms of the nonrelational properties of its parts, in addition to the lower-level laws that obtain over these properties, then it also follows from (i) and (ii) that the instantiation of some properties cannot be given a reductionist explanation (see Reductionism). Hence emergentism takes its place in contemporary philosophical parlance as a variety of nonreductionist physicalism (see Physicalism).

Paradigmatic examples of emergence often touted by early emergentists were taken from chemical bonding—for instance, the production of common salt from sodium and chloride. At the time, little was known about the microstructure of atoms, and consequently it seemed plausible that the phenomenon in question could be conceptualized only in terms of a *de novo* or fundamental physical law relating atomic interactions to macro-level phenomena, rather than one derivable on the basis of the atomic structure of sodium and chloride and generic laws of chemical bonding. (In fact, the success of quantum chemistry in providing reductions of this type remains disputed; see final section.) Although, given such a fundamental law, the instantiation of a macro-level property would be *determined* by the instantiation of a set of micro-level properties, this property would not be reductively explainable on the basis of the latter.

A stronger form of emergentism includes a third premise, that (iii) these properties that hold of certain complex physical entities are not *determined* by the nonrelational properties that hold of their individual parts and the relation of these parts to one another. This view is sometimes characterized as involving a failure of "microdeterminism" (following Klee 1984). Karl Popper (see Popper and Eccles 1977) argues for such a view on the basis of the probabilistic character of quantum mechanical laws. Finally, some emergentists hold a fourth premise, that (iv) the emergent properties of a whole system can affect the behavior of that system's own parts. This view, referred to as "downward causation" (Campbell 1974), "hierarchical downward control" (Sperry 1969), or "macrodeterminacy" (Weiss 1968), has played a significant role in the emergentist literature since the 1960s, although an intimation of this idea can be found in C. Lloyd Morgan's notion of the "dependence" of lower-level properties on those of the higher level (Lloyd Morgan 1923, 16).

Moreover, there are regional variants of these views. In Anglo-American philosophy of science and philosophy of mind, most references to the history of emergentism implicitly or explicitly refer to the British emergentist tradition, which traces its intellectual roots back to John Stuart Mill in the mid-nineteenth century and the doctrines of "emergent evolution" that were prominent among philosophers such as Scott Alexander and biologists such as Lloyd Morgan and W. M. Wheeler in the early twentieth century. However, some writers on emergentism also refer to the German organicist tradition, which traces its intellectual roots back to Immanuel Kant in the late eighteenth century and was prominent among German biologists and psychologists in the early twentieth century.

History of the Concept of Emergence

Mill's Heteropathic Effects

What is now referred to as British emergentism (McLaughlin 1992) begins with Mill's *A System of Logic* of 1843, in which he describes two classes of phenomena. The first is that class of effects produced by the mechanical mode of causation, or according to the "composition of causes" (1874, 267), that is, the principle that the effect produced by the joint action of several causes can be inferred by summating the effects that would have been produced by each of the agents acting separately. The paradigmatic example of the mechanical mode of causation is the composition of forces, in which the velocity and direction of a given particle at time t can be derived by summating over the velocities and directions of each of the particles that strike it at a prior time t' by vector addition.

This is contrasted with that class of effects produced by the chemical mode of causation. The laws governing this mode are referred to as *heteropathic laws* (269). By this expression Mill refers to the new uniformities that arise in those cases in which the combination of different substances produces a substance with new properties not possessed by any of the parts. For Mill, the production of heteropathic laws is a sufficient condition for the origination of higher levels of the scientific hierarchy. (Lewes [1875] is credited with coining the term "emergents"—to be distinguished from "resultants"—to refer to Mill's heteropathic effects.)

Mill is often ambiguous concerning whether the notion of a heteropathic effect should be understood primarily ontologically or epistemologically. His paradigmatic example of a heteropathic effect is chemical bonding, in which some of the properties of a compound are qualitatively different from those of its constituents and hence cannot be derived by mere summation of some known property of those constituents. It would appear from this example that, for Mill, the failure of additivity for the higher-level property is a consequence of the qualitative *novelty* of the property in question, and hence his form of emergentism would amount to an ontological doctrine that presupposes some account of the individuation of properties independent of their extension. This ontological doctrine would then presumably explain the failure of derivability by summation. However, Mill's view can also be interpreted epistemologically, in that what is to be considered a heteropathic effect is relative to a given state of knowledge at a particular time:

> There most of the uniformities to which the causes conformed when separate, cease altogether when they are conjoined; and we are not, at least in the present state of our knowledge, able to foresee what result will follow from any new combination, until we have tried the specific experiment. (1874, 267)

The notion of 'novelty' might then be understood in terms of the unexpectedness of the result of a specific experiment.

Some contemporary philosophers have questioned Mill's apparent restriction of emergence to *nonadditive* properties (Wimsatt 2001; McLaughlin 1992; Kim 1999). For example, in the case of probabilities, multiplication rather than addition is the appropriate operator for deriving the probability of the joint occurrence of independent events. Some philosophers have suggested that virtually *no* constraints be imposed upon the sort of mathematical function by which the value of a system-level property is derived from those of its constituents; so long as it is so derivable, it is not an emergent property (e.g., Kim 1999, 7). But this is too strong a criterion for an emergent property, since a nonlinear function mapping a system's input to its output suggests that the relation between the two is highly dependent upon the specific nature of the interactions that take place between the system components. Hence, if the nature of such interactions cannot be discerned, then the applicability of the function remains completely mysterious. These considerations suggest that the peculiarity that Mill attempted to conceptualize as the failure of "additivity" may be more accurately conceptualized in terms of an essential interaction between the parts of a system, or an interaction that is not discernible given the best available theoretical account of the parts and their relations.

Emergent Evolution

Mill's notion of heteropathic effects gave rise to the doctrine of *emergent evolution*, which gained some currency in the early twentieth century amongst philosophers and biologists such as Lloyd Morgan (Lloyd Morgan 1923), Scott Alexander (1920), and C. D. Broad (1925). According to these emergentists, over a cosmological time scale, certain forms of complexity are successively brought into existence that are novel, in the sense that they represent qualitatively new properties that are not the resultants of the properties that existed previously and are unpredictable in the sense that an exhaustive knowledge of the properties of the previous existents would not allow a prediction or derivation of the new properties before the fact of their appearance. In short, the emergent evolutionists took the two basic characteristics of Mill's heteropathic effects—novelty and unpredictability—and transposed them more explicitly into a cosmological time frame.

However, the emergentists of this period typically did not reject the idea that the appearance of emergent properties is causally *determined* by the lower-level properties (see Alexander 1920, 330; Broad 1925, 67), in the sense that anytime certain constituents *Pi* are arranged in relation *R* under the same conditions *C*, a higher-level property *Q* will be instantiated by the whole. Consequently, according to this classical viewpoint, an emergent property is one that can also be said to supervene upon the properties of its constituents (see Supervenience). However, the idea that emergence embodies a supervenience assumption has in more recent times been placed into question (see final section).

German Organicism

German organicism was an independent philosophical and scientific tradition with distinct roots, which slowly merged with the emergentist tradition during the mid-twentieth century. The organicist tradition often traces its roots to Immanuel Kant's *Critique of the Power of Judgment*. According to Kant, biological forms appear to the mind as inherently purposive rather than as the products of blind mechanism; the reciprocity of part and whole in the organism appeals to the faculty of judgment and provides transcendental justification for the use of teleological reasoning in biology (see especially §65). For Kant, such forms cannot be understood by analyzing the operation of each of its parts in isolation and then inferring the activity of the whole; rather, the existence and operation of each part can be understood only in terms of its contribution to the functioning of the whole. Like emergentism, organicism traces a middle path between mechanism and vitalism.

Within biology, organicism was adopted by such figures as Paul Weiss, Joseph Needham, and Ludwig von Bertalanffy. It found expression, for instance, in Weiss's (1968) concept of the morphogenetic field, as it did in psychology in the notion of the gestalt. The concept of the morphogenetic field was introduced to explain cell differentiation and specialization on the basis of the relative location of whole groups of cells within the organism, rather than on the basis of the intrinsic properties of individual cells and their relations to immediately adjacent cells. Gestalt psychologists such as Köhler held that, for instance, the perceptual patterns that organize the visual field *are* explainable on the basis of the patterns of electrical activity within the visual cortex but that the formation of these electrical patterns is governed by fundamental physical laws that do not involve reference to the individual neurons and their relations. Hence both of these examples satisfy premises (i) and (ii) of emergentism (see the introductory paragraph). (For a brief conceptual overview of the organicist tradition, see Gilbert and Sarkar 2000.) From this perspective it should not be surprising that the partial revival of interest in emergentism in Anglo-American philosophy of science was initiated less by philosophers than by philosophically oriented scientists entrenched in the organicist tradition, such as Michael Polanyi (1968), Weiss (1968), and Roger Sperry (e.g., 1969).

Criticism of Emergence

Criticism of the concept of emergence typically falls under three categories. The first type of criticism holds that "emergent properties" exist, but in some trivial or uninteresting sense, and that whatever philosophical interest it seems to possess is largely a product of conceptual confusion that can be resolved by proper explication of the concept. The second type of criticism (espoused most recently by Kim [1999]) holds that the concept of emergence is conceptually clear but that emergent properties cannot exist for a priori reasons. A third line of criticism, proposed by McLaughlin (1992), holds that emergence is both conceptually interesting and a priori possible but that emergent properties do not in fact exist, and their nonexistence is attested to by the overwhelming historical success of reductionist explanation.

The most important of the critiques that belong to the first category are those of Hempel ([1948] 1965) and Nagel (1961), both of whom argued on the basis of the deductive-nomological model of explanation that the failure of predictability (Hempel) or the failure of reducibility (Nagel) is *logically* trivial and hence does not warrant any important ontological conclusions. According to Nagel (1961, 369), emergence, like reduction, should be seen as a relation between theories, rather than properties. Specifically, it involves a relation between a (suitably axiomatized) "higher-level" theory, T_A (e.g., biology), and a "lower-level" theory, T_B (e.g., chemistry). To say, "P is an emergent property" is to say that the law statements of T_A that contain the predicate P cannot be derived from the law statements of T_B. But clearly, if the predicate P does not appear in T_B—as is often the case in theories at different levels—then nonderivability is a point of logic and is therefore trivial. In order for the derivation to go through, one must specify "bridge laws" or "translation rules" that connect the predicates of T_A with those of T_B via a series of conditionals, where such bridge laws are either lawlike generalizations or definitional stipulations. But once emergence claims are relativized to a given pair of theories and a given set of bridge laws, the very notion of *in principle* irreducibility or unpredictability appears to be incoherent. The most that such nonderivability allows one to infer is the (relatively uninteresting) claim that a given scientific theory is currently unable to explain a certain phenomenon.

However, this criticism is misguided to the extent that it centers upon formal rather than substantive facets of explanation. Nothing in Mill's account, or in that of the later emergentists, prohibits the postulation of bridge laws. Mill's view is that if P is an emergent property, then such bridge laws will find only a purely inductive justification; they will not be derivable in turn from a more fundamental theory, along with the relevant statements of definition. As such, these laws would appear to represent ultimate, inexplicable synthetic facts about the world, and this inexplicability would vitiate the purpose of the putative reduction (Broad 1925, 55). In other words, the emergentist would question the ontological status of the bridge laws themselves, rather than the success or failure of the formal derivation, in evaluating whether P is an emergent property.

Kim's (1999, 32) argument against the possibility of emergence falls under the second type of criticism, in that it seeks to expose an inherent tension between the *novelty* of emergent properties and their supervenient status. According to Kim, in order for a whole system to possess a "novel" property, that property must possess novel causal powers, or powers to bring about changes that cannot be attributed to the *emergence base* of that system, which consists of the nonrelational properties of the parts and the relations of those parts to one another. He also holds that the emergent properties of a system would supervene upon this emergence base, in the sense that the instantiation of the emergence base would determine the instantiation of the emergent property. But, because of this supervenience, it would appear that any putative scientific law L that refers to an emergent property can be replaced by a law L^* that refers not to the emergent property itself but to the emergence base upon which it supervenes. As a consequence, L^* would not lack any explanatory power possessed by L, and hence the reference to the emergent property would be superfluous in a scientific law. One way of countering Kim's argument, then, would be to reject the claim that emergent properties are supervenient, which is what some contemporary advocates of emergentism do, as will be discussed in the next section.

Contemporary Emergentism in the Philosophy of Science

Emergentist approaches in the philosophy of science today are oriented toward the evaluation of reductionist claims that appear within specific scientific contexts (see Primas 1998 for an overview of several problematic claims). Philosophers of science have also drawn upon examples from physics, chemistry, and biology in order to provide new explications of emergence. Three fairly recent explications of emergence are described below.

In the philosophy of physics, one significant area of attention involves nonseparable systems. In quantum mechanics, a nonseparable state is one the state vector of which cannot be represented as a tensor product of component state vectors (see Shimony 1987, which relates nonseparability in quantum mechanics to holism). Hence the behavior of a nonseparable system cannot be explained in terms of the behavior of its components, in addition to a fundamental compositional principle. Additionally, nonseparability in quantum mechanics leads to limitations in the reduction of chemistry to physics, insofar as it entails the use of approximations in the derivation of molecular structure on the basis of quantum mechanics (Woolley 1991; Jaeger and Sarkar 2003).

Some philosophers have taken quantum nonseparability as a model for constructing novel conceptions of emergence. According to one emergentist interpretation (Humphreys 1997), one should not speak of the "components," or "component properties," of a nonseparable system at all; rather one should say that the properties of the constituents have undergone a "fusion" such that they can no longer be meaningfully individuated. Generalizing this example, Humphreys (1997) proposes an abstract fusion operator as a way of explicating the concept of emergence; this, he argues, can also be used to explicate the slippery notion of downward causation. This explication of emergence can also be used to counter Kim (1999), since, in the case of nonseparability, there is no way of independently characterizing the emergence base in terms of the nonrelational properties of the parts of the system and their relations to one another.

A second interpretation (Teller 1986) holds that the components of a nonseparable system *can* be meaningfully individuated but that they stand in "inherent relations" to one another, that is, relations that do not supervene on the nonrelational properties of each part. In contrast to the first interpretation, this conception accepts the independent ontological status of relations; hence Teller (1986) refers to this view as "relational holism."

A third recent approach to elaborating the concept of emergence focuses upon the relation between theories and phenomena, rather than between parts and wholes of systems. Hence it is not, strictly speaking, necessary that an emergentist accept the ontological and epistemological premises ([i] and [ii]) outlined above. Batterman (2002) argues at length that emergence is best understood as involving the failure of a "smooth" or regular asymptotic limiting relation between two theories. As Nickles (1973) shows, in some exemplary cases of theory reduction, one theory or formula is shown to be a special case of another theory or formula when some parameter of the latter approaches a limiting value. An example is classical mechanics, which is a limiting case of the special theory of relativity when velocities approach zero. In some cases, this limiting relation between theories is "singular" rather than smooth or regular; that is, the behavior of the formula becomes highly irregular as the value of a parameter approaches a limit, or the limit itself is undefined. Berry (1994) and Batterman (2002) describe how certain natural phenomena, such as the state of a fluid at its critical point, or the appearance of supernumerary bows of a rainbow, are best described by the irregular behavior of formulae as they approach some limiting value. Supernumerary bows, for example, appear (or "emerge") only when the wavelength of light becomes very small; theoretically, this can be described as a singularity of certain wave-optical formulae as the wavelength parameter approaches zero (that is, as wave optics approximates ray optics). Such phenomena may be called emergent, although they do not appear to involve a special part/whole relation. Additionally, they call into question the putative reduction of ray optics to wave optics (Berry 1994).

Justin Garson

References

Alexander, S. (1920), *Space, Time, and Deity*, vol. II. London: Macmillan.

Batterman, R. W. (2002), *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.

Berry, M. (1994), "Asymptotics, Singularities, and the Reduction of Theories," in D. Prawitz, B. Skyrms, and D. Westerstahl (eds.), *Logic, Methodology and Philosophy of Science IX: Proceedings of the Ninth International Congress of Logic, Methodology and Philosophy of Science, Uppsala 1991*. Amsterdam: Elsevier, North-Holland, 597–607.

Broad, C. D. (1925), *The Mind and Its Place in Nature*. New York: Harcourt, Brace, and Company.

Campbell, D. (1974), "'Downward Causation' in Hierarchically Organised Biological Systems," in F. J. Alaya and T. Dobzhansky (eds.), *Studies in the Philosophy of Biology*. Berkeley and Los Angeles: University of California Press, 179–186.

Gilbert, S., and S. Sarkar (2000), "Embracing Complexity: Organicism for the 21st Century," *Developmental Dynamics* 219: 1–9.

Hempel, C. ([1948] 1965), "Studies in the Logic of Explanation," in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, 245–290.

Humphreys, P. (1997), "How Properties Emerge," *Philosophy of Science* 64: 1–17.

Jaeger, G., and S. Sarkar (2003), "Coherence, Entanglement, and Reductionist Explanation in Quantum Physics," in A. Ashtekar, R. S. Cohen, D. Howard, J. Renn, S. Sarkar, and A. Shimony (eds.), *Revisiting the Foundations of Relativistic Physics: Festschrift in Honor of John Stachel*. Dordrecht, Netherlands: Kluwer, 523–542.

Kim, J. (1999), "Making Sense of Emergence," *Philosophical Studies* 95: 3–36.

Klee, R. (1984), "Micro-Determinism and Concepts of Emergence," *Philosophy of Science* 51: 44–63.

Lewes, G. H. (1875), *Problems of Life and Mind*, vol. II. London: Kegan, Paul, Trench, Trubner, and Co.

Lloyd Morgan, C. (1923), *Emergent Evolution*. London: Williams and Norgate.

McLaughlin, B. (1992), "The Rise and Fall of British Emergentism," in A. Beckermann, H. Flohr, and J. Kim (eds.), *Emergence or Reduction: Essays on the Prospects of Nonreductive Physicalism*. Berlin: Walter de Gruyter, 49–93.

Mill, J. S. (1874), *A System of Logic*, 8th ed. New York: Harper and Brothers.

Nagel, E. (1961), *The Structure of Science*. New York: Harcourt, Brace and World.

Nickles, T. (1973), "Two Concepts of Intertheoretic Reduction," *Journal of Philosophy* 70: 181–201.

Polanyi, M. (1968), "Life's Irreducible Structure," *Science* 160: 1308–1312.

Popper, K., and J. Eccles (1977), *The Self and Its Brain*. Berlin: Springer.

Primas, H. (1998), "Emergence in the Exact Natural Sciences," *Acta Polytechnica Scandinavica* 91: 83–98.

Shimony, A. (1987), "The Methodology of Synthesis: Parts and Wholes in Low-Energy Physics," in R. Kargon and P. Achinstein (eds.), *Kelvin's Baltimore Lectures and Modern Theoretical Physics: Historical and Philosophical Perspectives*. Cambridge, MA: MIT Press, 399–423.

Sperry, R. W. (1969), "A Modified Concept of Consciousness," *Psychological Reviews* 76: 532–536.

Teller, P. (1986), "Relational Holism and Quantum Mechanics," *British Journal for the Philosophy of Science* 37: 71–81.

Weiss, P. (1968), "The Living System: Determinism Stratified," in A. Koestler and J. R. Smythies (eds.), *Beyond Reductionism*. London: Hutchinson, 3–42.

Wimsatt, W. C. (2001), "Emergence as Non-Aggregativity and the Biases of Reductionisms," *Foundations of Science* 5: 269–297.

Woolley, R. G. (1991), "Quantum Chemistry Beyond the Born-Oppenheimer Approximation," *Journal of Molecular Structure* 230: 17–46.

EMPIRICISM

Empiricism is the position according to which experience is the only source of warrant for one's claims about the world. Having assigned experience this exclusive role in justification, empiricists then have a range of views concerning the character of experience, the semantics of claims about unobservable entities, the nature of empirical confirmation, and the possibility of nonempirical warrant for some further class of claims, such as those accepted on the basis of linguistic or logical rules. Given the definitive principle of their position, empiricists can allow that one can have knowledge independent of experience only where what is known is not some objective fact about the world, but something about the way it is conceptualized or described. Some empiricists say that one can have knowledge of verbal equivalences or trivialities; some argue that any nonempirical tenets are not even properly called 'knowledge,' but should be seen as notions accepted on pragmatic rather than properly epistemic grounds. What no empiricist will allow is substantive a priori knowledge: According to empiricism, one can have no pure rational insight into real necessities or the inner structure of nature, but must rely on the deliverances of the senses for all information about external reality. Some versions of empiricism argue against the very notion of real necessities or metaphysical structure behind the phenomena; other versions take a more agnostic approach, arguing that if there is a metaphysical structure behind the phenomena, it is either out of epistemic reach or known only to the extent that it can be grasped through experience, rather than through rational reflection.

Early Modern Background

First published in 1689, John Locke's *Essay Concerning Human Understanding* sets out a version of empiricism whose basic framework remains an inspiration to contemporary advocates of the position. Expressing admiration for the accomplishments of Sir Isaac Newton and Boyle, Locke aims to show a similar respect for observation and theoretical simplicity in his investigation of the powers of the human mind. In contrast to the rationalist project of searching for the essence of the mind or the metaphysical principles behind the way it ought to work, Locke promises to pursue the "historical, plain method" of describing the type of process that would result in the ordinary formation of human knowledge (Locke [1689] 1975, 44). Locke contends that in this sequence of events one begins with a blank slate, a mind empty of ideas. The contrary postulation of innate ideas or principles is incompatible with what is observed in children and the dull-witted, Locke maintains, and in any event superfluous. Human cognition can be explained without helping oneself to the rationalist notion that some truths are built into the mind from the start; the positive task of providing such an explanation becomes the main project of Locke's *Essay*.

Locke maintains that all thought can be analyzed into ideas whose ultimate origin is in experience, broadly conceived to include both sensation (the passive reception of ideas from external objects through the senses) and reflection (the passive reception of ideas from the mind's introspective access to its own workings). Experience provides simple ideas (like those of 'blue,' 'sweet,' or 'pain'); the mind then manipulates and conjoins these simple ideas to form complex ideas (like the ideas of particular individual objects, modes, and relations). Because the mind is able to combine its ideas, the acquisition of knowledge is not restricted to the passive ingestion of ideas in experience; in fact the highest grade of certainty comes from assessing the internal structure of, and the relations among, complex ideas that are one's own construction. A lower degree of certainty accrues to knowledge of the external world, made possible in part by noting that certain ideas reliably come in clusters, which is presumed to indicate the presence of external substances, and also by one's consciousness of one's passivity in receiving ideas of sensation. While it may be the case that certain perceivable qualities necessarily coexist in certain substances (e.g., ductility and weight in gold) in virtue of the microscopic constitution of that substance, one's powers of perception are such that it is impossible to have the same kind of direct knowledge of this necessary coexistence as one has of the perceivable qualities themselves.

In Locke's theory, ideas received from experience are the only ingredients of thought, but many entities other than ideas get postulated during the course of the theory: the external objects causing ideas, powers inherent in those objects and causal relations among them, and the mind itself. Advanced some 50 years later, David Hume's version of empiricism exposes some of the difficulties of attempting to maintain this kind of mixed ontology within the empiricist framework (Hume [1739–1740] 1978). Hume is more careful than Locke to extract evidence for his theory of human cognition only from perceivable phenomena and to refrain from positing the kind of physical and metaphysical entities access to which would be unaccountable from an empiricist perspective. In the first wave of reaction to Locke, George Berkeley had already shown that even the apparently straightforward claim that one's ideas of sensation are caused by external objects could prove difficult for an empiricist to defend: If one were directly conscious only of one's ideas, with what right could one claim that these ideas resemble, and have their origin in, things of an entirely different kind that are not themselves directly present to the mind? Berkeley argues for a phenomenalist understanding of objects: The objects of which one is conscious are not independent matter but in fact collections of perceptions. Hume agrees with Berkeley that given the empiricist premise that one is only aware of one's perceptions, the postulation of independent matter is unjustifiable, but Hume also notes that people have a tendency to conceive of objects as having an independence and continuity that is not ascribed to perceptions. From the perspective of a consistent empiricist, Hume suggests that this tendency can be seen only as a blind instinct toward fabrication: Experience never delivers anything other than fleeting perceptions, so the sense of permanence is nothing more than an illusion, which Hume explains by pointing to the near resemblance of successive perceptions and the ease with which people can confuse resembling particulars for the same thing.

Causation receives a similar treatment: Where Locke had helped himself to a realist understanding of causation, Hume points out that causation is not itself perceived and cannot be construed as a pure relation of ideas. That no purely conceptual connection links a cause to an effect can be seen by reflecting on the ability one has to imagine a change in the course of nature. Like the stability of external objects, objective necessary connections among objects are an illusion, generated in this case by one's consciousness of one's instinctive (as opposed to rational) habit of expecting past patterns to continue. Where one has seen many events of type A followed by events of type B, one develops a mental custom of associating these ideas, and with this custom in place, the sight of type A compels the mind to think of B. The subjective sense of being pushed in this way gives rise to the idea of necessary connection, which then mistakenly projects onto nature and imagines it to be an objective relation among events.

If Hume's analysis is aimed at showing that such fundamental components of the commonsense worldview as enduring external objects and causation are illusory, he does not suggest that this philosophical result will overthrow that worldview; indeed, he argues that observation of the natural tendencies of the human mind shows that people will naturally continue in their instinctive patterns of thinking in terms of objective things and causes, however unjustified these instincts may seem from a philosophical standpoint. It is a difficult interpretive question to what extent this skeptical outcome should be read as a philosophical condemnation of ordinary claims to knowledge or as a demonstration of the shortcomings of either philosophical analysis in general or empiricism in particular.

One influential response to Hume was to see his skepticism as pointing to the inadequacy of the empiricist starting point. Immanuel Kant argued that one's thought about matters such as causation could not be understood without the postulation of something more than mere sensory perceptions as available to the mind; he maintained that one can make sense of empirical knowledge only if one sees sensory perceptions as entering a mind already possessed of a priori knowledge of the underlying causal structure of nature and the geometrical form of time and space. The exact nature and status of the metaphysical and geometrical commitments Kant envisaged is a matter of some controversy. For later advocates of the broadly Kantian style of response to empiricism (e.g., Reichenbach 1965), the complexity of the task of articulating a reasonable such set of a priori constraints was made particularly evident by such developments as the emergence of the theory of relativity (see Causality).

Early-Twentieth-Century Background

Until the twentieth century, geometry, or the study of the pure structure of space, had typically been seen as the paradigmatic example of an a priori discipline, and as an obstacle for empiricist accounts of knowledge. Einstein's use of non-Euclidean geometry in the theory of relativity made it hard to resist the conclusion that if geometry was a priori at all, it had this status only when considered as an uninterpreted deductive enterprise: The study of the structure of space itself could now be taken as either an empirical matter or a matter of the postulation of conventions rather than the discovery of objective facts. The reexamination of the status of questions once considered intuitive or rational was a significant source of inspiration for logical positivism, originating in Germany and Austria in the 1920s. Positivism drew inspiration from the development of Frege and Russell's symbolic logic and the new clarity it brought to the problem of the foundations of mathematics; at the same time, the legacy of Ernest Mach's eliminative empiricism was also a powerful if short-lived force behind the movement.

The relation between positivism and empiricism is a complex matter. It is clear that the positivists thought that all substantive questions about the world were to be answered by empirical science, but it is less clear that their conception of empirical science was straightforwardly empiricist. Some examination of the details of positivism is in order here.

The positivists conceived of philosophy as an enterprise of clarifying and making explicit the conceptual, linguistic, and logical structures of science, rather than as a means of discovering further characteristics of reality at a deeper metaphysical level than the empirical phenomena. The positivists hoped for a clean divide between the material questions about nature to be answered by the empirical sciences and the formal questions about science to be answered by philosophy. Given this formal approach, it is not surprising that the positivists cast the central problems of epistemology in linguistic terms. Locke's causal picture of sensation saddled him with a metaphysics not easily defended from within empiricism; the positivists aimed to avoid metaphysics altogether and take the question of the relation between experience and theory as a question about the proper form of observation reports and their formal relations to other sentences in the language of science. So Moritz Schlick writes in "The Foundation of Knowledge": "I think it a great improvement in method to try to aim at the basis of knowledge by looking not for the primary *facts* but for the primary *sentences*" (Schlick 1959, 212). These primary or protocol sentences are seen as idealized records of basic experience, cast in a vocabulary of observational terms and separated sharply from the higher-level theoretical claims whose confirmation they supply. Positivists divided into several factions over the question of the form of these statements. On Schlick's "foundationalist" side of the debate, a protocol statement aims to capture the content of what Schlick called a "confirmation," or decisive moment, of experience, whose certainty is beyond doubt; other parts of the system of science are ultimately justified by their relations to these confirmations, but the confirmations themselves are justified by the character of experience itself, and not by anything further within the system of science (see Schlick, Moritz). In opposition to Schlick, Otto Neurath proposed a fallibilist approach to protocol statements: A protocol statement is, like any other statement in the system of science, subject to rejection in light of considerations of overall coherence (see Neurath, Otto). Schlick has difficulty explaining the relation between basic confirmations and their linguistic expressions in protocol statements without recourse to metaphysics. Neurath has difficulty explaining how his solution maintains a special role for experience, or how he maintains empiricism without leaving himself open to the charge that science and fantasy could be equally well grounded given just sufficient internal consistency.

While Rudolf Carnap's original position was closer to Schlick's, he soon moved to adopt what he took to be a neutral stance, declaring that the question of the form of protocol sentences is "not answered by assertions but rather by postulations. . . . [T]he task consists in investigating the consequences of these various possible postulations and in testing their practical utility" (Carnap [1932] 1987, 458). Rather than supposing that something in the nature of reality determines the correct syntactical form and role of observation statements in science, Carnap now maintained that this is not a question of fact with a single correct answer, but a question about which postulation will be found most convenient for one's purposes. Carnap's work went on to exhibit an increasing emphasis on conventions adopted for pragmatic reasons.

The exact extent of Carnap's allegiance to empiricism is subject to debate (see Friedman 1999; Sarkar 2001). Carnap does not start from the position that the justification of empirical science is in doubt until science can be shown to be derived from the contents of experience, nor does he think that the immediately given has a specially certain or unproblematic epistemic status. In *The Logical Structure of the World*, Carnap tries to show how scientific concepts could be reduced to relations among moments of experience, but he claims that he could have taken other basic elements, like space-time points or even physical entities such as subatomic particles, as his starting point: His aim is strictly to analyze the internal logical structure of science rather than to justify science by appeal to something better grounded. By his own admission, Carnap's analysis of the internal logical structure of science was incomplete, most crucially in its failure to exhibit the dispensability of the basic relation of recollected similarity. Carnap's work in the decade after *Logical Structure* shows further departures from the verificationist empiricism of early positivism. While early positivists had claimed that every scientific term could be explicitly defined in terms of observable properties, in his "Testability and Meaning" (1936 and 1937) Carnap argues that some theoretical terms have a less direct relation with observation. Because dispositional concepts such as 'solubility,' for example, need to be understood in terms of relationships among various possible test conditions and observable outcomes, sentences involving terms of this sort cannot just be translated into sentences using the original observational vocabulary (see Carnap, Rudolf).

A form of positivism that lies squarely in the empiricist tradition is presented in A. J. Ayer's (1936) *Language, Truth and Logic*. Ayer insists on a phenomenalist account of external objects and a verificationist theory of meaning. According to Ayer, only two kinds of statements have literal significance and the possibility of truth or falsity: synthetic statements, identified as those statements that can be rendered more or less probable by some specifiable course of experience, and analytic statements, whose acceptability is wholly determined by the syntactic rules for the symbols they contain. All other statements, and in particular the statements of traditional metaphysics, are not even false, but meaningless. Philosophy itself is seen as falling on the analytic side of the line: Epistemology is concerned with the rules governing the use of symbols, and aims to identify the formal relations between the various strings of symbols that constitute observational and theoretical statements in the language of science (see Ayer, Alfred Jules).

Ayer's version of empiricism was one of the first targets of a wave of arguments that led to the decline of positivism by the middle of the twentieth century. Phenomenalism was attacked as incoherent (see Chisholm 1948); Nelson Goodman ([1954] 1979) argued that confirmation could not be explained in syntactic terms; Wilfred Sellars (1956) urged that empiricism's view of what is given in experience made experience an inadequate basis for knowledge of the world; and Willard Van Quine (1953) argued that the positivists had no acceptable way of drawing their distinction between analytic and synthetic statements. Quine's criticism proved particularly influential in the subsequent development of empiricism.

Empiricism after Positivism

Quine's "Two Dogmas of Empiricism" attacked both the positivist notion of a sharp distinction between analytic and synthetic sentences and what Quine calls the doctrine of reductionism, according to which each synthetic sentence is associated with a fixed set of actual or possible experiences tending to confirm or discredit that sentence. On the first front, Quine argues that various positivist efforts to identify the distinctive features of analytic sentences have either been inadequate to distinguish the set of sentences the positivists needed to take as analytic or slipped into an empty circularity, in which, for example, analyticity is understood with the help of the notion of cognitive synonymy, and cognitive synonymy is either left unexplained or itself defined in terms of what is analytically true. On the question of reductionism, Quine finds a lesson in Carnap's failure to reduce individual statements about the physical world to statements about immediate experience, and recalls Duhem's claim that one is always free to maintain a theory in the face of apparently contrary evidence by amending an auxiliary hypothesis. According to the slogan that has become known as the Quine-Duhem thesis, "Our statements about the external world face the tribunal of sense experience not individually but only as a corporate body" (Quine 1953, 41) (see Duhem Thesis; Quine, Willard Van; Underdetermination of Theories).

Quine intended his essay strictly as an attack on the positivist version of empiricism, and not on empiricism itself. In the final section, "Empiricism Without the Dogmas," experience is clearly identified as the only source of information for theories about the world, but the relation between experience and theory is not as the positivists had thought. Beliefs about everything from general physical laws to mundane claims about particular objects form a single system, the parts of which are amended in response to recalcitrant experience and kept in line with each other in accordance with rules of logic that are themselves part of the web. Nothing is immune to revision, and everything is revised on the same basis of accommodating experience, so that there is no difference in principle between changing a logical law to simplify quantum mechanics and changing from a geocentric to a heliocentric cosmology, or revising any other empirical claim. In place of the formal positivist approach to confirmation, Quine introduces a relation of 'germaneness' in his account of the relation between sensory evidence and the theory it supports. A body of sensory experience is more germane to one claim than to others when this experience will leave one more likely in practice to revise this particular claim. Rather than engaging in the study of how an ideal scientific language would be formulated, or how one ought to reform one's thinking, the epistemologist is directed to engage in an empirical study of the relationship between the actual input of sensory stimulation and the output of theoretical utterances. Following through with this program would require epistemology ultimately to become a chapter of psychology.

Quine insisted throughout his career that this naturalist position counted as a form of empiricism, but this classification is controversial. Indeed, Donald Davidson (1973–1974) argues that a natural extension of Quine's argument will do away with the contrast between form and content and leave nothing recognizable as empiricism. Also, while Quine contends that there is a normative element in his position, insofar as it leaves room for people to be criticized for having beliefs that accommodate their sensory experience poorly, it is clear that Quine's naturalism does not have the same normative ambition of traditional empiricism. Traditional empiricism was concerned with the question of what one ought to believe, or how common ways of thinking might be reformed to respect the limits of warrant; Quine's naturalism aims to take cognition as a given object of empirical inquiry, and does away with the traditional conception of warrant (see Hookway 1994). For Quine, the question is always about what sentences people *do* revise in practice, and not about what sentences *would be right* to revise, whether people actually do so or not. Whether Quine is an "empiricist" will depend in part on how one wants to use the term. If one emphasizes Quine's advocacy of empirical methods for the study of knowledge itself, then it may seem appropriate to classify his epistemological naturalism as a development continuous with the main thrust of empiricism; indeed, Quine is sometimes faulted for not having gone far enough in using the empirical data he recommends as useful in epistemology. On the other hand, if one sees epistemology as an enterprise that is aimed at figuring out what justifies beliefs, then it is hard not to see Quine's naturalism as constituting a change of topic rather than a development of earlier empiricism.

The version of empiricism that constitutes the most influential contribution to traditional epistemology since the collapse of positivism has been put forward by Bas van Fraassen, in support of the view of science he calls "constructive empiricism." According to van Fraassen, the positivists were mistaken in assuming that once empiricists take experience as the sole source of warrant, they are required to reduce everything to experience or to reinterpret statements about unobservable entities as abbreviations for more complex statements about observable phenomena. Empiricism does set limits on what one can see oneself as rationally obliged to believe, but by invoking a distinction between acceptance and belief, van Fraassen is able to defend an empiricist approach to science without requiring a positivist reformulation of the language of theories. When one accepts a theory, and commits oneself to a certain research program, one must believe what the theory says about observables, that is, one must believe that the theory is empirically adequate; but one does not have to believe the whole theory, including what it says about unobservables. Allowing this agnosticism about the unobservable makes accepting less committal than believing, but van Fraassen argues that science can be understood without the stronger realist stance; nothing that matters is lost by seeing science as aiming just at empirical adequacy, rather than full-blown truth. Equally, nothing is gained by the stronger realist position if van Fraassen is right, other than the need to contend with, and explain epistemic access to, various items of metaphysical baggage like causes and laws, realistically construed.

Van Fraassen allows that theories may have virtues that go beyond empirical adequacy (perhaps simplicity or explanatory power), but such informative virtues do not make the theory more likely to be true. Indeed, the more informative a theory is, the more risk it runs of being false; if one chooses informative theories over their less committal counterparts, it can only be for pragmatic reasons and not because these theories are more likely to be true. In van Fraassen's empiricism, scientists need never accept ampliative rules of inference (like inference to the best explanation [IBE]) as forcing them to go beyond the limits of observation: If positing the real existence of electrons would explain some observable phenomenon, this is not in itself a reason to take the step of believing that the unobservable electrons exist. Respecting the limits of his warrant, scientists may rationally stick to the more modest position that all observable phenomena are *as they would be if* the electron theory were true.

Van Fraassen shares with the positivists a sense of the epistemic significance of the line between what is observable and what is not, but instead of aiming to find a syntactical way of drawing the line, say, by developing a purely observational vocabulary, he argues that the problem can be naturalized: Scientific theories themselves can show how the realm of the observable is delimited. According to constructive empiricism, a scientific theory shows a picture of how the world could be, giving a set of models corresponding to various initial conditions. The theory itself can then specify parts of these models (the "empirical substructures") as potentially representing observable phenomena. A theory is empirically adequate if it has a model in which the observable phenomena can be embedded.

Van Fraassen himself notes that while belief in a theory's empirical adequacy is weaker and therefore safer than belief in its truth, it is not without risk: In claiming that a theory is empirically adequate, one is still going out on a limb and committing oneself to the truth of claims about states of affairs that are not observed by oneself, or have not yet been observed, or will never actually be observed, and so on. If one's motivation were just to maintain the weakest possible beliefs compatible with the evidence, one should shrink in the direction of a solipsism of the present moment rather than adopting the scientific rationality of constructive empiricism. So van Fraassen's position does not enable one to be maximally certain of one's beliefs. He has argued that his aim is rather to develop a characterization of the aim of science, or the standards for what counts as success or failure in that enterprise; if scientists do not restrict admissible evidence to what is observed by themselves alone, then no adequate account of science can give supreme epistemic significance to that special class of evidence.

This is not to suggest that van Fraassen sees his constructive empiricism as a sociological summary of the attitudes of working scientists. In particular, van Fraassen is ready to acknowledge that scientists may often believe that their theories are not merely empirically adequate but true, even with respect to unobservables. Because of the way van Fraassen defines rationality, he does not have to classify such thinking as irrational: His conception of rationality is permissive, rather than prescriptive. In this view, the scientist does not need to be rationally compelled to believe something in order for the scientist's belief to count as rational; rather, she may believe anything as long as she is not rationally compelled to believe otherwise. Rationality requires one to maintain logical consistency and accept the testimony of the senses, but if such minimal limits are respected, it neither requires nor forbids one from making conjectures about what lies beyond sensory evidence. In this view, then, the main upshot of an empiricist conception of rationality is negative: If warrant comes only from experience, rationality can never require belief in entities and characteristics of reality to which one lacks empirical access.

Criticisms of Constructive Empiricism

The most direct way to attack van Fraassen's empiricist view of science would be to identify a properly epistemic (as opposed to merely pragmatic) reason to believe in the claims that science makes about entities that lie below the threshold of observation. Many critics of van Fraassen have attempted to defend the rationality (as opposed to the mere practical convenience) of abduction or IBE. The best-known move here is Hilary Putnam and Richard Boyd's "no miracle argument" (NMA), according to which it is only by taking scientific theories to be true or approximately true that the success of science will be anything other than miraculous. It would be a tremendously strange coincidence, they argue, if all observable phenomena were just as though quarks existed and yet in fact they did not exist. This argument would have more force against an eliminative empiricist who would actually forbid belief in the unobservable. Against van Fraassen, the realists need to establish not just that belief in quarks is rationally permissible (he already grants this) but that it is rationally required. The main difficulty the NMA faces in establishing that conclusion is that it appears to be an argument with the very same abductive form as is in question (see Fine 1991). The argument urges that the truth of scientific theories is the best explanation for the phenomenon of their success; but even if that is so, unless one is already convinced that one is entitled to infer that whatever is the best explanation of a phenomenon is for that reason likely to be true, then one has no reason to accept the realist conclusion (see Putnam, Hilary).

A number of empiricist positions are intended to suggest that sound arguments in support of IBE are unlikely to be forthcoming. According to the "pessimistic induction," it is a mistake to infer the truth of a scientific theory from its acceptability as an explanation of the known phenomena, because of the many historical examples of theories that were explanatory successes in their day but have since been shown to be false. From the past course of events, there is no reason to believe that the theories now found persuasive as explanations of the phenomena are in fact true descriptions of things seen and unseen. In response to this argument, realists have noted that doubts about whether a current theory is exactly right may not provide a reason to withhold belief in the entities posited by that theory. Many theories that are shown to be false are superseded by theories that continue to use the same basic framework of entities, although there is some question about whether the realist can present a historical argument about the reasons for past predictive successes without presupposing the legitimacy of abduction (for a detailed historical discussion, see Psillos 2000). In addition, there is a more abstract and general form of the pessimistic induction available to the empiricist. According to the *argument from the bad lot*, the label "inference to the best explanation" is misleading, because there is no guarantee that one is in a position to choose the *best* explanation: One's choice is from among the explanations scientists have in fact been able to concoct so far, a range of alternatives that might in fact fail to include the true story. One can at most think of oneself as choosing the best available story, rationally weighing various rival theories only on the basis of evidence about observable phenomena.

The "conjunction objection" to constructive empiricism constitutes a quite independent move (Boyd 1973; Putnam 1978; Friedman 1983). It may be correct that in terms of vulnerability to recalcitrant evidence, a single theory's truth is never more credible than its empirical adequacy, but by taking theories to be true, one logically has the right to conjoin them, and the conjoined theory ($T_1 + T_2$) can have richer empirical consequences, which can give additional confirmation to its each of its conjuncts T_1 and T_2 taken separately. In addition, the larger unified theory can give the kind of integrated explanation of phenomena at which science (arguably) must aim. Meanwhile, accepting that two theories are empirically adequate does not automatically give one the right to conjoin them (they may, for example, include contradictory statements about unobservables), and even where they can be conjoined, the claim that "T_1 is empirically adequate and T_2 is empirically adequate" will have fewer observational consequences than ($T_1 + T_2$). It is open to the empiricist to challenge the realist idea that science aims at such unified explanations rather than unifying, where it does, as a pure consequence of the search for empirical adequacy; it is also possible to challenge the extent to which science does in fact engage in this kind of unification, or whether in fact later theories are used to correct earlier ones, rather than being straightforwardly conjoined with them (see van Fraassen 1980, Ch. 4).

Other points in the empiricist program that have attracted critical attention include the issue of modal concepts of possibility and necessity, even as they figure in van Fraassen's own statement of his position (Rosen 1994; Ladyman 2000), and the question of whether empiricism can give an adequate characterization of experience (Nagel 2000).

In raising doubts about whether the truth might always lie outside of the range of theories available, van Fraassen is sometimes seen as risking a collapse into skepticism. If warrant is so restricted that one can never have rational grounds to believe in any unobservable entity, no matter how well it would explain observations, then it may seem that by similar reasoning one will never be rationally compelled to believe anything as strong as the empirical adequacy of a theory, or even anything at all beyond the present testimony of the senses. Conversely, if van Fraassen wants to support the rationality of believing that certain theories are empirically adequate (true in all they say about the observable, and not just about what is presently observed), or even that perceived objects continue to exist after one leaves the room, then perhaps he is already committed to the admissibility of ampliative rational rules. Against the idea that continuously existing tables and trees are posited as the best explanation of given sense data, van Fraassen (1989) argues that philosophers have given ample arguments to show that one's awareness of the world cannot be a matter of making inferences from a body of raw sense data. What are perceived are not sense data but the observable parts of an objective world: "[W]e can and do see the truth about many things: ourselves, trees and animals, clouds and rivers—in the immediacy of experience"(178). Experience itself can be understood only "in the framework of observable phenomena ordinarily recognized" (1980, 72). This marks a reversal from the earlier empiricist strategy of attempting to show how the framework of observable phenomena could be constructed out of the ideas of experience.

In this version, empiricism is insulated from skepticism by setting its focus on the manner in which beliefs are updated, and not on their initial formation. According to van Fraassen (1989), "It is possible to remain an empiricist without sliding into skepticism, exactly by rejecting the skeptics' pious demands for justification where none is to be had" (178). Once one is committed to the general framework of observable phenomena, one will be in a position to examine critically the ways in which beliefs are changed, but there is no useful prospect of a critical examination of one's initial commitments. Critics of empiricism can wonder whether this pessimism about the scope of epistemology is justified and whether van Fraassen is right to characterize people's initial position as involving no commitments other than those to observables. It has also been suggested that what is in dispute between empiricism and realism may not be decidable on the basis of considerations acceptable to both sides, and this has generated some skepticism about the legitimacy of this conflict.

Skepticism About Empiricism

Both the empiricist and the realist are committed to the project of giving a philosophical analysis of the aim of science. But Arthur Fine argues that there is something wrong with that project. According to Fine (1986), realists and empiricists are mistaken in supposing that science has a single essence amenable to philosophical examination. There is nothing in scientific practice itself, Fine argues, that requires the possession of a philosophical theory of the point of science, and nothing in the deliverances of scientific inquiry yields an answer to whether empiricism or realism is correct. As an alternative, Fine advocates what he calls the natural ontological attitude, according to which one allows science to "speak for itself" and refrains from attempting to construct a notion of truth that goes beyond that "already in use in science." Of course, both realists and empiricists take themselves to be articulating exactly that conception of truth that is already in use in science. But Fine's contention is that they do not have any neutral or unprejudiced perspective from which to pass judgment on what science involves.

One of Fine's central criticisms of empiricism is that the empiricist's effort to create a special epistemic status for claims about observables could be based only on a priori commitments that do not square well with the basic orientation of empiricism. Observations alone do not force upon one any particular epistemic attitude to observation. If Fine is right about that, then the empiricist has some reason to resist the naturalist's suggestion that the claims advanced in epistemology are, like the claims of empirical science, themselves warranted only by experience (see van Fraassen 1995 for an argument along these lines). Empiricism is then a theory about what claims are warranted within science; the separate question of what claims are warranted within epistemology would lie beyond the scope of empiricism itself.

JENNIFER NAGEL

The author acknowledges the helpful input of Anjan Chakravartty, University of Toronto.

References

Ayer, A. J. (1936), *Language, Truth and Logic*. London: Gollancz, 1936.

Boyd, Richard (1973), "Realism, Underdetermination, and a Causal Theory of Evidence," *Noûs* 7: 1–12.

——— (1984), "The Current Status of Scientific Realism," in Jarrett Leplin (ed.), *Scientific Realism*. Berkeley and Los Angeles: University of California Press, 41–82.

Carnap, Rudolf ([1928] 1967), *Der logische Aufbau der Welt*. Translated by Rolf George as *The Logical Structure of the World*. Berkeley and Los Angeles: University of California Press.

——— ([1932] 1987), "On Protocol Sentences" (translated by R. Creath and R. Nollan), *Noûs* 21 (1987): 457–470. Originally published as "Über Protokollsätze," *Erkenntnis* 3: 107–142.

——— (1936),"Testability and Meaning," *Philosophy of Science* 3: 419–471.

——— (1937),"Testability and Meaning—Continued," *Philosophy of Science* 4: 1–40.

Chisholm, Roderick (1948), "The Problem of Empiricism," *Journal of Philosophy* 45: 512–517.

Davidson, Donald (1973–4), "On the Very Idea of a Conceptual Scheme," *Proceedings and Addresses of the American Philosophical Association* 67: 5–20.

Fine, Arthur (1986), "Unnatural Attitudes: Realist and Instrumentalist Attachments to Science," *Mind* 95: 149–179.

——— (1991), "Piecemeal Realism," *Philosophical Studies* 61: 79–96.

Friedman, Michael (1999), *Reconsidering Logical Positivism*. New York: Cambridge University Press.

——— (1983), *Foundations of Space-Time Theories*. Princeton, NJ: Princeton University Press.

Goodman, Nelson ([1954] 1979), *Fact, Fiction and Forecast*, 4th ed. Cambridge, MA: Harvard University Press.

Hookway, Christopher (1994), "Naturalized Epistemology and Epistemic Evaluation," *Inquiry* 37: 465–485.

Hume, David ([1739–40] 1978), *A Treatise of Human Nature*, 2nd ed. Edited by L. A. Selby-Bigge, revised P. H. Nidditch. Oxford: Clarendon.

Ladyman, James (2000), "What's Really Wrong with Constructive Empiricism: Van Fraassen and the Metaphysics of Modality," *British Journal for the Philosophy of Science* 51: 837–856.

Locke, John ([1689] 1975), *An Essay Concerning Human Understanding*. Edited by P.H. Nidditch. Oxford: Clarendon.

Nagel, Jennifer (2000), "The Empiricist Conception of Experience," *Philosophy* 75: 345–376.

Psillos, Stathis (1999), *Scientific Realism: How Science Tracks Truth*. New York: Routledge.

Putnam, Hilary (1978), *Meaning and the Moral Sciences*. Boston: Routledge and Kegan Paul.

Quine, Willard Van (1953), "Two Dogmas of Empiricism," in *From a Logical Point of View*, 2nd ed. Cambridge, MA: Harvard University Press.

Reichenbach, Hans (1965), *The Theory of Relativity and A Priori Knowledge*. Berkeley and Los Angeles: University of California Press.

Rosen, Gideon (1994), "What Is Constructive Empiricism?" *Philosophical Studies* 74: 143–178.

Sarkar, Sahotra (2001), "Rudolf Carnap," in *A Companion to Analytic Philosophy*. London: Blackwell, 94–109.

Schlick, Moritz (1959), "The Foundation of Knowledge," in A.J. Ayer (ed.), *Logical Positivism*. New York: Free Press, 209–227.

Sellars, Wilfrid (1956), "Empiricism and the Philosophy of Mind," in Herbert Feigl and Michael Scriven (eds), *Minnesota Studies in the Philosophy of Science*, vol. 1. Minneapolis: University of Minnesota Press, 253–329.

van Fraassen, Bas (1980), *The Scientific Image*. Oxford: Clarendon Press.

——— (1985), "Empiricism in the Philosophy of Science," in P. M. Churchland and C. A. Hooker (eds.), *Images of Science: Essays on Realism and Empiricism*. Chicago: University of Chicago Press.

——— (1989), *Laws and Symmetry*. Oxford: Clarendon.

——— (1995), "Against Naturalized Epistemology," in Paolo Leonardi and Marco Santambrogio (eds.), *On Quine: New Essays*. Cambridge: Cambridge University Press.

See also **Epistemology; Instrumentalism; Logical Empiricism; Realism**

EPISTEMOLOGY

Epistemology is often identified as the theory of knowledge, and epistemologists have often taken this to mean the giving of an analysis of the concept of knowledge. This is motivated, in large part, by skeptical worries about the possibility of knowledge. The project of answering the skeptic has once again emerged as an important theme in epistemology, largely due to the impact of contextualist analyses of knowledge (Cohen 1988; DeRose 1995; Lewis 1996).

While analyses of knowledge are certainly a major theme of much of contemporary epistemology, they by no means exhaust its central concerns. Of particular relevance for the philosophy of science are high-level theories about the general structure of epistemic justification and theories of epistemic change (belief revision or theory change).

The Gettier Problem

Conventional wisdom, pre-1963, had it that the concept of knowledge admitted of a straightforward analysis: A subject *S* knows that *p* iff (“if and only if”) *S* believes that *p, S* is justified in so believing, and *p* is true. This justified true belief (JTB) analysis, however, is far from adequate. Gettier (1963) offered alarmingly simple counterexamples to the JTB analysis, establishing that justified true belief cannot be sufficient for knowledge. Consider the case of Jones. Jones has impeccable evidence that Smith owns a Ford—he sees Smith driving a Ford, Smith is always talking about his great Ford, and so on. So Jones justifiedly believes that Smith owns a Ford. Jones has also recently taken a logic class and has done very well. He realizes that “Smith owns a Ford” entails that Smith owns a Ford or Brown is in Barcelona. So he adopts this latter, disjunctive belief on the basis of his belief about Smith and the Ford. It seems as though Jones is justified in this disjunctive belief if he is justified in believing that Smith owns a Ford. Now, unbeknownst to Jones, Smith actually does *not* own a Ford, but by sheer coincidence Brown *is* in Barcelona. Hence, Jones has a justified true belief—namely, that Smith owns a Ford or Brown is in Barcelona—that does not seem to count as knowledge. His justification for the belief is not aligned in the “right way” with the facts that make the content of the belief true. The Gettier problem is that of giving an analysis of knowledge (or an analysis of the locution “*S* knows that *p*”) that makes clear what the right way is.

An immediate thought is that what has gone wrong in the case of Jones is that he is reasoning from a false belief. The natural suggestion, then, is to add to the JTB account a fourth condition to the effect that *S* knows that *p* only if *S*’s justification for *p* does not rely on any false beliefs. This simple fix, however, is not adequate, and this fact was apparently known to Bertrand Russell much earlier (see Russell, Bertrand). Adapting an example from Russell (1912), suppose that Jones walks through the town square every morning on his way to work. Every morning Jones looks at the clock tower, forms the relevant belief about the current time, and adjusts his pace to make sure he stays on schedule. The clock is a remarkably reliable timepiece, as everyone in town knows. Now, on a particular morning, Jones walks through the square, sees that the clock reads 8:15, and forms the belief that it is 8:15 A.M. In fact, it *is* 8:15 A.M., but unbeknownst to Jones, the clock has stopped! There was an electrical storm the night before, and lightning struck the clock precisely at 8:15 P.M. Jones seems to be justified in his belief that it is 8:15 A.M. and his belief is true, yet it does not seem that Jones knows that it is 8:15 A.M. His belief that it is 8:15 A.M. is not (or not obviously) inferred from any false belief, nevertheless this does not count as knowledge, since his justification for the belief does not connect up with its truth in the so-called right way.

The Gettier problem is seductive for its simplicity, but devilishly hard to solve. Gettier-type counterexamples tend to proliferate, so that given a particular candidate theory constructed with an eye to avoiding the original cases, it succumbs to a close variant. And patched to handle the variant, the theory typically falls prey to a variant of the variant. (See Plantinga 1993 for a detailed survey of attempts at providing a fourth condition to the JTB analysis.) A related project attempts to solve the Gettier problem (and also skeptical worries) by analyses of the locution “*S* knows that *p*” in which justification plays no essential role. So-called tracking accounts of knowledge (Nozick 1981), certain forms of contextualism (Lewis 1996), and what might be called "neighborhood reliabilism" (Williamson 2000) fit more or less uncontroversially into this broad category.

Theories of Justification

Theories of epistemic justification have traditionally been at the center of work in epistemology. This is largely because of the double duty that justification is asked to play. On the one hand, it is thought by many to be a necessary condition of knowledge, and many theories of justification aim at answering the Gettier problem. But "justification" is also a term of epistemic appraisal, signifying the positive epistemic status of a belief, and so the contours of this concept are of independent epistemological interest. Pretheoretically and from an epistemic point of view, justified beliefs are those that it is permissible to hold. Theories of epistemic justification aim at codifying and systematizing this rough-and-ready intuitive gloss by specifying the structure and nature of justification—delivering predictions about how exactly a belief is justified. The class of theories of justification can be neatly categorized in a number of ways. One way is by a series of distinctions (Pollock 1979; Pollock and Cruz 1999). The first distinction is between doxastic and nondoxastic theories. Justification of a particular belief is at least in part a function of what else an agent believes. Doxastic theories insist that in addition, justification is only a function of the agent's beliefs. Nondoxastic theories deny this. The second distinction is the internalist–externalist divide. This is a difference over the sorts of things that can be justifiers of beliefs. Internalist theories insist that justifiers must be "internal states," in some sense internally accessible to the cognitive agent in question. Externalist theories deny this. All doxastic theories are internalist, but not all internal states are doxastic, since it is possible for such states to be nevertheless relevant to the justification of a belief. Alternatively, one can describe the space of theories of justification as being carved according to the sources of the justifiers and the structure that those justifiers must have when related to other beliefs (Alston 1986). That is, the space can be carved along the dimensions of the internalist–externalist divide (the source of justifiers) and the structure of the justification relationship.

Foundationalism

Doxastic theories start from the assumption that whether an agent *S*'s belief is justified is a function of, and only a function of, what else *S* believes—*S*'s doxastic state at the time. After all, an agent's information about the world (at least the information about the world that is relevant to the concerns of epistemic justification) is codified by the beliefs that the agent has about it. Now, the project of a theory of justification is to say what beliefs one ought to hold. But then what information could be relevant, aside from an agent's beliefs, in deciding what one should believe? The assumption that nothing else could be relevant is just the assumption that epistemic justification is a function exclusively of one's doxastic state.

Foundationalist theories start from the intuition that some beliefs serve as reasons for others and that justification is a function of this "reason-for" relation. Now, belief in candidate belief *p* is justified only if the beliefs upon which *p* depends (in *S*'s doxastic state) are also justified. But, the foundationalist argues, this tracing of reasons must either come to an end or continue ad infinitum (perhaps by running in a circle). Such infinite chains would be of no help in developing a theory of justification for finite beings (or else would be viciously circular), so the tracing of reasons must come to an end with beliefs that are self-justifying, or *basic*. The foundationalist maintains that there is a proper subset of an agent's beliefs that are basic in this way, epistemologically privileged and standing in no need of (exogenous) justification. There is a recursive structure to the nature of epistemic justification, with the basic beliefs laying the foundation: *S*'s belief that *p* is justified iff *p* is a basic belief, or else it is properly based on or legitimately inferred from other justified beliefs. This schema can be turned into a full-blown theory once the foundationalist gives an account of what beliefs are basic (Are there enough of them? In what sense are they self-justifying?) and what sorts of relations count as "proper" inference relations between beliefs.

The foundationalist theories of the early twentieth century foundered on the nature of the reason-for relation. It was thought that the only good reasons were either deductive or inductive. But if that is right, then it seems impossible to give a foundationalist account of justified perceptual beliefs. Suppose *S* enters a room, sees an object that appears red, and forms the belief that there is a red object. This belief is intuitively justified but does not seem to be entailed by her perceptual information, the agent's other beliefs, or any combination of the two. Attempts at analyzing perceptual content in such a way as to guarantee such an entailment while staying within the doxastic framework (the phenomenalist project of [the early] Rudolf Carnap, [the early] Willard Van Quine, Goodman, and others) met with insuperable difficulties and were abandoned (see Carnap, Rudolf; Quine, Willard Van Orman). Similarly, *S*'s belief cannot, in general, be justified by appeal to inductive reasons. The suggestion would be that *S*'s belief that there is a red object is justified because *S* believes that she is in circumstances relevantly similar to other circumstances in which *S* has had similar perceptual evidence that turned out to be justified-belief-about-red-object circumstances. But this will not in general do, since in those cases *S*'s justified belief could not have been inductively justified. But clearly *S*'s belief is justified, and arguably this is a function of some reasons *S* has for it. So there must be epistemically good, noninductive, defeasible reasons. Establishing the existence of general defeasible reasons and their role in a theory of justification is one of the most important contributions to epistemology, independently discovered by Chisholm (1966) and Pollock (1967, 1974).

There are, in addition to doxastic foundationalist theories, nondoxastic relatives. Pollock and Cruz (1999) propose such a theory, which they call "direct realism." Their theory is largely foundationalist in structure, with defeasible reasoning at its core. While the details of the theory of defeasible reasoning have evolved significantly, the general framework is closely related to the foundationalist theory developed, for example, in Pollock (1974). However, they reject the doxastic assumption. They argue that any belief (even appearance beliefs) can be held for bad reasons without the agent in question realizing that the relevant belief is being held for a bad reason. A theory of justification ought to predict that in such cases those beliefs are unjustified. If this is right, then it follows that there can be no epistemically privileged class of beliefs. Their solution to this problem is to allow subdoxastic states to enter (as antecedents) into the reason-for relation—for instance, the having of a perceptual image is a defeasible reason for an agent to believe. And, of course, the reason can be defeated. Justifiers, according to this view, are still internal states, but they are not all beliefs.

Coherence Theories

While foundationalist theories assert that some nonempty proper subset of an agent's beliefs are basic, coherence theories deny this. (Thus a doxastic theory is a coherence theory iff it is not a foundationalist theory.) Coherence theories deny the existence of an epistemologically privileged class of beliefs, insisting instead that all of an agent's beliefs have the same fundamental status. Justification is then a function of how well a given belief fits in, or "coheres," with the rest of an agent's doxastic state. The main task for a coherence theorist is to give an account of the coherence relation.

One can taxonomize the class of coherence theories according to distinctions along two dimensions. Along one dimension, an analysis of the coherence relation will specify the structure of the relation and how it relates to other beliefs: For a candidate belief *p*, is it a linear relationship between *p* and some other belief(s) that matters (much like the foundationalist's tracing of reasons), or is it a holistic relationship between *p* and all of the agent's other beliefs? Along a second dimension, analyses of the coherence relation must specify its role, which may be positive (in which case a belief cohering with an agent's doxastic state constitutes positive grounds for the belief to be justified) or relevant only in a negative way (in which case the absence of coherence is a reason to get rid of some beliefs). This yields four types of coherence theory in logical space: positive linear, positive holistic, negative linear, and negative holistic.

Positive linear theories insist that coherence is a linear relationship between a candidate belief *p* and another belief *q* (or perhaps a smallish subset of an agent's beliefs). This linear relationship is just a version of the reason-for relation. Note that since such theories deny that linear appeals between beliefs must always come to an end (else they would be foundationalist theories), positive linear coherence theories must admit either that such an application of reasons can go on infinitely or else that the reason-for relation may be nontrivially cyclic. On the face of it, this seems an unpalatable dilemma: Either some legitimate justificatory chains go on infinitely or some legitimate justificatory chains run in a circle. The first horn connects with the intuition that reasons seem to be only a conduit through which justification flows—reasons cannot, in other words, generate justification, but only allow beliefs to be justified conditional on good starting points. The second horn connects with the intuition that circular reasoning can never lead to a justified belief. Sequences like the following seem clearly bad from an epistemic point of view: Suppose one's only reason for thinking that *p* is that q_1 is a reason for it, and one's only reason for thinking that q_1 is that q_2 is a reason for it, . . . , and one's only reason for thinking that q_n is that *p* is a reason for it. In such cases, how could *S*'s belief that *p* be justified? These considerations are often taken together and labeled as the "regress problem" for coherence theories of justification (Sosa 1980). It is important to note that they properly apply only to (some) positive linear coherence theories, and not to the class of coherence theories in general.

Lehrer (2000) proposes a positive linear coherence theory of justification. In his theory, a belief *p* coheres with a background doxastic system just in case it is more reasonable to accept *p* than any of its "competitors" (roughly, propositions negatively relevant to *p*). One can conclude that accepting *p* is reasonable in this way by appeal to what he calls the "trustworthiness argument": Suppose *S* is trustworthy in what she accepts (for purely epistemic goals). Call this belief *(T)*. Then *S* is trustworthy in accepting the candidate acceptance *p* (for purely epistemic goals). But then it must be reasonable to accept that *p* (otherwise *S* would not be trustworthy in *S*'s acceptance of it). But how does one accept *(T)*, the first premise of this argument? Lehrer says it is a "keystone belief," made reasonable by applying the trustworthiness argument to *(T)*. So the picture is that coherence reduces to reasonableness, which is a linear relation. All justified beliefs end up appealing to *(T)*. But *(T)* is not a basic belief, and itself is linearly supported by appeal to the trustworthiness argument and itself. This is what Lehrer calls the "virtuous loop."

There are a number of positive holistic coherence theories, notably those advocated by Lehrer (1974) and BonJour (1985). Positive holistic coherence theories still insist that coherence is positively relevant to justification but claim that the coherence relation is not linear. The picture is that an agent's doxastic system is a "web" of interrelated beliefs with coherence between *p* and a doxastic system being a function of the entire system (without *p*, of course). This function might be a measure of how well *p* would be explained by the truth of the other beliefs or how well the other beliefs "mutually support" *p* or some function of the agent's subjective probability assignment with respect to *p*. The most general sort of difficulty for this class of theories is making sense of the difference between a belief being justified and a belief being (merely) justifiable. For instance, *S*'s doxastic state may exhibit the proper coherence between *p* and the rest of *S*'s beliefs without *S* being able to tell, perhaps due to complexity considerations, that this is so. Is *S*'s belief justified? More worrisome are cases in which *S*'s belief in *p* is spontaneous or ungrounded (due, say, to an epistemically serendipitous brain lesion) but nevertheless fits in well with *S*'s other beliefs. Such a belief could be justified, since it *ex hypothesi* exhibits the right properties to fit in well with *S*'s other beliefs, but it is far from clear whether *S* believes *p* in an epistemically permissible way.

Negative linear and holistic coherence theories insist that coherence plays only a negative role in assessing whether an agent is justified in believing some candidate belief *p*. Such theories are motivated by the Neurath metaphor: Doxastic states are raftlike and free-floating without an anchor, each plank held in place by its relationship to all the others, with repairs and maintenance taking place at sea (Sosa 1980) (see Neurath, Otto). The picture is that doxastic agents find themselves with a plethora of existing beliefs, and any such belief is automatically *prima facie* justified. The role of coherence is a maintenance tool to shape and refine one's doxastic state by weeding out beliefs that have lost this *prima facie* status. Though a conceptual possibility, there are in fact no extant negative linear coherence theories. Harman (1986) offers the clearest example of a negative holistic coherence theory. According to this view, coherence is overall explanatory coherence. So suppose an agent *S* believes that *p*. This makes *S* automatically *prima facie* justified in believing that *p*. *S* should stop believing that *p* if and when *p* no longer coheres with the rest of *S*'s doxastic state—that is, if and when *S*'s other beliefs make it hard to explain how *p* could be true.

A difficulty facing negative coherence theories, and so negative holistic coherence theories, is that they insist that coherence (and, more broadly, reasoning in general) plays exclusively a negative, undermining role with respect to epistemic justification. As Pollock and Cruz (1999) point out, this is equivalent to the thesis that all of an agent's beliefs are *prima facie* justified, no matter how the agent acquired them. But that seems counterintuitive in cases in which an agent holds a belief, say *p*, on the basis of wishful thinking. *S* may have no specific reason for thinking that belief in *p* is no good, and so may have no negative, undermining lack of coherence that forces *S* to withdraw the belief. Nevertheless, such a belief seems unjustified, in which case it cannot be an undefeated *prima facie* justified belief, as the negative holistic coherence theory appears to predict. In principle, just as there are nondoxastic foundationalist theories, there is room in logical space for nondoxastic coherence theories of justification. However, this presently remains unexplored territory.

Externalist Theories of Justification

Foundationalist theories, both doxastic and nondoxastic, and their coherence counterparts take justification to be a function exclusively of the internal states of the epistemic agent in question. Externalist theories of justification deny this. Some external states—states of the world or modal facts about such states—are relevant to justification. Put another way: Internalists claim that justification supervenes on internal states of the cognizer in question (and so justification is invariant across manipulations in which the internal states are fixed but the facts about the outside world vary), and externalists deny this claimed supervenience.

A common motivation for externalism is the pretheoretic connection between epistemic justification and knowledge. Why would *S* want justified beliefs, as opposed to unjustified beliefs? The immediate thought is that, *qua* epistemic agent, *S* wants to believe what is true and avoid believing what is false. The point of having justified beliefs, then, is instrumental to attaining truth while avoiding error. And so an analysis of justification should have as a consequence that if a belief is justified, it is in some sense likely to be true (where this likelihood is a measure of actual success, not subjective probability). Justification is not merely a matter of a belief 's chance of being true, understood as a result of a stochastic process. Rather, the view is that beliefs are formed and generated by cognitive processes, and (external, perhaps modal) properties of these processes determine the justificatory status of the beliefs they output.

The externalist position picks out a large region of logical space, but most of the extant externalist theories of justification, as a matter of fact, are versions of reliabilism. Broadly, reliabilist theories take belief to be justified just in case they (or their producers) are reliable indicators that the belief in question is true. The most straightforward reliabilist theory is process reliabilism (Goldman 1979). The analysis (schema) is simple: *S*'s belief that *p*, produced by belief-forming process *M*, is justified iff *M* is a reliable process. To turn this into a full account, the process reliabilist must offer a criterion for reliable belief-forming processes. Different process reliabilist accounts then differ (in complexity and scope) depending on the complexity and scope of the analyses of reliable belief-forming processes. The basic analysis is simply that a belief-forming process *M* is reliable iff the actual ratio of true beliefs to false beliefs produced by *M* is sufficiently high. In such an account of reliability, *S*'s perceptual belief that she sees a red apple is justified iff the module forming *S*'s perceptual beliefs tends to produce true beliefs significantly more often than false ones.

As appealing as this simple form of process reliabilism is, it faces major difficulties, which (from the reliabilist's point of view) point toward more sophisticated accounts of reliability. First, tying justification to actual truth ratios does not rule out what might be called "accidental" reliability. For example, suppose *S* has the following belief-forming process: On a certain occasion *t*, *S* decides to believe that it is sunny in a distant location at $t + S$, say Amsterdam, if the toss of a certain coin at *t* turns up heads. *S* flips the coin and it turns up heads, whereupon *S* believes that it is sunny in Amsterdam. In fact, suppose it *is* sunny in Amsterdam. Then *S*'s belief-formation process has maximal reliability, but clearly it is not a justified belief. BonJour (1985) raises similar worries by applying the basic process reliabilist account to the case of a reliable clairvoyant. Accidental reliability, in other words, seems just as bad as accidental truth. Some more subtle account of reliability seems to be in order, one in which modal stability of the truth ratio plays an important role. The "normal-worlds reliabilism" in Goldman (1986) aims, at least in part, at providing such an account. In this more sophisticated version, actual truth ratios are replaced by truth ratios across so-called normal worlds, defined as compatible with the agent's general beliefs about the general presumed structure of the actual world.

Another difficulty for reliabilist accounts as a class, which has proved to be the impetus for much further research, is the generality problem. Take the case of process reliabilism. For any belief, there are a multitude of ways of circumscribing which belief-forming process generated it. For instance, a perceptual belief about the color of an object (say, a red apple) can be described as the output of a color-vision module, color-vision-normally-situated-on-Earth module, color-vision-any-way-situated-on-Earth module, and so on indefinitely, with ever more general descriptions of the process. But it can equally be described as *S*'s-color-vision module, *S*'s-color-vision-since-1983 module, and so on, down to *S*'s-color-vision-module-used-on-this-occasion-to-view-this-apple. Which of these processes is the right one by which to judge the justificatory status of *S*'s belief? Since reliability is being treated as an objective conditional probability (i.e., the probability that a belief is true given it is produced by process *M*), the answer matters a great deal.

A total evidence requirement seems to compel the description to be as specific as possible, but then this conditional probability goes to 1 (if the apple is the color S believes it to be) or to 0 (if not), trivializing the account, since it would render a belief justified iff true. Determining the appropriate reference class of worlds raises essentially the same issue for normal-worlds reliabilism. Indeed any account that relativizes justificatory status of a belief to circumstances in the local environment must face this problem.

A final puzzle for reliabilism, and for externalist theories of justification more generally, is the problem posed by good epistemic agents who find themselves in epistemically unfortunate circumstances (Cohen 1984). Consider two epistemic agents S_1 and S_2. S_1 is a very careful reasoner, always forming beliefs on the basis of excellent evidential support and in short is as epistemically responsible as any normal cognizer. S_2, on the other hand, is a very sloppy epistemic agent, always forming beliefs on the basis of wishful thinking, fancy, and faulty reasoning. As a happenstance, S_1 and S_2 have a lot of beliefs in common—that is, there are many beliefs about the physical world that they share, which form the set C. Now, as it happens, S_1 and S_2 inhabit (unbeknownst to them) an Evil Demon world—their beliefs about the world are systematically and radically false, and in particular all beliefs in C are false. The facts are being manipulated by the whimsy of an Evil Demon whose aim is to deceive them. For any belief in C, can there be any difference in justificatory status between the two agents? Reliabilist theories, insofar as the two cognizers *ex hypothesi* have identical actual and counterfactual truth ratios (namely, 0), treat any two beliefs of the respective agents as unjustified. But this seems implausible, since there is an important epistemological difference between S_1's belief that, say, most swans are white (formed on the basis of a standard induction with a large inductive base) and S_2's belief that most swans are white (a randomly occurring thought to S_2). The difference seems to be a difference in justification.

The extent to which such worries upset the externalist project is an open question. Some externalists just deny the purported intuitive difference above. It is also unclear to what extent this example depends on particular incarnations of externalist theories like process reliabilism. Goldman (1986), in fact, seems to advocate exploring both of these themes, remarking that many do not find the process reliabilist prediction obviously counterintuitive (113, n. 32), but that perhaps the best way to understand reliability is in terms of normal worlds (113).

Epistemic Change

While theories of justification aim at a characterization construed as a property of beliefs in a given epistemic state, there is a related project that has received much attention, that of characterizing justification construed as a property of transitions between sets of beliefs, or, more generally, as a property of transitions between epistemic states. The main desideratum for such theories is to say in a precise way how an epistemic state ought to change under the impact of some new information—that is, to provide a characterization of rational or justified belief change. From a structural point of view, this is just the problem of theory change in science: How ought a theory to be changed in light of new, perhaps connecting, information? (See Gärdenfors 1988 and Hansson 1999 for introductions, surveys, and references to belief dynamics research.)

From a logical point of view, a main task in belief dynamics research is to specify a model of the revision process. In general terms, a model specifies four things: a set of epistemic states, a language L suitable for expressing epistemic commitments, a canonical relation of epistemic commitment (relating states to formulas of L), a language L_0 (possibly different from L) of possible epistemic inputs (these represent the impetus for change), and a revision function over the set of states. Given a set of beliefs, say, the deductive closure of $\{ p, p \mathbin{!} q \}$, logic alone will not in general tell one how to revise it. If one in such a state learns, to one's surprise, that not-q, then one has to either stop believing p or stop believing $p \mathbin{!} q$. The task, then, is to attempt to codify in very general terms what one should do in all such cases—what the rational mandates are on the transitions between possible epistemic states.

The most well known model of belief revision is the AGM model, named after its trio of developers Alchourrón, Gärdenfors, and Makinson (1985). The AGM model takes epistemic states to be deductively closed subsets of a language of classical propositional logic. They then place constraints on rational revision functions by listing postulates that such functions ought to obey. The postulates are largely driven by codifying the intuition that rational changes of belief should be minimal changes of belief. It turns out that the functions that meet the AGM postulates are exactly those that can be described in the following way. Fix a space W of possible worlds. Think of an epistemic state as a subset of that space—intuitively, just those worlds consistent with what an agent believes. Suppose that the agent can assign a system of spheres around such a subset. The idea is that given a state *X*, an agent assigns relative implausibility of the worlds in *W*; the farther out the sphere containing a world, the more implausible the agent finds that world relative to *X*. If one has to revise *X* to reflect the new information that °, then one should adopt as candidates for the actual world the set of least implausible °-worlds. So one adopts as one's new epistemic state the closest °-worlds in terms of the system of spheres centered around *X*. In order to be prepared to repeat the process, one would now also need to adopt a new system of spheres, one that is centered around the new epistemic state.

Theories of belief revision differ in at least four important ways from theories that lie within the Bayesian tradition (see Bayesianism). First, revision models tend to treat 'belief' as flat footed and full, whereas in the Bayesian tradition, *degrees* of belief are of primary interest. Second, the representation here is entirely "qualitative" in the sense that an epistemic state is represented by a set of possible worlds (or, what amounts to the same thing, a logically closed set of sentences) plus a comparative ranking. Epistemic states in the Bayesian tradition, on the other hand, assume a much richer representation of states, full probability distributions over the language. In this sense the revision models discussed here are more general than their Bayesian counterparts, since they require less structure to be assumed on the agents' states. Third, sets of beliefs are deductively closed in belief revision models, whereas this is true in the Bayesian tradition only for special (and one assumes rare) subsets of beliefs in which all members have probability 1. In this sense, Bayesian models are the more general of the two. And, fourth, in belief revision models, there is no special problem of revising by some fact that an agent previously had ruled out, whereas this is a notoriously difficult problem in the Bayesian tradition.

From a philosophical point of view, there are many unsettled debates surrounding theories of belief revision. One such debate is borrowed and adapted from the literature on justification taken as a property of beliefs: Is rational belief change constrained by coherence principles or foundationalist principles? Foundationalist belief change takes seriously, in one way or another, the foundationalist intuition: Agents hold some beliefs just because they hold others, and this difference makes a difference to the landscape of belief revision. For, if *S* believes *q* just because she believes *p*, then if she is forced to give up her belief in *p*, she should also give up her belief in *q*. Coherence theories of belief revision deny that such asymmetrical relations play an interesting role in belief change (they typically try to explain away the problematic examples); instead they insist that the guiding aim in rational belief change is information conservation.

The AGM model is a coherence theory of belief dynamics in this sense. Unlike the case for classical theories of justification, there is at present no clear consensus on just what counts as a foundationalist model (this seems to be a much wider class than that of coherence models), or on the characterization of theories that take the foundationalist intuition seriously, or on whether coherence theories or their foundationalist counterparts are to be preferred (Harman 1984; Hansson 1999; Pollock and Gillies 2000; Gillies 2004). Other open questions in belief dynamics include the status of conditionals and epistemic modalities in belief revision models, how revision relates to questions in qualitative decision theory, the relationship between revising an epistemic state (to reflect an agent learning it was mistaken about some fact) and updating it (to reflect that the world has changed), and generally how the rational dynamics of epistemic states relates to dynamics in other cognitive domains.

Anthony S. Gillies

References

Alchourrón, C. E., P. Gärdenfors, and D. Makinson (1985), "On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision," *Journal of Symbolic Logic* 50: 510–530.

Alston, W. (1986), "Internalism and Externalism in Epistemology," *Philosophical Topics* 14: 179–221.

BonJour, L. (1985), *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Chisholm, R. (1966), *Theory of Knowledge*. Englewood Cliffs, NJ: Prentice-Hall.

Cohen, S. (1984), "Justification and Truth," *Philosophical Studies* 46: 279–296.

——— (1988), "How to Be a Fallibilist," *Philosophical Perspectives* 2: 91–123.

DeRose, K. (1995), "Solving the Skeptical Problem," *Philosophical Review* 104: 1–52.

Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, MA: MIT Press.

Gettier, E. (1963), "Is Justified True Belief Knowledge?" *Analysis* 23: 121–123.

Gillies, A. S. (2004), "New Foundations for Epistemic Change," *Synthese* 138: 1–48.

Goldman, A. (1979), "What Is Justified Belief?" in G. Pappas (ed.), *Justification and Knowledge*. Dordrecht, Netherlands: D. Reidel.

——— (1986), *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Hansson, S. O. (1999), *A Textbook of Belief Dynamics*. Boston: Kluwer Academic Publishers.

Harman, G. (1984), "Positive versus Negative Undermining in Belief Revision," *Noûs* 18: 39–49.

——— (1986), *Change in View*. Cambridge, MA: MIT Press.
Lehrer, K. (1974), *Knowledge*. Oxford: Oxford University Press.
——— (2000), *Theory of Knowledge*, 2nd ed. Boulder, CO: Westview Press.
Lewis, D. (1996), "Elusive Knowledge," *Australasian Journal of Philosophy* 74: 549–567.
Nozick, R. (1981), *Philosophical Explanations*. Oxford: Oxford University Press.
Plantinga, A. (1993), *Warrant: The Current Debate*. Oxford: Oxford University Press.
Pollock, J. L. (1967), "Criteria and Our Knowledge of the Material World," *Philosophical Review* 76: 28–62.
——— (1974), *Knowledge and Justification*. Princeton, NJ: Princeton University Press.
——— (1979), "A Plethora of Epistemological Theories," in G. Pappas (ed.), *Justification and Knowledge*. Dordrecht, Netherlands: D. Reidel.
Pollock, J. L., and J. Cruz (1999), *Contemporary Theories of Knowledge*, 2nd ed. New York: Rowman and Littlefield.
Pollock, J. L., and A. S. Gillies (2000), "Belief Revision and Epistemology," *Synthese* 122: 69–92.
Russell, B. (1912), *The Problems of Philosophy*. Oxford: Oxford University Press.
Sosa, E. (1980), "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge" [in French], in T. E. Uehling and H. K. Wettstein (eds.), *Midwest Studies in Philosophy*, vol. 5: *Studies in Epistemology*. Minneapolis: University of Minnesota Press.
Spohn, W. (2001), "A Brief Comparison of Pollock's Defeasible Reasoning and Ranking Functions," *Synthese* 131: 39–56.
Williamson, T. (2000), *Knowledge and Its Limits*. Oxford: Oxford University Press.

ERROR

See **Statistics, Philosophy of**

EVOLUTION

The term "evolution" is used to describe both the fact of common descent—that all organisms living today have descended from common ancestry—and the process of descent, or the ways in which species have diversified over time. The question of whether evolution is fact or theory trades on this dual use of the term; it is both a fact that evolution has gone forward, and there is a theory—evolutionary theory—that describes the process of change over time. Those who raise the fact-or-theory question may have in mind another question, however: Is the evidence sufficient to support the claim of common descent? Over 150 years of natural history, paleontology, biogeography, developmental biology, and molecular genetics have provided ample evidence for evolution in this sense (for a review, see Ridley 1996). Biologists do not argue about the fact of common descent; they do, however, argue about which mechanisms have operated in specific cases and which patterns and processes occur most frequently. Some of these debates will be discussed below. The structure of this article will be as follows. First, there will be a very brief overview of the history of evolutionary theory since Darwin; then a discussion of evolutionary theory today; and finally a review of debates among philosophers of biology, many of which originated in historical debates among biologists about evolutionary theory and its interpretation.

History of Evolutionary Theory

The term "evolution," from the Latin *evolutio* (*evolvere*, to unroll or unfold), was first used in the scientific context to describe the process of embryological development in an individual organism. The term was chosen in light of some early embryologists' views that development was simply the unfolding of preexisting parts. Lyell first applied the English word "evolution" as early as 1832 to theories of species change. Only by the late nineteenth century did Herbert Spencer use the term to characterize Darwin's theory of the origin of species (Bowler 1975).

The idea of evolution, that diversity of life may have arisen by a natural process, is much older than the 1830s. The germ of this idea was arguably nascent in Lucretius, Descartes, and Hume. For example, in Hume's *Dialogues Concerning Natural Religion* ([1779] 1947, 185), Cleanthes attributes to Philo the view that

> No form, [of plant or animal] . . . can subsist, unless it possess those powers and organs requisite for its subsistence: some new order or economy must be tried, and so on, without intermission; till at last some order, which can support and maintain itself, is fallen upon.

This characterization of origins of new varieties is arguably a predecessor to the idea of natural selection: Options are tried until something works. The novelty in Darwin's contribution to evolutionary biology was not the suggestion that natural, rather than supernatural, causes explained the diversity and adaptation of species. Rather, it consisted in the discovery of a very simple mechanism by which this process moved forward: natural selection.

Natural selection is the differential success in survival and reproduction of some entity (gene, organism, population) due to differences in adaptation to environmental conditions (see Natural Selection). Darwin ([1859] 1964) introduced this idea in *On the Origin of Species*. There, he summarizes evidence for the following empirical generalizations:

1. There exists variation among members of a species.
2. There are more organisms born than survive and reproduce.
3. Certain traits are correlated with an individual's propensity to survive and reproduce.
4. Many such traits are heritable.

From this set of generalizations, Darwin concludes the following: "Every slight modification, which in any way favored the individuals of any species, by better adapting them, would tend to be preserved." Darwin ([1859] 1964) called this his "hypothesis of natural selection" (61). In Darwin's work, the individual is the primary unit of selection. However, he does appeal to competition among groups in attempting to explain the evolution of altruism in human populations. Darwin argued that this simple process could yield not only novel adaptations within species but, over time, the diversity of species seen today. All species, thus, share common ancestry. The idea of common descent was not original to Darwin, nor was it (at least in the scientific community) altogether controversial when proposed by him. Lamarck, St. Hilaire, and many others, including Darwin's grandfather, Erasmus Darwin, had speculated on what was then called the "transmutation" of species. What was novel to Darwin was the view that natural selection was the main mechanism of descent. In later editions of the *Origin*, however, Darwin placed greater emphasis on other mechanisms, through which changes were somehow induced by "effects of the environment," taken up, and passed on to offspring.

Darwin spent five years traveling and collecting specimens of living and fossil animals and plants in South America as a companion to the captain on the *Beagle*. The *Origin* was an attempt to address the following questions, in part inspired by Darwin's travels: Why do animals and plants vary as they do relative to different geographies and climates? Why do species on islands, such as the Galápagos, seem so closely related to species on the mainland, and yet vary in certain ways? And finally, what explains the fact that fossil species share so many characteristics with living species at the same locations? Darwin deliberated on these questions for several years after returning to London. There, he was surrounded by other scientists (several of whom he put to work in cataloguing and identifying his specimens), such as Lyell, Grant, Owen, and Hooker, who were interested in similar questions about the history of the Earth and the diversity of animal life and its causes. Darwin spent a number of years publishing scientific work as well as popular work on his travels; however, he was reluctant to publish his work on the origin of species. Many of his colleagues were extremely skeptical of evolutionary hypotheses, some of which were published in popular presses and motivated by political, and not always scientific, purposes. However, after reading a manuscript on the topic by Wallace (1858), "On the Tendency of Varieties to Depart Indefinitely from the Original Type," Darwin was prompted to put forward the theory of evolution by natural selection. Wallace had independently arrived at a very similar hypothesis to Darwin's, that diverse types of organisms have greater or lesser success at survival and reproduction and that cumulatively, this leads to change in the constitution of populations.

One of the main difficulties for Darwin's theory is that he did not have an adequate view of heredity. Darwin mistakenly thought that heredity was a process of "blending," thus opening himself to the objection that the effects of selection would be "swamped," or reversed, in every generation (see Population Genetics). The rediscovery of Mendel and subsequent development of genetics supplied a mechanism for inheritance that treated traits as discrete, rather than blending, thus resolving the difficulties for Darwin's theory of inheritance (see Genetics). Further, the consequences of Mendelism and the notion of heritability were given precise quantitative formulations with the development of population and quantitative genetics. Haldane, Fisher, and Wright developed mathematical models to represent evolutionary change as occurring in genotypic frequencies, due to mutation, migration, selection, drift, and assortative mating. The early population geneticists demonstrated how, even with very small selective differences, over time, large evolutionary changes could be effected. The mathematical models of evolutionary genetics provided evolutionary biology with a firm theoretical foundation.

This mathematization of evolution, along with the collective efforts of paleontologists, systematists, and geneticists, was the basis of a new "synthetic" theory of evolution (for an overview, see Mayr and Provine 1980). Beginning in the 1930s and 1940s, biologists developed a consensus on the main mechanisms of evolution. Essentially, the evolutionary synthesis was an agreement that independent evidence from diverse fields showed that the pattern of diversity today and in the fossil record could be explained as a result of gradual change, due primarily to selection in response to environmental changes over the course of geological history. This view has been called "neo-Darwinian," insofar as it can be traced back to Darwin's view that the gradual changes seen accumulating in species, such as in domestic races, would gradually lead to new varieties and, eventually, to the diversity of life seen today.

The evolutionary synthesis was in part a reaction to challenges from two different sectors. First, because the gaps between species seemed so significant, some biologists (notably Goldschmidt 1940) argued that macroevolution, or change among species, was a fundamentally different process than microevolution, or change within species. Goldschmidt claimed that macroevolution required a change in the genetic makeup of organisms that was different in *kind* from those changes resulting from simple mutation, selection, migration, assortative mating, and drift. He wrote:

> Microevolution by means of micromutation leads only to diversification within the species. . . . [T]he large step from species to species is neither demonstrated nor conceivable on the basis of micromutations. (396–397)

According to Goldschmidt, there is a "bridgeless gap" between species. The geographic races are thus not incipient species, as Darwin had argued. The road to novel species was not via simple mutation and selection, but via what Goldschmidt called "systemic mutations."

In response to Goldschmidt and other "macromutationists," the authors of the synthesis held that novel species arose by the very same mechanisms observed to cause change within species (Dobzhansky 1937; Mayr 1942; Simpson 1944). As evidence for this perspective, Dobzhansky demonstrated that genetic variation among species is not different in kind than variation within species. The very same mutations, ranging from base-pair changes to transversions and translocations, may be observed both within and among species lineages. Further, Mayr (1964) documented in great detail how species varied in ways that correlated with features of their environment, such as soil type, vegetation, microclimate, and topography, and how these variations gave rise to geographic races. These races may sometimes (though not always) give rise to incipient species, by the very same mechanism (selection) that yields changes within species. Geographic isolation of a group of individuals by a feature of the environment (e.g., mountain range, river) or a catastrophic event (e.g., a flood) will change a population over time in such a way that it becomes reproductively isolated, or unable to interbreed with its parent population. Mayr's preferred mechanism of speciation is called allopatry.

The founders of the evolutionary synthesis were also reacting to the development of molecular biology (see Dietrich 1998; Beatty 1990). Many biologists, particularly Mayr and Dobzhansky, felt that the growth of molecular biology challenged the work of "classical" evolutionists. In the late 1950s, some molecular biologists argued that the major questions of phylogeny and taxonomy could be solved by molecular methods; "chemical phylogenetics" and "protein taxonomy" would replace classical methods of natural history, biogeography, and classical systematics. What some have called the dogmatism of the synthesis was thus arguably prompted in part by a battle for institutional power and resources between the "young Turks" (molecular biologists such as Watson, Zuckerkandl, Pauling, Margoliash, and Fitch) and the old guard (Mayr and Dobzhansky). Mayr argued that the new molecular methods, while useful, could not supersede or serve to answer the same questions about the diversification of species that natural history could. In order to understand the pattern and process of speciation, one needed to investigate not only species' biochemical or genetic makeup, but also how climate and biogeography could yield different patterns of species diversity.

Critics of the synthesis argued that proponents were too narrow in their views about the pattern and process of evolutionary change (Gould 1983). According to Gould (2002), the founders of the synthesis emphasized gradual change, as opposed to punctuation followed by stasis, and ignored theoretical and empirical contributions from embryologists and developmental biologists. Punctuated equilibrium is the view that there are long periods of very little change in the fossil record, followed by rapid transformation of species (Gould and Eldredge 1977). However, some have argued that Gould's criticism rests on a false account of the history of the synthesis (see Charlesworth, Lande, and Slatkin 1982). Nonetheless, there have been perceived challenges to the "synthetic" view of evolution from molecular biology, paleontology, systematics, and developmental biology in the past 50 years. For instance, Kimura (1968) and King and Jukes (1969) argued that many changes at the molecular level are neutral with respect to selection. Kimura's claim was not just that most changes are selectively neutral, but that most evolutionary changes are due to the random fixation of neutral (or nearly neutral) alleles. At first, this was perceived as a challenge to what many felt was a consensus that selection was the main force driving evolutionary change. However, many of these apparent challenges have been integrated into mainstream evolutionary theory.

Evolutionary Theory Today

There has effectively been a new synthesis between classical evolutionary biology and molecular evolution. Molecular biologists, systematists, paleontologists, and developmental biologists work together to understand the pattern and process of evolutionary change at the molecular level and above. Since the 1960s, new technologies have enabled evolutionary biologists to study evolution at the molecular level. Molecular evolutionary biologists may observe and quantify the amount of genetic divergence within and among species (Nei and Kumar 2000). Molecular methods are used by biologists of every stripe, from ecologists to developmental biologists. Today, systematists and paleontologists routinely sequence samples of genetic materials from their specimens to determine rates of change and time since the most recent common ancestor. (However, most think that such inferences cannot be based purely on sequence data.)

The synthetic theory has been moderated by new evidence and research. For example, it is no longer universally agreed that speciation is primarily a gradual process, that the origin of species requires changes in many genes, or that speciation in sympatry, or within the range of the ancestral population, is an extremely infrequent occurrence. There are examples of species differences due to as little as two genes (e.g., mimicry in butterflies [Sheppard et al. 1985]). Modes of speciation other than allopatry have been recognized and are gaining acceptance—notably, speciation in sympatry (Coyne and Orr 2004). Some argue against what they claim is one of the central theses of the synthesis, that rates of macroevolutionary change are uniform rather than punctuated by stasis followed by rapid bursts of change (Gould and Eldredge 1977).

Finally, the investigation of evolution has come to employ not only the tools of genetics, ecology, systematics, and paleontology, but also developmental biology. Evolutionary developmental biology, or the study of the evolution of developmental systems, has become an active area of investigation (see Carroll, Grenier, and Weatherbee 2001). Goldschmidt's challenge has thus been addressed insofar as biologists are now more attentive to the effects of major developmental changes, such as heterochrony (changes in rates of development), as playing a role in the generation of novel species.

Philosophical Issues

What is the structure of evolutionary theory? Theoretical population genetics uses the laws of probability and idealized mathematical models to generalize about evolving populations. Theoretical population genetics has thus been described as the "dynamics" of the theory (Sober 1984), akin to Newton's dynamical theory of physics. In this view, the mathematical models of theoretical population genetics describe the main "forces" effecting change in the genetic constitution of populations over time, or how migration, selection, drift, assortative mating, and mutation effect the genetic constitution of populations from one generation to the next. So, the mathematical theory of evolution is constituted by a family of models describing how different types of causal factors effect evolutionary change.

What does it mean to speak of selection as a force, a mechanism, or a cause? It is clear that selection, for instance, is not a cause of genetic changes in populations in the way that a baseball striking a window causes it to shatter. It is a cause of change at the level of populations rather than individuals. But what does it mean to say that there are "population-level causes"? What is the relationship between causal processes at the individual and at the population level? These sorts of concerns about the interpretation of causal terms and concepts in biology have generated some controversy in the philosophical literature. Some have argued that given the statistical character of evolutionary theory, the language of "forces" is inappropriate.

However, deployment of the force metaphor is not inconsistent with the recognition that evolution is a statistical theory. On behalf of Sober's "forces" view, Wright's classic paper ([1931] 1986) is a reminder of how it is a useful shorthand to speak of evolutionary theory as a theory of forces. Wright recognized that selection and drift were statistical processes. Yet, he argued that they act "deterministically," in the sense that they both deterministically decrease genetic heterogeneity. Yet, they are also "statistical" or "indeterministic" processes, insofar as possession of this or that selectively advantageous trait does not guarantee survival or reproduction, but selection coefficients describe an "average" survivorship or fecundity relative to one's cohort. Nonetheless, Wright describes selection, drift, and linkage as forces that decrease genetic variability in a population, and he describes recombination, migration, and large population size as forces that tend to increase genetic variability. This is not to say that these are 'forces' in the Newtonian sense, but rather that on average, selection systematically increases homozygosity, as does decreased population size. Smaller populations are, on average, less genetically diverse than larger populations. Wright took it to be the case that there needs to be an appropriate "balance" of forces leading to both genetic homogeneity and heterogeneity in order for populations to be "plastic" enough to evolve: "A balance between factors of homogeneity and heterogeneity may provide a more favorable condition for evolution than either factor by itself" (Wright [1931] 1986, 146). Of course, Wright was well aware that, as noted earlier, large population size is not a 'force' in the Newtonian sense; rather, it is a condition that makes it the case that chance events are less likely to change the genetic constitution of a population. And selection is not, strictly speaking, a force that *directs* evolution, but a consequence of the differential survival of individual organisms due to their differences in adaptation relative to their local environments. However, Wright usefully deployed the analogy with Newtonian physics to illuminate how adaptive evolution required a combination of factors or forces operating in combination. Specifically, Wright believed that a combination of isolation, drift, and intra- and interdemic selection was an optimal balance of forces for generating adaptation.

This appeal to the notion of 'force' did not prohibit Wright's simultaneous conviction that evolution was a statistical process. He was particularly sensitive to the fact that given populations of interbreeding organisms are finite and that chance, or sampling error, must play a key role in changes in gene frequency in a population from one generation to the next. In populations of small size, drift (or sampling error) will govern changes in genetic constitution to a greater extent than will selection. (More precisely, the strength of the selection pressure determines how small the population would need to be for drift to be the primary factor.) In larger populations and over the long term, even a very small difference in fitness between organisms possessing genotype x and genotype y may yield dramatic changes in the constitution of the population. These generalizations have something like the structure of a law. However, one might argue that this particular generalization is not a law of nature, but at bottom a fact about probability, that is, it is reducible to something like the claim that when one flips a coin biased toward heads ten times, one is not as likely to be able to determine that it is biased as when one flips the coin a hundred times. Any finite population will be subject to drift; or chance or "sampling" error play an inevitable role in the change in the genetic constitution of populations. This is what some people mean when they say that evolution is subject to "chance," or that evolution is irreducibly "probabilistic" in character. It is an interesting philosophical question whether and in what sense drift is a cause of evolutionary change. Yet, to concede this is not to deny that Wright's talk of forces is an effective, albeit metaphorical, way of describing the balance of factors that contribute to genetic variability in populations. Wright's work deserves praise as an example of how one can usefully combine multiple metaphorical ways of describing and explaining evolution.

Finally, to return to the question that opened this essay: What is the relationship between evolutionary theory and the evidence in support of it? The support for common descent is indirect and depends on several lines of evidence. Nevertheless, the fact of common descent is not a matter of controversy, nor is the thesis that natural selection has been a major factor at work in descent with modification. Ultimately, the most well supported theory is the one that, all evidence taken into consideration, best explains the phenomena than any alternative. It is uncontroversial (at least in the scientific community) that this is true for evolution. The evolution of life is the most likely explanation of a wide range of data—not simply the diversity and adaptation of life, but also the uniformity of life from the molecular level on up. Consider the following. Heredity is controlled by DNA and RNA in organisms as diverse as viruses and humans. The genetic code, or the codons that determine amino acids that make up proteins, which themselves control all the functions in living cells, is uniform across species (with very few exceptions). The same genes (cytochrome C, hemoglobin) can be found across species, and the relatedness among species (time since most recent common ancestor) is correlated with the number of substitutions of nucleotides in these sequences. The embryonic stages of development of diverse vertebrates such as chickens, dogs, and chimpanzees parallel almost exactly. All of these observations are not only well explained by the hypothesis of evolution, but evolutionary theory also entails precise predictions—for instance, about rates of change in various sequences—that are borne out by the evidence. Sequences of DNA and RNA with important functions are strongly conserved, whereas nonfunctional portions of the genome have a relatively quick rate of turnover.

Moreover, many lineages, and natural populations, have been so well studied over so many generations that biologists have been able to describe the ecological conditions affecting change in populations, the rates of change, and the relative significance of selection and drift. An example is Darwin's finches (Geospizinae) on the Galápagos Islands. Gibbs and Grant (1987), for instance, have demonstrated that large adult size is favored under drought conditions and that smaller sizes tend to increase when there is an increase in rainfall. Thus, there should be no question whether Darwin's theory has been borne out by the evidence to date.

The pattern and process of evolution, and how to study it, is more than a useful case study for classic questions in the philosophy of science about theory structure, confirmation, and explanation. Evolution is also the explanation for how *Homo sapiens* came to exist, and so is potentially relevant to questions in ethics, moral psychology, philosophy of mind, and epistemology (see Evolutionary Epistemology; Evolutionary Psychology). In addition, questions that arise within the science of evolutionary biology itself are not simply empirical, but also conceptual, and so repay philosophical examination. For instance, debates in evolutionary biology about the possibility of the evolution of altruism have moved forward in part via critical philosophical examination of the models in question (Sober and Wilson 2001) (see Altruism). In addition, the question of how to define species, the problem of reconstructing the history of species lineages, determining the relative significance of drift and selection in evolving lineages, examining the relationship between micro- and macroevolution, as well as the relationship between molecular and evolutionary biology, are all active areas of investigation in biology that may be usefully served by philosophical inquiry (see Molecular Biology; Species). More generally, evolutionary biology and its sister disciplines of genetics and ecology are important case studies for broader questions in the philosophy and history of science about modeling and idealization, determinism, reduction, explanation, theoretical unification, instrumentalism and realism, and scientific progress.

ANYA PLUTYNSKI

The author acknowledges the helpful input of Gary Hatfield.

References

Beatty, John (1990), "Evolutionary Anti-Reductionism: Historical Reflections," *Biology and Philosophy* 5: 199–210.

Bowler, Peter J. (1975), "The Changing Meaning of 'Evolution,'" *Journal of the History of Ideas* 36: 95–114.

Carroll, Sean B., Jennifer K. Grenier, and Scott D. Weatherbee (2001), *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Cambridge, MA: Blackwell Science.

Charlesworth, Brian, Russell Lande, and Montgomery Slatkin (1982), "A Neo-Darwinian Commentary on Macroevolution," *Evolution* 36: 474–498.

Coyne, Jerry, and H. Allen Orr (2004), *Speciation*. Sinauer Associates Publishing.

Darwin, C. ([1859] 1964), *On the Origin of Species*. Edited by Ernest Mayr. Cambridge, MA: Harvard University Press.

Dietrich, Michael (1998), "Paradox and Persuasion: Negotiating the Place of Molecular Evolution within Evolutionary Biology," *Journal of the History of Biology* 31, 85–111.

Dobzhansky, T. (1937), *Genetics and the Origin of Species*. New York: Columbia University Press.

Gibbs, Leslie, and Peter Grant (1987), "Oscillating Selection on Darwin's Finches," *Nature* 327: 511–513.

Goldschmidt, Richard (1940), *The Material Basis of Evolution*. New Haven, CT: Yale University Press.

Gould, Stephen Jay (1983), "The Hardening of the Modern Synthesis," in M. Grene (ed.), *Dimensions of Darwinism*. Cambridge: Cambridge University Press, 71–93.

——— (2002), *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.

Gould, Stephen J., and Niles Eldredge (1977), "Punctuated Equilibria and the Tempo and Mode of Evolution Reconsidered," *Paleobiology* 3: 115–151.

Hume, David ([1779] 1947), *Dialogues Concerning Natural Religion*. Edited by Norman Kemp Smith. London and New York: Thomas Nelson and Sons.

Kimura, Motoo (1968), "Evolutionary Rate at the Molecular Level," *Nature* 217: 624–626.

King, Jack, and Thomas Jukes (1969), "Non-Darwinian Evolution," *Science* 164: 788–798.

Mayr, E. (1942), *Systematics and the Origin of Species*. New York: Columbia University Press.

——— (1964), *Animal Species and Evolution*. Cambridge, MA: Harvard University Press.

Mayr, Ernest, and William Provine (1980), *The Evolutionary Synthesis: Perspectives on the Unification of Biology*. Cambridge, MA: Harvard University Press.

Nei, Masatoshi, and Sudhir Kumar (2000), *Molecular Evolution and Phylogenetics*. Oxford: Oxford University Press.

Oyama, Griffiths, and R. D. Gray (2001), *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge, MA: MIT Press.

Richards, Robert (1994), "Evolution," in E. Fox Keller and E. Lloyd (eds.), *Key Words in Evolutionary Biology*. Cambridge, MA: Harvard University Press.

Ridley, Mark (1996), *Evolution*. Cambridge, MA: Blackwell Science.

Sheppard, Paul, J. R. Turner, K. S. Brown, W. W. Benson, and M. C. Singer (1985), "Genetics and the Evolution of Muellerian Mimicry in Heliconius Butterflies," *Philosophical Transactions of the Royal Society of London, B: Biological Sciences* 308: 433–610.

Simpson, G. C. (1944), *Tempo and Mode in Evolution*. New York: Columbia University Press.

Sober, Elliott (1984), *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, MA: MIT Press.

Sober, E., and D. S. Wilson (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Wallace, A. R. (1858), "On the Tendency of Varieties to Depart Indefinitely from the Original Type," *Journal of the Proceedings of the Linnaean Society, Zoology* 3: 53–62.

Wright, Sewall ([1931] 1986), "Evolution in Mendelian Populations," in W. B. Provine (ed.), *Evolution: Selected Papers by Sewall Wright*. Chicago: University of Chicago Press. Originally published in *Genetics* 16: 97–159.

See also **Adaptation and Adaptionism; Biological Information; Ecology; Evolutionary Psychology; Fitness; Natural Selection; Population Genetics; Species**

EVOLUTIONARY EPISTEMOLOGY

Evolutionary epistemology refers to a variety of approaches to the theory of knowledge that emphasize the evolutionary dynamic of knowledge acquisition and evaluation. It therefore has its roots in Charles Darwin's *On the Origin of Species* and *The Descent of Man* and Herbert Spencer's independently developed evolutionary theory of just about everything. The motivating idea is that the capacity to know and knowledge itself, human or otherwise, is a product of evolutionary forces. Any attempts to analyze or understand the nature of knowledge must take this fact into account.

Almost immediately upon the publication of Darwin's work, others began to extend the Darwinian insights to the problem of knowledge. Chief among these were the American pragmatists Charles Peirce, William James, Chauncey Wright, and John Dewey, the psychologist James Mark Baldwin, and many others, including Friedrich Nietzsche and the author Samuel Butler. Much of the work on evolutionary epistemology in the twentieth century derives from the work of Konrad Lorenz, Donald Campbell, Karl Popper, and Jean Piaget (for historical references, see Campbell 1974; Bradie 1986).

Evolutionary insights, models, and metaphors have been brought to bear on a bewildering array of diverse issues, from the evolution of organisms themselves to the development of scientific knowledge. Some have urged that biological evolution in and of itself is a knowledge process. Others have urged that conceptual change is an evolutionary process that mimics the process of natural selection. The literature is rife with analogies and metaphors. Some are drawn from biology to characterize the growth of knowledge. Some are drawn from models of knowing to inform our understanding of biological evolution. Some see biological natural selection and the evolution of scientific understanding as two examples of a single process. Others see selective processes everywhere and promote a view that has come to be labeled universal Darwinism (Plotkin 1993; Cziko 1995; Blackmore 1999).

Evolutionary epistemologies are often seen to be related to or a variant of so-called naturalized epistemologies.

Two Programs

It is useful to distinguish between two interrelated yet arguably distinct projects. On the one hand, there are projects that aim to explain or understand the development of the physical and psychological mechanisms by means of which animals and humans come to acquire and process information about the world. These have been labeled evolution of epistemic (or epistemological) mechanisms (EEM). On the other hand, there are projects that aim to understand the nature and development of the content, norms, and methods of information systems, knowledge corpuses, and scientific theories or traditions. These have been labeled evolution of epistemic theories (EET) (Bradie 1986). Not everyone accepts this division, but it is useful to keep the two projects distinct. For one thing, they do not stand or fall together. EEM programs involve the application of evolutionary biological methods to the study of the development of brains, sensory organs, nervous systems, motor systems, and the like, which are, as far as is known, the *sine qua non* for sentient and sapient creatures. As such, they share the cachet of the success and general acceptance of a broadly Darwinian view of the evolution of characters and traits. EET programs, on the other hand, trade on analogies or metaphors drawn from evolutionary biology and may well turn out to be false or unfruitful characterizations of the development of knowledge. Even those, like David Hull, who argue that the two processes are exemplars of a single overriding model admit that the details of the mechanisms promoting change in the two cases are not identical (Hull 1988). Therefore, the one could turn out to be essentially right and the other essentially wrong. At the moment, it is clear that EEM programs are probably basically right, though filling in the details is fraught with all the problems and more that plague phylogenetic reconstructions. Brains and their cultural products, unlike bones, do not fossilize easily. If one includes the problem of reconstructing the phylogeny of the evolution of mental capacities, the difficulties become formidable indeed. The verdict is not yet in with respect to the various attempts to reconstruct the development of human knowledge in terms of evolutionary models (see Evolutionary Psychology).

In addition to the distinction between the EEM and EET programs, there is the distinction between *phylogenetic evolution* and *ontogenetic development*. In order to understand the structure of the human brain, for instance, two separate though related questions can be asked, both of which can be couched as, Why do human beings have the kind of brains they do? Such a question, as Ernest Mayr (1961) pointed out, can be given either proximate or ultimate answers. The ultimate, or phylogenetic answer, will turn on the contingencies of the evolution of the brain in the human lineage. The proximate, or ontogenetic answer, will turn on the details of the interaction between the genetic make-up of particular human beings and the ambient environment in which they develop. Both questions are part of the EEM program. EET questions can be similarly partitioned. One can ask, for example, about the development of human understanding of the nature of motion from Aristotle through Descartes, Newton, Einstein, to the present. This question, in effect, is asking about the phylogeny of a particular strand of human understanding about the nature of the universe. On the other hand, one may inquire into the development of a given individual's knowledge and understanding of the nature of motion as he or she develops from child to adult. Such questions, in effect, are asking about the ontogeny of a particular strand of human understanding in particular individuals. With these distinctions in hand, it is time to turn to an examination of some representative examples of each.

EEM Phylogenetic Projects

These include attempts to reconstruct the emergence of the biological substrate that serves as the basis for sentience, cognition, and knowledge. A philosophical example of this is Popper's notorious view of three world stages (Popper 1972; Popper and Eccles 1977). Popper correctly pointed out that according to the best modern theories, the early universe was composed of matter, energy, and radiation. Before life could emerge, suitable planets had to be formed. When life did finally emerge on the planet Earth, natural selection kicked into gear, and lineages began to proliferate and diversify. At some point, organismic brains evolved the capacity for sentience. At some later point, consciousness emerged. At this point, Popper claims, there were two worlds. World 1 was purely physical. World 2 was the world of consciousness. How did consciousness evolve? No one knows, but Popper's evolutionary hypothesis is that it emerged as an adaptation that conferred selective advantages on those that possessed it. Another long period ensued until the emergence of sapient creatures capable of knowing. At this point, a World 3, the world of objective knowledge, came into being (Popper 1972; Popper and Eccles 1977). The rationale remained the same: Creatures who could command objective knowledge had a selective advantage over those who could not. The metaphysics of this view are extravagant but the sentiment is clear. Popper went on to argue that the evolutionary process in all three independent but mutually interacting worlds was the same. In the view argued for here, the evolution of life and minded creatures is a result of Darwinian evolution broadly construed. These are EEM processes. Once humans evolved to the point where they could codify and develop their knowledge of the physical world, the evolution of that understanding was no longer a matter of biological selection but involved other mechanisms that may or may not have had relevant structural similarities to evolution by natural selection. This is EET territory. A less extravagant picture along similar lines is drawn by Daniel Dennett (1995).

EEM Ontogenetic Projects

EEM ontogenetic projects are concerned with the development, in the individual organism, of the physical structures that support cognitive and epistemic activities. For much of the twentieth century, developmental biology took a back seat to evolutionary biology. The developmental pathways were too complicated to sustain fruitful investigation. In any case, the processes underlying development seemed quite different from those underlying phylogenetic evolution. This thinking has changed in recent years with the emergence of neural Darwinism as developed by G. M. Edelman, Jean-Pierre Changeaux, and their colleagues (Changeaux 1985; Edelman 1987). The basic idea involves modeling ontogenetic development using the variational and population models characteristic of evolutionary biology. In particular, the neuronal structure of the brain is no longer construed as a lockstep unfolding of instructions hardwired in the genes. Rather, a variety of neural pathways are constructed and then some are selected and others atrophy. The result is that the brain structures of different individuals, even those with identical genetic endowments, turn out to be unique.

The signature of research on EEM projects is a focus on the development of the physical and psychological mechanisms that enable organisms to gain information and knowledge about the world they live in. What they do with this information, at least in the case of human beings, is to construct bodies of knowledge that are then acquired by individuals as they mature, are communicated among individuals, and are transmitted from one generation to another. There have been a number of proposals on how to construct evolutionary models of what may be called the dynamics and kinematics of conceptual change. This leads to the arena of EET. These projects can also be divided into phylogenetic investigations into the transmission of information from one generation to another and ontogenetic investigations into the means by which individuals come to acquire and process information.

EET Phylogenetic Projects

The phylogenetic models of the growth of knowledge have tended to focus on the growth of scientific knowledge. More recently, "universal Darwinism" and the so-called science of memetics have been postulated as models of conceptual development in general (Cziko 1995). Evolutionary models of scientific change run up against a formidable objection at the onset. The growth of scientific knowledge appears to be progressive, directed, and converging on truth. Biological evolution appears to be nonprogressive, nondirected, and focused on survival and reproductive fitness. Some evolutionary models of science, notably Thomas Kuhn's and Stephen Toulmin's, bite the bullet and opt for a nonconvergent theory of the growth of science (Kuhn 1973; Toulmin 1972). Other models, notably those proposed by Popper, Campbell, and David Hull, seek to finesse these worries in various ways.

Kuhn's model portrays science as a series of periods of "normal science" punctuated by periods of scientific "revolutions." In the revolutionary stages, many of the specific methods, theories, and norms associated with the previous stage of normal science are called into question. What happens during these revolutions is something Kuhn likened to a "gestalt shift," as sometimes radically new perspectives are tried out and adopted. Near the end of *The Structure of Scientific Revolutions*, Kuhn notes the similarity to the competition among varieties that characterizes the selective processes of biological evolution (Kuhn 1973, 172f). The winning scientific perspectives are analogous to the survivors of selection. He draws the obvious conclusion: Just as biological evolution is not seen as progressing toward some global goal, so perhaps the view of scientific progress as a series of stages leading to a "permanent fixed scientific truth" should be reexamined. This was basically a throwaway line at the end of his book, but it raised a firestorm of criticism from those who saw Kuhn as a defender of an invidious relativism. Kuhn complained that he had been misunderstood. Despite the turmoil created by revolutionary stages that threatened to upend the standards and practices of normal-science traditions, there still were general scientific values such as predictive accuracy and problem-solving capacity that served as transcendent guidelines for evaluating research programs separated by a revolutionary chasm. Kuhn denied he was a relativist but did not renege on his rejection of a global sense of progress for science.

Toulmin's evolutionary model of science also rejects the unidirectionality of scientific change and the notion of global progress. Toulmin's (1972) book *Human Understanding*, the first of a projected trilogy, lays out an ambitious project for interpreting the history of ideas in terms of a form of epistemological Darwinism. The general Darwinian model of variation within populations as the material on which selection acts is, for Toulmin, just "one illustration of a more general form of historical explanation; and . . . this same pattern is applicable also, on appropriate conditions, to historical entities and populations of other kinds" (Toulmin 1972, 135). Science, in this view, develops in a two-step process, with the same structure as evolution by natural selection. At each stage in the historical development of science, a pool of intellectual variants—theories, laws, techniques/procedures, and norms—exist along with a selection process that determines which variants survive and which die out (Toulmin 1967, 465). The constraints on theory development imposed by nature are only one selective factor among many in the evolution of scientific knowledge. The net result is a picture of the evolution of scientific knowledge that provides no promise of 'progress.' The details concerning the nature of the selective forces and an explanation of how they worked were left for further volumes, which never appeared.

The roots of Popper's version of evolutionary epistemology can be found in his 1935 classic, *The Logic of Scientific Discovery*, which first appeared in English in 1959. There one finds the first glimmerings of what came to be codified as the method of "conjectures and refutations," explicitly couched in Darwinian terms. In laying out the demarcation criterion based on the falsifiability of scientific conjectures, Popper (1961) writes:

> What characterizes the empirical method is its manner of exposing to falsification, in every conceivable way, the system to be tested. Its aim is not to save the lives of untenable systems but, on the contrary, to select the one which is by comparison the fittest, by exposing them all to the fiercest struggle for survival. (42)

Later, he argues that our choice of one theory over another is a reflection of our choice of that "theory which best holds its own in competition with other theories; the one which by natural selection, proves itself the fittest to survive" (Popper 1961, 108). In 1974, he still took Darwinism to be a "metaphysical research program" (Schilpp 1974, 133–43). He later recanted biological Darwinism and cemented his Darwinian approach to epistemology (Popper 1984), arguing that "the evolution of scientific knowledge is, in the main, . . . a Darwinian process. The theories become better adapted through natural selection: they give us better and better information about reality (they get nearer and nearer to the truth)" (Popper 1984, 239). It is not clear whether he was now construing the evolutionary model of scientific change to be itself a testable and potentially falsifiable hypothesis. If so, he would be proposing a 'science' of science in the sense advocated by Hull (1988). But this reading would make Popper more of an epistemological naturalist than the textual evidence warrants.

Campbell (1974) coined the term *evolutionary epistemology* in his influential review of the literature. Campbell developed a model he called blind variation and selective retention (BVSR), which was designed to cover both biological evolution and conceptual change. Popper was very sympathetic to Campbell's view and held it to be in almost complete agreement with his own. Both views intertwine elements of the evolution of cognitive capacities (an EEM project) and the evolution of science (an EET project). The heart of Campbell's view, developed in a series of papers that began before he was aware of Popper's views, is the construal of the evolution of organisms as the result of a nested hierarchy of levels of biological and conceptual development. The key to this process is the subsumption of organismic evolution and conceptual change under the rubric of *problem solving*. So, the earliest forms of life and most basic organisms first must develop techniques for finding nourishment and sustenance. The organisms move about randomly in search of food. At the next stage, various "vicarious" sensory modalities evolve that allow for exploration of the environment without the organisms having to move into potential danger. The more advanced modalities, overlapping to an extent, include the development of habits, instincts, visually supported thought, mnemonically supported thought, "socially vicarious exploration" (including the development of observational learning and imitation), and the development of language (Munz 1993). This in turn allows for the emergence and accumulation of culture, of which the development of science is one aspect. There are several aspects of Campbell's picture that have appeared problematic to his critics. For one thing, empirical confirmation of his nested hierarchy view is yet forthcoming, as Campbell himself admitted. For another, Campbell insists that conjectures or tentative solutions to the problems faced by organisms be "blind." Some have seen this as inconsistent with the apparent intentionality of scientific research.

Hull (1988) takes a Darwinian approach to scientific change very seriously. He proposes, in effect, to develop an empirical hypothesis about the development of knowledge along selectionist lines. Rather than interpreting scientific change as *merely* analogous to biological evolution, he argues that both biological evolution and conceptual development are examples of a common selectionist structure. For his analysis, Hull borrowed a useful distinction, first introduced by Richard Dawkins (1976), between "replicators" and "vehicles." Replicators are what get handed down from one generation to the next, and vehicles are what serve as the packages containing the replicators in any given generation. Hull replaced the term "vehicle" with "interactor" to emphasize the fact that the selection forces that pick out the fittest variants work in virtue of the interaction between vehicles and their environments. For Hull, the interactors in science are the scientists themselves, who compete with one another for success. Success is measured by inclusive conceptual fitness or the measure of how widespread one's views become. What get replicated are their ideas, theories, conjectures, and methods. In contemporary science, most workers are members of a research group, or *deme*, and these groups compete with one another as well.

Hull's model, unlike Popper's, is clearly constructed along naturalistic lines. In addition, Popper's three-world view and his rejection of the *justified-true-belief* picture of knowledge leads him, unlike Hull, to downplay the role of scientific agents (the scientists) in the growth of objective knowledge. The process of conjectures and refutations that Popper sees as the core of the scientific method involves competition among *hypotheses* or *conjectures*, not among scientists. Another virtue, then, of Hull's approach over Popper's is the emphasis Hull places on the social dimension of science.

The Darwinian models of science proposed by Campbell, Popper, Toulmin, and Hull all share a commitment to a selectionist model of scientific theory change. Not all those who invoke Darwin, however, see a corresponding commitment to such a model. Nicholas Rescher argues for what he calls "methodological Darwinism," as opposed to "thesis Darwinism" (Rescher 1977 and 1990). In his view, it is methods that compete with one another for acceptance, not theses or theories. When Michael Ruse, an early skeptic about the virtues of evolutionary epistemology, changed his mind, he too rejected the idea that embracing epistemological Darwinism entails embracing a natural selection model of theory change (Ruse 1996).

EET Ontogenetic Projects

Evolutionary models of the ontogenesis of knowledge in individual organisms have also been constructed. Briefly, as Campbell (1974) has noted, these views have their roots in nineteenth-century philosophers and psychologists. B. F. Skinner's theory of operant conditioning has obvious affinities to the theory of natural selection, as he himself has noted (Skinner 1981).

Jean Piaget's extensive writings on "genetic epistemology" develop themes with clear connections to the concerns and interests of evolutionary epistemologists. His 1971 book *Biology and Knowledge: An Essay on the Relations Between Organic Regulations and Cognitive Processes* provides a useful introduction to his ideas. As the title suggests, Piaget argues that cognitive processes are rooted in and are extensions of the fundamental organic "autoregulatory" feedback processes that are the basis of organismic existence. Living organisms, he argues, are basically interactive systems that adapt to their environments by means of the "assimilation" and "accommodation" of new elements into their structural organization. These autoregulatory systems operate at all organic and psychological levels—evolutionary, ontogenetic, physiological, cognitive, and psychological. Piaget suggests that the central problem about knowledge is the relationship between subjects and objects. This relationship, in his view, corresponds directly with the relationship between organism and environment (Piaget 1971, 99). The relationship of organism to environment is an interactive one. Transposed to the realm of knowledge, this undercuts, in Piaget's view, any "copy theory" of knowledge, whereby whole bodies of knowledge are acquired, communicated, and transmitted intergenerationally (see "EEM Ontogenetic Projects" above). Knowledge, for Piaget, is active and regulatory. He thus urges the development of appropriate cybernetic models that incorporate both Darwinian and Lamarckian elements. Piaget's work in evolutionary epistemology has not received the attention it deserves.

More recently, Henry Plotkin, Susan Blackmore, and others have begun to argue for a "science of memetics," based on the idea that conceptual change can be modeled as the differential replication of cultural units, or "memes" (Plotkin 1993; Blackmore 1999). Blackmore, in particular, argues that memetic evolution has become decoupled from genetic evolution. Memes were created to enhance the fitness of these vehicles, but once generated they take on a life of their own and become replicators in their own right. Blackmore sees implications for both the evolution of culture and the development of the big brains necessary to create culture in the first place. These programs cut across all the distinctions drawn here and have implications for the ontogeny and phylogeny of both epistemic mechanisms and the conceptual systems that they produce. It is too early to pass judgment on this project, but should it prove fruitful, it has implications for a general analysis of the evolution of culture.

Evolutionary Epistemology and the Tradition

The evolutionary approach to epistemology is most closely allied with naturalistic approaches to epistemology. The focus is on the biological conditions of knowing and the dynamics of conceptual change. Therefore, it has seemed to some critics to be "epistemology" in name only. Critics have charged that evolutionary epistemology fails to address the traditional normative issues, such as the nature of justification and the reliability of evidence. Thus, it is beside the point or involves changing the question. There is no doubt that evolutionary approaches to epistemology entail a radical reevaluation of what it means to do "proper epistemology." It is appropriate to note here that John Dewey argued that one of the consequences of taking Darwin seriously would be to restructure the kinds of questions that philosophers ask and the kinds of answers they deem appropriate (Dewey 1910). Not all are prepared to be so accommodating. In Jaegwon Kim's (1988) view, if epistemologists abandon the task of providing justifications, they have abandoned epistemology. Campbell's approach was to argue that evolutionary epistemology was "descriptive" and hence complementary to the traditional normative approach. Others stand ready to abandon the tradition and the search for justifications altogether (Radnitsky 1987). Hull concurs in part, although he allows a role for contextually articulated epistemic norms that arise from the practice of science itself. If, on the other hand, norms are construed as instrumental procedural and methodological rules, then a selectionist account can be given of them. In the marketplace of ideas, those rules that promote the development of successful strategies for coping with the environment, including the development of successful scientific theories and inferential practices, will be at a selective advantage over those that do not. The norms that emerge are justified by the fact that their deployment does lead to successful practices. From a naturalistic and evolutionary standpoint, one can ask for nothing more.

Students of the human epistemic condition stand at a fork in the road. In one direction lies the tradition that denies the relevance to "real" epistemology of any or most of the considerations discussed above. In the other lie research projects that seek to integrate the latest work on evolutionary biology, psychology, and computer modeling into a philosophically sophisticated understanding of the nature of knowledge, how it is acquired, and how it is transmitted. (Those wishing to pursue these latter issues should see the extensive bibliography in Cziko and Campbell 1997.)

Michael Bradie

References

Barkow, Jerome H., Leda Cosmides, and John Tooby (eds.) (1992), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Blackmore, Susan (1999), *The Meme Machine*. Oxford: Oxford University Press.

Bradie, Michael (1986), "Assessing Evolutionary Epistemology," *Biology and Philosophy* 4: 401–459.

——— (1994), "Epistemology from an Evolutionary Point of View," in Elliott Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 453–475.

Campbell, D. T. (1974), "Evolutionary Epistemology," in P. A. Schilpp (ed.), *The Philosophy of Karl Popper*. LaSalle, IL: Open Court.

Changeux, Jean-Pierre (1985), *Neuronal Man: The Biology of Mind*. New York: Pantheon Books.

Cziko, Gary (1995), *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution*. Cambridge, MA: MIT Press.

Cziko, Gary, and D. T. Campbell (1997), *Selection Theory Bibliography*. http://faculty.ed.uiuc.edu/g-cziko/stb/Default.asp?bhcd2=1033275860

Dawkins, Richard (1976), *The Selfish Gene*. Oxford: Oxford University Press.

Dennett, Daniel (1995), *Darwin's Dangerous Idea*. New York: Simon & Schuster.

Dewey, John (1910), *The Influence of Darwin on Philosophy and Other Essays in Contemporary Thought*. New York: Henry Holt & Co.

Edelman, Gerald M. (1987), *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.

Hull, David (1988), *Science As a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.

Kim, Jaegwon (1988), "What Is 'Naturalized Epistemology'?" in James E. Tomberlin (ed.), *Philosophical Perspectives 2. Atascadero*, CA: Ridgeview Publishing, 381–405.

Kuhn, Thomas S. (1973), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Mayr, Ernest (1961), "Cause and Effect in Biology," *Science* 134: 1501–1506.

Munz, Peter (1993), *Philosophical Darwinism: On the Origin of Knowledge by Means of Natural Selection*. London: Routledge.

Piaget, Jean (1971), *Biology and Knowledge: An Essay on the Relations Between Organic Regulations and Cognitive Processes*. Translated by Beatrix Walsh. Chicago: University of Chicago Press.

Plotkin, Henry (1993), *Darwin Machines and the Nature of Knowledge*. Cambridge, MA: Harvard University Press.

Popper, Karl (1961), *The Logic of Scientific Discovery*. New York: Science Editions.

——— (1972), *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.

——— (1984), "Evolutionary Epistemology," in J. W. Pollard (ed.), *Evolutionary Theory: Paths into the Future*. London: John Wiley & Sons Ltd.

Popper, Karl, and John Eccles (1977), *The Self and Its Brain*. New York: Springer International.

Radnitzky, Gerard, and W. W. Bartley (1987), *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*. La Salle, IL: Open Court.

Rescher, Nicholas (1977), *Methodological Pragmatism*. Oxford: Basil Blackwell.

——— (1990), *A Useful Inheritance: Evolutionary Aspects of the Theory of Knowledge*. Savage, MD: Rowman & Littlefield.

Ruse, Michael (1996), *Taking Darwin Seriously*. Oxford: Basil Blackwell.

Skinner, B. F. (1981), "Selection by Consequences," *Science* 213: 501–504.

Toulmin, Stephen (1967), "The Evolutionary Development of Natural Science," *American Scientist* 55: 456–471.

——— (1972), *Human Understanding: The Collective Use and Evolution of Concepts*. Princeton, NJ: Princeton University Press.

See also **Evolutionary Psychology; Naturalism**

EVOLUTIONARY PSYCHOLOGY

The evolutionary study of the mind in the twentieth century has been marked by three self-conscious movements: classical ethology, sociobiology, and Evolutionary Psychology (capitalized to indicate that it functions here as a proper name). Classical ethology was established in the years immediately before the Second World War, primarily by Konrad Lorenz and Niko Tinbergen (Burckhardt 1983). Interrupted by the war, the movement blossomed in the early 1950s, when ethologists established major research institutes in most developed countries and had a major impact on the broader culture through popular science writing. From the outset, ethology sought to apply its methods for the comparative study of animal behavior to human beings, something that was especially prominent in popular works written by ethologists. Lorenz's *On Aggression* (1966a) is perhaps the best known of these, but several other ethologists wrote books advocating the application of the new evolutionary science of the mind to problems of international conflict and social unrest. The ethologist who focused most on human beings in his empirical research was Lorenz's student Irenaus Eibl-Eibesfeldt, who throughout the 1960s and 1970s sought to document innate, universal behavior patterns in *Homo sapiens* through photography and film (Eibl-Eibesfeldt 1989).

Classical ethology was largely displaced in the 1970s by sociobiology, a movement that sought to apply to humans a set of new mathematical techniques for the study of animal behavior (Wilson 1975). During the 1960s behavioral ecologists had come to view animal behaviors primarily as strategies adopted in competitions among and within species. Models of these competitive interactions could be constructed using evolutionary game theory, and the predictions of these models could be tested against actual behavior. Animal behaviors were expected to correspond to "evolutionarily stable strategies," that is, to equilibria in the relevant game-theoretic models. The game-theoretic approach had the advantage that it did not require knowledge of the neural mechanisms underlying behavior (or the genetic mechanisms underlying its transmission). The early ethologists' "hydraulic model" of neural mechanisms had collapsed during the 1950s as it became clear that the prewar neuroscience on which the model was based had not been borne out by further investigation. The hydraulic model also failed to accommodate many of the new behavioral phenomena uncovered as ethology matured (Hinde 1956). No new model of similar generality was available to replace the hydraulic model and its relatives, making a method that dealt directly with behavior highly desirable.

Sociobiologists also argued that their approach was intrinsically more scientific than classical ethology because it made predictions about behavior and tested them, rather than merely describing behavior and explaining it. This led to the hope that evolutionary models could guide psychological research and point it toward important phenomena that would otherwise be misunderstood or overlooked, an idea that remains central to today's Evolutionary Psychology, which includes advocates of sociobiology among its leading figures. Among them is Jerome Barkow (1979), who expressed this viewpoint succinctly in his title "Classical Ethology: Empirical Wealth, Theoretical Dearth." But despite such oppositional rhetoric, there was considerable continuity of practice and personnel between ethology and sociobiology. Richard Dawkins and Desmond Morris, for example, key figures in the popularization of sociobiology, were students of Niko Tinbergen and regarded sociobiology as a continuation of the tradition he had established (see the introduction to Dawkins, Halliday, and Dawkins 1991).

At the end of the 1980s sociobiology itself came under attack from a new movement calling itself Evolutionary Psychology (Barkow, Cosmides, and Tooby 1992; Crawford, Smith, and Krebs 1987). The Evolutionary Psychologists argued that the whole project of explaining contemporary human behaviors as a direct result of adaptive evolution was misguided (Symons 1992). The contemporary environment is so different from that in which human beings evolved that their behavior probably bears no resemblance to the behavior that was important in evolution. This problem had been identified by many of the best-known critics of sociobiology (e.g., Kitcher 1985), but Evolutionary Psychology followed it up with a positive proposal. Evolutionary theory should be used to predict which behaviors *would have been* selected in postulated ancestral environments. Human behavior today can be explained as the output of mechanisms that evolved to produce those ancestral behaviors when these mechanisms operate in their very different, modern environment. Furthermore, the diverse behaviors seen in different cultures may all be manifestations of a single, evolved psychological mechanism operating under a range of local conditions, an idea that originated in an offshoot of sociobiology known as Darwinian anthropology (Alexander 1979). Refocusing research on the Darwinian algorithms that underlie observed behavior, rather than on the behavior itself, lets the Evolutionary Psychologist "see through" the interfering effects of environmental change and cultural difference to an underlying human nature (see Natural Selection).

Adherents of today's Evolutionary Psychology normally present their approach as something very novel, typically describing it as "the new science of the mind" (Cosmides and Tooby 2001). They allege that the social and behavioral sciences have until recently been dominated by the "standard social science model" (SSSM), which denies the existence of any evolved features of the mind. The SSSM grew out of the liberal political agendas of the 1960s, which aimed to change traditional social behavior: '"Not so long ago jealousy was considered a pointless, archaic institution in need of reform. But like other denials of human nature from the 1960s, this bromide has not aged well," as Stephen Pinker puts it on the dustjacket of a work of Evolutionary Psychology (Buss 2000). But the claim that Evolutionary Psychology is a rebellion against an antibiological consensus in the social and behavioral sciences is at best a considerable exaggeration. Instead, Evolutionary Psychology represents the latest stage of a tradition of evolutionary psychology dating back at least to Lorenz. Nor is the public prominence of Evolutionary Psychology entirely new. Lorenz was as successful a popular author in the 1950s and 60s as Richard Dawkins was in the 1970s and 80s. Furthermore, in some important respects, Evolutionary Psychology actually represents a return to the positions of classical ethology. Classical ethologists thought that modern human behavior was the (often maladaptive) result of ancient, evolved mechanisms operating in radically new environments. They also shared the "modular" conception of the mind, described below. Most importantly, classical ethology and Evolutionary Psychology offer very similar critiques of conventional psychology. Lorenz's complaints were directed against those he liked to call "American behaviorists." The laboratory-based search for general laws of learning seemed to him as misguided as dropping automobiles from buildings under controlled conditions and writing down the results. Without an evolutionary perspective, he argued, psychology does not know what it is looking for, and when it finds something, it does not know what it is looking at (e.g., Lorenz 1966b, 274). In the same way, advocates of Evolutionary Psychology argue that empirical psychology without an evolutionary perspective has no way to determine whether it is studying meaningful units of behavior or mental functioning:

> Cognitive scientists will make far more rapid progress in mapping this evolved architecture if they begin to seriously incorporate knowledge from evolutionary biology and its related disciplines . . . into their repertoire of theoretical tools, and use theories of adaptive function to guide their empirical investigations. (Tooby and Cosmides 1998, 195)

Evolutionary Psychology and Cognitive Science

The classical ethologists based their ideas about mental mechanisms on the neuroscience of the interwar years. Similarly, Evolutionary Psychology has turned to "classical" cognitive science, with its guiding idea that the mind is computer software implemented in neural hardware (Fodor 1983; Marr 1982). Evolutionary Psychology argues that the representational, information-processing language of classical cognitive science is ideal for describing the evolved features of the mind. Behavioral descriptions of what the mind does are useless because of the problem of changing environments, described above. Neurophysiological descriptions are inappropriate, because behavioral ecology does not predict anything about the specific neural structures that underlie behavior. Models in behavioral ecology predict which behaviors would have been selected in the ancestral environment, but they cannot distinguish between different mechanisms that produce the same behavioral output. Hence, if one accepts the conventional view in cognitive science that indefinitely many different neural mechanisms could potentially support the same behavior, it follows that behavioral ecology predicts nothing about the brain except which information-processing functions it must be able to perform:

When applied to behavior, natural selection theory is more closely allied with the cognitive level of explanation than with any other level of proximate causation. This is because the cognitive level seeks to specify a psychological mechanism's function, and natural selection theory is a theory of function (Cosmides and Tooby 1987, 284). It is thus slightly confusing that Evolutionary Psychologists talk of discovering psychological "mechanisms," a term that suggests theories at the neurological level. What 'mechanism' actually refers to in this context is a performance profile—an account of what output the mind will produce given a certain range of inputs (see Cognitive Science).

The fact that evolutionary reasoning yields expectations about the performance profile of the mind fits neatly with the explanatory framework of classical cognitive science. According to the influential account by David Marr (1982), explanation in cognitive science works at three mutually illuminating levels. The highest level concerns the tasks that the cognitive system accomplishes—for example, recovering the shape and position of objects from stimulation of the retina. The lowest level concerns the neurophysiological mechanisms that accomplish that task—the neurobiology of the visual system. The intermediate level concerns the functional profile of those mechanisms, or as it is more usually described, the computational process that is implemented in the neurophysiology. Hypotheses about the neural realization of the computational level constrain hypotheses about computational processes: Psychologists should propose only computational models that can be realized by neural systems. Conversely, hypotheses about computational processes guide the interpretation of neural structure: Neuroscience should look for structures that can implement the required computations. Similar relations of mutual constraint hold between the level of task description and the level of computational processes. But there remains something of a puzzle as to how the highest level—the task description—is to be specified other than by stipulation. It seems obvious that the task of vision is to represent things around us, but what makes this true? According to Evolutionary Psychology, claims about task descriptions are really claims about evolution. The overall task of the mind is survival and reproduction in the ancestral environment, and the subtasks performed by parts of the mind correspond to separate adaptive challenges posed by the ancestral environment. For example, it would have been useful for the ancestors of humans to be able to see, so it is predictable that humans will have a visual system. This kind of thinking becomes useful when the function of a psychological mechanism is not as blindingly obvious as in the case of vision. What, for example, is the task description for the emotional system, or for individual emotions such as jealousy or grief? Evolutionary Psychology argues that in such cases it should be evolutionary thinking that sets the agenda for cognitive science, telling it what to look for and how to interpret what it finds.

The Massive Modularity Thesis

One of the best-known features of Evolutionary Psychology is the massive modularity thesis, or the "Swiss army knife" model, according to which the mind contains few if any general-purpose cognitive mechanisms. The mind is a collection of separate modules, each designed to solve a specific adaptive problem, such as mate recognition or the enforcement of female sexual fidelity. The flagship example of a mental module is the *language acquisition device*, the mechanism that allows human infants to acquire a language in a way that, it is widely believed, would not be possible using any general-purpose learning rules (Pinker 1994). Other well-known examples include the perceptual input devices for which the modularity concept was originally introduced (Fodor 1983). The massive modularity thesis is an example of the kind of evolutionary guidance for cognitive science described in the last section. Evolutionary Psychology argues that evolution would favor multiple modules over domain-general cognitive mechanisms because each module can be fine-tuned for a specific adaptive problem. So, cognitive scientists should look for domain-specific effects in cognition and conceptualize their work as the search for and characterization of mental modules.

The Monomorphic Mind Thesis

The leading Evolutionary Psychologists John Tooby and Leda Cosmides have argued strongly for the monomorphic mind thesis, or "psychic unity of humankind" (Cosmides, Tooby, and Barkow 1992, 72). This thesis states that any differences that exist in the cognitive adaptations of individual humans or human groups are not due to genetic differences. Psychological differences are always, or almost always, due to environmental factors that trigger different aspects of the same developmental program. If true, this would make cognitive adaptations highly atypical, since most human traits display considerable individual variation related to differences in genotype. All human beings have eyes, but these eyes exhibit differences in color, size, shape, acuity, and susceptibility to various forms of degeneration over time, all due to differences in genotype. It has been known for half a century that wild populations of most species contain substantial genetic variation, and humans are no exception.

Tooby and Cosmides (1990) offer one main argument for the conclusion that the genes involved in producing cognitive adaptations will be the same in all human individuals:

> Complex adaptations necessarily require many genes to regulate their development, and sexual recombination makes it combinatorially improbable that all the necessary genes for a complex adaptation would be together at once in the same individual, if genes coding for complex adaptations varied substantially between individuals. Selection, interacting with sexual recombination, enforces a powerful tendency towards unity in the genetic architecture underlying complex functional design at the population level and usually the species level as well. (393)

The authors apply this argument to only psychological adaptations, but its logic extends to all traits with many genes involved in their etiology. The argument fails because it assumes that development is a mechanical consequence of the exact sequence of genes on each chromosome. What Cosmides and Tooby seem to have overlooked is the phenomenon described by C. H. Waddington in the 1940s as "developmental canalization": Development is buffered against genetic variation, as well as against environmental variation (Waddington 1959). This is why surprisingly many gene knock-out experiments produce negative results. Disabling a gene known to be involved in a developmental pathway frequently produces no effect (a null phenotype), because development contains positive and negative feedback mechanisms that increase transcription of the required gene product from the other allele, initiate transcription from another gene copy, or initiate transcription of a different gene product that can produce the same outcome (Freeman 2000). On a larger scale, it has become a commonplace amongst evolutionary developmental biologists that complex phenotypic features of organisms can be conserved over evolutionary time despite changes in the specific genes used to construct them and even in the general form of the developmental pathway by which they are constructed (Raff 1996; Wagner 1994) (see Developmental Biology).

One reason for the popularity of the doctrine of the monomorphic mind is probably as a bulwark against racism. If all human beings have substantially the same genes, then racial differences are superficial and modifiable. But no such bulwark is necessary. If it is assumed that variation in evolved human phenotypes roughly mirrors the known variation in human genotypes, then it follows that the vast majority of traits are pancultural and that the differences among human groups are dwarfed by the differences among individuals within those groups (Cavalli-Sforza, Menozzi, and Piazza 1994).

Alternatives to Evolutionary Psychology

The Evolutionary Psychology movement has been as controversial as it has been successful. Jerry Fodor, one of the originators of Evolutionary Psychology's preferred framework for cognitive science, rejects the massive modularity thesis and has expressed considerable skepticism about the value of evolutionary thinking as a heuristic for cognitive science (Fodor 2000). Many researchers accept that evolutionary thinking can and should transform psychology and cognitive science but disagree, often radically, with Evolutionary Psychology's specific program for accomplishing this transformation. Several recent collections of papers present the views of such evolutionary psychologists (Heyes and Huber 2000; Holcomb 2001; Scher and Rauscher 2002). Finally, a "developmentalist" tradition in animal behavior research with its roots in classical ethology and comparative psychology has criticized both sociobiology and Evolutionary Psychology for failing to integrate the Darwinian study of behavior with the study of how behavior develops (Gottlieb 1997; Bjorklund and Pellegrini 2002). Accessible introductions to this tradition are provided by Patrick Bateson and Paul Martin (1999) and David Moore (2001).

Paul E. Griffiths

References

Alexander, R. (1979), *Darwinism and Human Affairs*. Seattle: Washington University Press.

Barkow, J. H. (1979), "Human Ethology: Empirical Wealth, Theoretical Dearth," *Behavioral and Brain Sciences* 2: 27.

Barkow, J. H., L. Cosmides, and J. Tooby (eds.) (1992), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press.

Bateson, P. P. G., and P. Martin (1999), *Design for a Life: How Behavior and Personality Develop*. London: Jonathan Cape.

Bjorklund, D. F., and A. D. Pellegrini (2002), *The Origins of Human Nature: Evolutionary Developmental Psychology*. Washington DC: American Psychological Association.

Burckhardt, R. W. (1983), "The Development of an Evolutionary Ethology," in D. S. Bendall (ed.), *Evolution: From Molecules to Men*. Cambridge: Cambridge University Press, 429–444.

Buss, D. M. (2000), *The Dangerous Passion: Why Jealousy Is As Essential As Love and Sex*. New York: Simon and Schuster.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza (1994), *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Cosmides, L., and J. Tooby (1987), "From Evolution to Behaviour: Evolutionary Psychology As the Missing Link," in J. Dupré (ed.), *The Latest on the Best: Essays on Optimality and Evolution*. Cambridge, MA: MIT Press, 277–307.

——— (2001), *What Is Evolutionary Psychology? Explaining the New Science of the Mind*. New Haven, CT: Yale University Press.

Cosmides, L., J. Tooby, and J. H. Barkow (1992), "Introduction: Evolutionary Psychology and Conceptual Integration," in J. H. Barkow, L. Cosmides, and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford and New York: Oxford University Press, 3–15.

Crawford, C., M. Smith, and D. Krebs (eds.) (1987), *Sociobiology and Psychology: Ideas, Issues and Applications*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dawkins, M. S., T. R. Halliday, and R. Dawkins (eds.) (1991), *The Tinbergen Legacy*. London: Chapman and Hall.

Eibl-Eibesfeldt, I. (1989), *Human Ethology*. New York: Aldine de Gruyter.

Fodor, J. A. (1983), *The Modularity of Mind: An Essay in Faculty Psychology*. Cambridge, MA: Bradford Books/MIT Press.

——— (2000), *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

Freeman, M. (2000), "Feedback Control of Intercellular Signaling in Development," *Nature* 408: 313–319.

Gottlieb, G. (1997), *Synthesizing Nature-Nurture: Prenatal Roots of Instinctive Behavior*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Heyes, C. M., and L. Huber (2000), *The Evolution of Cognition*. Cambridge, MA: MIT Press.

Hinde, R. A. (1956), "Ethological Models and the Concept of 'Drive,'" *British Journal for the Philosophy of Science* 6: 321–331.

Holcolmb, H. H. III. (ed.) (2001), *The Evolution of Minds: Psychological and Philosophical Perspectives*. Dordrecht, Netherlands: Kluwer.

Kitcher, P. (1985), *Vaulting Ambition*. Cambridge, MA: MIT Press.

Lorenz, K. (1966a), *On Aggression*. Translated by M. K. Wilson. New York: Harcourt, Brace and World.

——— (1966b), "Evolution of Ritualisation in the Biological and Cultural Spheres," *Philosophical Transactions of the Royal Society of London* 251: 273–284.

Marr, D. (1982), *Vision*. New York: W. H. Freeman.

Moore, D. S. (2001), *The Dependent Gene: The Fallacy of "Nature versus Nurture"*. New York: W. H. Freeman/Times Books.

Pinker, S. (1994), *The Language Instinct: The New Science of Language and Mind*. New York: William Morrow.

Raff, R. (1996), *The Shape of Life: Genes, Development and the Evolution of Animal Form*. Chicago: University of Chicago Press.

Scher, S., and M. Rauscher (2002) (eds.), *Evolutionary Psychology: Alternative Approaches*. Dordrecht, Netherlands: Kluwer.

Symons, D. (1992), "On the Use and Misuse of Darwinism in the Study of Human Behavior," in J. H. Barkow, L. Cosmides, and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press, 137–159.

Tooby, J., and Cosmides, L. (1990), "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments," *Ethology and Sociobiology* 11: 375–424.

——— (1998), "Evolutionizing the Cognitive Sciences: A Reply to Shapiro and Epstein," *Mind and Language* 13: 195–204.

Waddington, C. H. (1959), "Canalisation of Development and the Genetic Assimilation of Acquired Characters," *Nature* 183: 1654–1655.

Wagner, G. P. (1994), "Homology and the Mechanisms of Development," in B. K. Hall (ed.), *Homology: The Hierarchical Basis of Comparative Biology*. New York: Academic Press, 273–299.

Wilson, E. O. (1975), *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press.

See also **Adaptation and Adaptionism; Cognitive Science; Evolution; Evolutionary Epistemology; Natural Selection; Naturalism**

EXPERIMENT

Over the past two decades the historical development of experimental science has been studied in detail. One focus has been on the nature and role of experiment during the rise of the natural sciences in the sixteenth and seventeenth centuries. Earlier accounts of this so-called Scientific Revolution emphasized the universalization of the mathematical method or the mechanization of the worldview as the decisive achievement. In contrast, the more recent studies of sixteenth- and seventeenth-century science stress the great significance of a new experimental practice and a new experimental knowledge. Major figures were Bacon, Galileo, and Boyle. The story of the controversy of the latter with Hobbes has been made a paradigm of the recent history of scientific experimentation by Shapin and Schaffer (1985). In this controversy, the legitimacy of experiment as a way to knowledge was at issue. While Hobbes defended the "old" axiomatic-deductive style of the geometric tradition, Boyle advocated the more modest acquisition of probable knowledge of experimental "matters of fact." According to Shapin and Schaffer, what was at stake were simultaneously the technical details of Boyle's air-pump experiments, the epistemological justification of experimental knowledge, and the social legitimacy of the new experimental style of doing science. While Shapin and Schaffer emphasize the novelty of sixteenth- and seventeenth-century experimentation, some others have questioned this claim by arguing that Hellenistic antiquity, the Arab world, and the Middle Ages all have significant experimental traditions—for instance, in the areas of optics and alchemy.

A more wide-ranging account of the role of experimentation in the emerging natural sciences has been proposed by Thomas Kuhn (see Kuhn, Thomas). He argues that the rise of modern physical science resulted from two simultaneous developments (Kuhn 1977). On the one hand, a radical conceptual and worldview change occurred in what he calls the classical, or mathematical, sciences, such as astronomy, statics, and optics. On the other, the novel type of Baconian, or experimental, sciences emerged, dealing with the study of light, heat, magnetism, and electricity, among other things (see Scientific Change). An important additional claim put forward by Kuhn is that it was not before the second half of the nineteenth century that a systematic interaction and merging of the experimental and mathematical traditions took place. An example is the transformation of the Baconian science of heat into an experimental-mathematical thermodynamics during the first half of the nineteenth century. At about the same time, the interactions between (at first, mainly experimental) science and technology increased substantially. Important results of this scientification of technology were chemical dye stuffs and artificial fertilizers.

Starting in the second half of the nineteenth century, systematic experimentation also took root in various other sciences. This happened in medicine, in particular in physiology, somewhat later in psychology, and still later in the social sciences. A characteristic feature of many experiments in those sciences is a strong reliance on statistical methods. Thus far, most philosophers of science have focused on experimentation in physics, chemistry, and biochemistry, while many analyses of statistical experiments can be found in the methodological literature of the medical, psychological, and social sciences (see e.g., Campbell and Stanley 1963). In this article, the focus will be on the philosophical approach to experimentation.

The Philosophy of Scientific Experimentation

A central feature of experimentation is the manipulation of, and the interference with, material things. Historians and philosophers, however, have focused on science as mostly a theoretical activity, a matter of thinking and reasoning. Thus, although the logical empiricists acknowledged the importance of observation and experiment, they took these activities mostly for granted and concentrated their studies on the philosophical problems of theories and theoretical knowledge (see Logical Empiricism).

Yet, historically some authors—including scientists and philosophers—did write about the nature and function of scientific experimentation. Among the better known examples are Bacon's and Galileo's advocacy of the experimental method. Mill (around the middle of the nineteenth century) and Mach (late nineteenth and early twentieth centuries) provided some methodological and epistemological analyses of experimentation. Bernard promoted and analyzed the use of the experimental method in medicine. His *Introduction to the Study of Experimental Medicine* (Bernard [1865] 1957) influenced a number of twentieth-century French writers, including Duhem, Bachelard, and Canguilhem. While those authors addressed some aspects of experimentation in their accounts of science, a substantial and coherent tradition in the philosophy of scientific experimentation did not yet arise.

Such a tradition did spring up in Germany, in the second half of the twentieth century. Within this German tradition, two approaches may be distinguished. One developed the pioneering work of Dingler (1928), who emphasized the action and production character of experimentation, and hence its kinship to technology. One of the aims of his operationalist approach (see Bridgman, Percy) was to show how the basic theoretical concepts of physics, such as length and mass, could be grounded in concrete experimental actions. This part of Dingler's philosophy was taken up and systematically developed by a number of other German philosophers, including Lorenzen, Holzkamp, and Janich. More recently, the emphasis on the methodical construction of theoretical concepts in terms of experimental actions has given way to more self-contained accounts of scientific experimentation and a more culturalistic interpretation of its results (see Janich 1996).

A second approach within the German tradition took its departure even more directly from the kinship between experiment and technology. The major figure here is the early Habermas. In his work from the 1960s, Habermas ([1968] 1978) conceived of (empirical-analytical) science as "anticipated technology," the crucial link being experimental action. In the spirit of Marx, Heidegger, and Marcuse, Habermas's aim was to develop not merely a theory of (scientific) knowledge but rather a critique of technocratic reason. More recently, attempts have been made to connect this German tradition to Anglo-Saxon philosophy of experiment (Radder 1996, Ch. 2) and to contemporary social studies of science and technology (Feenberg 1999). Recent work on science as technology by Lelas (2000) can be characterized as broadly inspired by this second branch of the German tradition.

In the English-speaking world, a substantial number of studies of scientific experimentation have been written since the mid-1970s. They resulted from the Kuhnian "programs in history and philosophy of science" (see Kuhn, Thomas). In their studies of (historical or contemporary) scientific controversies, sociologists of scientific knowledge often focused on experimental work (e.g., Collins 1985), while so-called laboratory studies addressed the ordinary practices of experimental scientists (e.g., Latour and Woolgar 1979). An approach that remained more faithful to the history and philosophy of science started with Hacking's argument for the relative autonomy of experimentation and his plea for a philosophical study of experiment as a topic in its own right (Hacking 1983). It includes work by Franklin, Galison, Gooding, and Rheinberger, among many others (see the volumes edited by Gooding, Pinch, and Schaffer 1989; Buchwald 1995; Heidelberger and Steinle 1998).

More recently, some philosophers argue that a further step should be taken by combining the results of the empirical and historical study of experiment with more developed theoretical-philosophical analyses (see Radder 2003). A mature philosophy of experiment, they claim, should not be limited to summing up its empirical features but should attempt to provide a systematic analysis of experimental practice and experimental knowledge. The latter is often lacking in the sociological and historical literature on scientific experimentation.

Action and Production and Their Philosophical Implications

Looking at the role of experiments within the overall practice of science, there is one feature that stands out. In order to perform experiments, whether they are large-scale or small-scale, experimenters have to intervene actively in the material world; moreover, in doing so they produce all kinds of new objects, substances, phenomena, and processes. More precisely, experimentation involves the material realization of the experimental system (that is to say, the object[s] of study, the apparatus, and their interaction) as well as an active intervention in the environment of this system. In this respect, experiment contrasts with theory even if theoretical work is always attended with material acts (such as the typing or writing down of a mathematical formula). Hence, a central issue for a philosophy of experiment is the question of the nature of experimental action and production, and their ontological, epistemological, and methodological implications.

Clearly, not just any kind of intervention in the material world counts as a scientific experiment. Quite generally, one may say that successful experiments require, at least, a certain stability and reproducibility, and meeting this requirement presupposes a measure of control of the experimental system and its environment as well as a measure of discipline of the experimenters and the other people involved in realizing the experiment.

Experimenters employ a variety of strategies for producing stable and reproducible experiments (see e.g., Bhaskar 1978; Franklin 1986; Janich 1996; Radder 1996). One such strategy is to attempt to realize "pure cases" of experimental effects. For example, in some early electromagnetic experiments carried out in the 1820s, Ampère investigated the interaction between an electric current and a freely suspended magnetic needle. He systematically varied a number of factors of his experimental system and examined whether or not they were relevant, that is to say, whether they had a destabilizing impact on the experimental process.

Furthermore, realizing a stable object/apparatus system requires knowledge and control of the (actual and potential) interactions between this system and its environment. Depending on the aim and design of the experiment, these interactions may be *necessary* (and hence required), *permitted* (but irrelevant), or *undesirable* (because disturbing). Thus, in his experiments on electromagnetism, Ampère anticipated a potential disturbance exerted by the magnetism of the Earth. In response, he designed his experiment in such a way that terrestrial magnetism constituted a permitted rather than a disturbing interaction.

A further aspect of experimental stability is implied by the notion of reproducibility. Investigating the questions of *what* should be reproducible and *by whom* leads to different types of experimental reproducibility, which can be observed to play different roles in experimental practice. A successful application of the strategy of reproducing an experiment is an achievement that may depend on certain idiosyncratic aspects of a local situation. Yet, a purely local experiment that cannot be carried out by other experimenters and in other experimental contexts will, in the end, be unproductive for science.

Laboratory experiments in physics, chemistry, and biochemistry often allow one to control the objects under investigation to such an extent that the relevant objects in successive experiments may be assumed to be in identical states. Hence, when statistical methods are employed, it is primarily to further analyze or process the data (see e.g., the error-statistical approach in Mayo 1996) (see Statistics, Philosophy of). In contrast, in field biology, medicine, psychology, and social science such a strict experimental control is often not feasible. To compensate for this, statistical methods in these areas are used directly to construct groups of experimental subjects that are presumed to possess identical average characteristics. It is only after such groups have been constructed that one can start the investigation of hypotheses about the research subjects. One can phrase this contrast in a different way by saying that in the former sciences, statistical considerations bear mostly upon linking experimental data and theoretical hypotheses, while in the latter, it is often the case that statistics play a role already at the stage of producing the actual individual data (see Psychology, Philosophy of; Social Sciences, Philosophy of the).

The action and production aspect of scientific experimentation carries implications for ontological and epistemological questions. A general ontological lesson, already drawn by Bachelard, appears to be this: The action and production character of experimentation entails that the actual objects and phenomena themselves are, at least in part, materially realized through human intervention. Hence, it is not just the knowledge of experimental objects and phenomena but also their actual existence and occurrence that prove to be dependent on specific, productive interventions by the experimenters. This fact gives rise to a number of important philosophical issues. If experimental objects and phenomena have to be realized through active human intervention, does it still make sense to speak of a "natural" nature, or does one merely deal with artificially produced laboratory worlds? If one does not want to endorse a full-fledged constructivism, according to which the experimental objects and phenomena are nothing but artificial, human creations (see Social Constructionism), one needs to go beyond an actualist ontology and introduce more differentiated ontological categorizations (see Realism). In this spirit, various authors (e.g., Bhaskar 1978) have argued that an adequate ontological interpretation of experimental science needs some kind of dispositional concepts, such as powers, potentialities, or tendencies. These human-independent dispositions would then enable the human construction of particular experimental processes.

Next to such ontological problems, the interventionist character of experimentation engenders a number of epistemological questions. At least for those who assume the ontological independence of nature, a further important question is whether scientists, on the basis of artificial experimental intervention, can acquire knowledge of a human-independent nature (see also Epistemology; Realism). Some philosophers claim that at least in a number of philosophically significant cases, such "back inferences" from the artificial laboratory experiments to their natural counterparts can be justified. Another approach accepts the constructed nature of much experimental science but stresses the fact that its results acquire a certain endurance and autonomy with respect to both the context in which they have been realized in the first place and later developments. In this vein, Baird (2004) offers a neo-Popperian account of "objective thing knowledge," the knowledge encapsulated in material things, such as Watson and Crick's material double-helix model or the indicator of Watt and Southern's steam engine (see Popper, Karl Raimund).

Another epistemologically relevant feature of experimental science is the distinction between the working of an apparatus and its theoretical accounts. In actual practice it is often the case that experimental devices work well, even if scientists disagree on how they do so. This fact supports the claim that variety and variability at the theoretical and ontological levels may well go together with a considerable stability at the level of the material realization of experiments. This claim can then be exploited for philosophical purposes—for example, to vindicate entity realism (Hacking 1983) or referential realism (Radder 1996).

At times, scientists devise and discuss so-called thought experiments (see Brown 1991). Such experiments—in which the crucial aspect of action and production is missing—are better conceived as not being experiments at all but rather as particular types of theoretical argument, which may or may not be materially realizable in experimental practice. Furthermore, recent scientific practice shows an ever-increasing use of "computer experiments." These involve various sorts of hybrids of material intervention, computer simulation, and theoretical and mathematical modeling techniques. Often, more traditional experimental approaches are challenged and replaced by approaches based fully or primarily on computer simulations (sometimes this replacement is based on budgetary considerations only). This development raises important questions for the philosophy of scientific experimentation. Prominently, there is the epistemological question of the justifiability of the results of the new approaches. Should experiments that involve a substantial material component remain the standard, or are simulated experiments equally reliable and useful?

A further issue prompted by these computer experiments concerns the nature of philosophy of science itself. Apparently, practicing scientists do not mind calling such computational procedures "experiments." This raises the question of how philosophers' notions of 'experiment' should relate to scientists' usages? Of course, this is just one example of a quite general hermeneutical issue: To what extent should philosophers take into account the concepts and interpretations of the people who are being studied (in this case, scientists)? Answers to this question will depend on the conception of philosophy one adheres to. Those philosophers who advocate a more descriptive approach will tend to follow the scientists' usages, while those who favor a more theoretical or normative approach will emphasize the legitimacy of employing their own terminology.

The Relationship Between (Experimental) Science and Technology

Traditionally, philosophers of science have defined the aim of science as, roughly, the generation of reliable knowledge of the world. Moreover, as a consequence of explicit or implicit empiricist influences, there has been a strong tendency to take the production of experimental knowledge for granted and to focus on theoretical knowledge. However, if one takes a more empirical look at the sciences, at both their historical development and their current condition, this approach must be qualified as one-sided. After all, from Archimedes' lever-and-pulley systems to the cloned sheep Dolly, the development of (experimental) science has been intricately interwoven with the development of technology (see Tiles and Oberdiek 1995). Experiments make essential use of (often specifically designed) technological devices, and conversely, experimental research often contributes to technological innovations. Moreover, there are substantial conceptual similarities between the realization of experimental and of technological processes, most significantly the implied possibility and necessity of the manipulation and control of nature. Taken together, these facts justify the claim that the science/technology relationship ought to be a central topic for the philosophy of (experimental) science.

One obvious way to study the role of technology in science is to focus on the instruments and equipment employed in experimental practice. Many studies have shown that the investigation of scientific instruments is a rich source of insights for a philosophy of scientific experimentation (Gooding, Pinch, and Schaffer 1989; Heidelberger and Steinle 1998; Radder 2003). One may, for example, focus on the role of visual images in experimental design and explore the wider problem of the relationship between thought and vision (see Visual Representation). Or one may investigate the problem of how the cognitive function of an intended experiment can be materially realized, and what this implies for the relationship between technological functions and material structures. Or one may study the modes of representation of instrumentally mediated experimental outcomes and discuss the question of the epistemic or social appraisal of qualitative versus quantitative results.

In addition to such studies, several authors have proposed classifications of scientific instruments or apparatus. One suggested distinction is between instruments that represent a property by measuring its value (e.g., a device that registers blood pressure), instruments that create phenomena that do not exist in nature (e.g., a laser), and instruments that closely imitate natural processes in the laboratory (e.g., an Atwood machine).

Such classifications form an excellent starting point for investigating further philosophical questions on the nature and function of scientific instrumentation. They demonstrate, for example, the inadequacy of the empiricist view of instruments as mere enhancers of human sensory capacities. Yet, an exclusive focus on the instruments as such may tend to ignore two things. First, an experimental setup often includes various "devices," such as a concrete wall to shield off dangerous radiation, a support to hold a thermometer, a spoon to stir a liquid, curtains to darken a room, and so on. Such devices are usually not called instruments, but they are equally crucial to a successful performance and interpretation of the experiment and hence should be taken into account. Second, a strong emphasis on instruments may lead to a neglect of the environment of the experimental system, especially of the requirement to control the interactions between the experimental system and its environment. Thus, a comprehensive view of scientific experimentation needs to go beyond an analysis of the instrument as such by taking full account of the specific setting in which the instrument needs to function.

Finally, there is the issue of the general philosophical significance of the experiment/technology relationship. Some of the philosophers who emphasize the importance of technology for science endorse a "science-as-technology" account. That is to say, they advocate an overall interpretation in which the nature of science—not just experimental but also theoretical science—is seen as basically or primarily technological (see e.g., Dingler 1928; Habermas [1968] 1978; Lelas 2000). Other authors, however, take a less radical view by criticizing the implied reduction of science to technology and by arguing for the *sui generis* character of theoretical-conceptual and formal-mathematical work. Thus, while stressing the significance of the technological (or perhaps more precisely, the action and production dimension of science), these views nevertheless see this dimension as complementary to a theoretical dimension.

The Role of Theory in Experimentation

This brings us to a further central theme in the philosophy of scientific experimentation: the relationship between experiment and theory (see Theories). The theme can be approached in two ways. One approach addresses the question of how theories or theoretical knowledge may arise from experimental practices. Thus, Franklin (1986) has developed an epistemology of experiment by arguing that following established strategies for producing stable and reproducible experiments provides a good reason for believing in the validity of the experimental results. Hon (2003) has put forward a classification of experimental error and argued that the notion of error may be exploited to elucidate the transition from the material, experimental processes to propositional, theoretical knowledge.

A second approach to the experiment/theory relationship examines the question of the role of existing theories, or theoretical knowledge, within experimental practices. Over the last two decades, this question has been debated in detail. Are experiments, factually or logically, dependent on prior theories, and if so, in which respects and to what extent? The remainder of this section reviews some of the debates on this question.

The strongest version of the claim that experimentation is theory dependent says that all experiments are planned, designed, performed, and used from the perspective of one or more theories about the objects under investigation. In this spirit, von Liebig and Popper, among others, claimed that all experiments are explicit tests of existing theories. This view completely subordinates experimental research to theoretical inquiry. However, on the basis of many studies of experimentation published during the last two decades, it can be safely concluded that this claim is most certainly false. For one thing, quite frequently the aim of experiments is just to realize a stable phenomenon or a working device. Yet, the fact that experimentation involves much more than theory testing does not, of course, mean that testing a theory may not be an important goal in particular scientific settings.

At the other extreme, there is the claim that experimentation is basically theory free. The older German school of "methodical constructivism" (see Janich 1996) came close to this position. A somewhat more moderate view is that in important cases, theory-free experiments are possible and do occur in scientific practice. This view admits that performing such "exploratory" experiments does require some ideas about nature and apparatus, but not a well-developed theory about the phenomena under scrutiny. Hacking (1983) and Steinle (1998) make this claim primarily on the basis of case studies from the history of experimental science. Heidelberger (2003) aims at a more systematic underpinning of this view. He distinguishes between theory-laden and causally based instruments and claims that experiments employing the latter type of instruments are basically theory free.

Another view admits that not all concrete activities that can be observed in scientific practice are guided by theories. Yet, according to this view, if certain activities are to count as constituting a genuine experiment, they require a theoretical interpretation (Morrison 1990; Radder 1996; Hon 2003). More specifically, performing and understanding an experiment depends on a theoretical interpretation of what happens in materially realizing the experimental process. In general, quite different kinds of theory may be involved, such as general background theories, theories of the (material, mathematical, or computational) instruments, and theories of the phenomena under investigation.

One argument for such claims derives from the fact that an experiment aims to realize a reproducible correlation between an observable feature of the apparatus and a feature of the object under investigation. The point is that materially realizing this correlation and knowing what can be learned about the object from inspecting the apparatus depends on theoretical insights about the experimental system and its environment. Thus, these insights pertain to those aspects of the experiment that are relevant to obtaining a reproducible correlation. It is not necessary, and in practice it will usually not be the case, that the theoretical interpretation offers a full understanding of any detail of the experimental process.

A further argument for the significance of theory in experimentation notes that a single experimental run is not enough to establish a stable result. A set of different runs, however, will almost always produce values that are, more or less, variable. The questions then are: What does this fact tell us about the nature of the property that has been measured? Does the property vary within the fixed interval? Is it a probabilistic property? Is its real value constant, and are the variations due to random fluctuations? In experimental practice, answers to such questions are based on an antecedent theoretical interpretation of the nature of the property that has been measured.

Regarding these claims, it is important to note that in actual practice, the theoretical interpretation of an experiment will not always be explicit and the experimenters will not always be aware of its use and significance. Once the performance of a particular experiment or experimental procedure becomes routine, the theoretical assumptions drop out of sight: They become like an (invisible) window to the world. Yet, in a context of learning to perform and understand the experiment or in a situation where its result is very consequential or controversial, the implicit interpretation will be made explicit and subjected to empirical and theoretical scrutiny. This means that the primary locus of the theoretical interpretation is the relevant epistemic community and not the individual experimenter.

Further Issues for the Philosophy of Scientific Experimentation

As was explained before, the systematic philosophical study of scientific experimentation is a relatively recent phenomenon. Hence, there are a number of further issues that have received some attention but merit a much more detailed account. In concluding this article, three such issues will be briefly discussed.

The first bears upon the notion of (scientific) experience. If one takes into account the fact that in many cases scientific experience is experimentally realized experience, the empiricist view that reduces scientific experience to sense perception or even to visual sensation needs to be revised. Of course, this view has already been challenged by the claim that all observation is theory laden (see Empiricism; Observation; Perception). Yet, a systematic study of the action and production aspects of experimentation will lead to a more radical criticism. The studies that have been done so far suggest that gaining scientific experience depends not just on intersubjectively communicable language but also on human agency and that it requires particular skills that cannot be supposed to be simply universally available.

A second subject that merits more attention from philosophers of science is the nature and role of experimentation in the social and human sciences, such as economics, sociology, medicine, and psychology. Practitioners of those sciences often label substantial, or even large, parts of their activities "experimental." So far, this fact is not reflected in the philosophical literature on experimentation, which has focused primarily on the natural sciences. Thus, a challenge for future research is to connect the primarily methodological literature on experimentation in economics, sociology, medicine, and psychology with the philosophy of science literature on experimentation in natural science.

One subject that will naturally arise in philosophical reflection upon the similarities and dissimilarities of natural and social or human sciences is the problem of the double hermeneutic. Although it is true that the nature of this problem has been transformed by the more recent philosophical accounts of the practices of the natural sciences, the problem has by no means been resolved. The point is this: In experiments on human beings, the experimental subjects, in addition to the scientists, will often have their own interpretation of what is going on in these trials, and this interpretation may influence their responses over and above the behavior intended by the experimenters. As a methodological problem (of how to avoid "biased" responses), this is of course well known to practitioners of the human and social sciences. However, from a broader philosophical or sociocultural perspective, the problem is not necessarily one of bias. It may also reflect a clash between a scientific and a commonsense interpretation of human beings. In the case of such a clash, social and ethical issues are at stake, since the basic question is, Who is entitled to define the nature of human beings: the scientists or the people themselves? In this form, the methodological, ethical, and social problems of the double hermeneutic will continue to be a significant theme for the study of experimentation in the human and social sciences.

This brings us to a last issue. The older German tradition explicitly addressed wider normative questions surrounding experimental science and technology. The views of Habermas, for example, have had a big impact on broader conceptualizations of the position of science and technology in society. Thus far, the more recent Anglophone approaches within the philosophy of scientific experimentation have dealt primarily with more narrowly circumscribed scholarly topics. Insofar as normative questions have been taken into account, they have been mostly limited to epistemic normativity—for instance, to questions of the proper functioning of instruments or the justification of experimental evidence. Questions regarding the connections between epistemic and social or ethical normativity are hardly addressed.

Yet, posing such questions is not far-fetched, and they often relate to ontological, epistemological, or methodological concerns quite directly. For instance, those experiments that use animals or humans as experimental subjects are confronted with a variety of normative issues, often in the form of a tension between methodological and ethical requirements. Also normatively relevant are (1) the ontological issue of the artificial and the natural in experimental science and (2) science-based technology. Consider, for example, the question of whether experimentally isolated genes are natural or artificial entities. This question is often discussed in environmental philosophy, and different answers to it entail different environmental ethics and politics. More specifically, the issue of the contrast between the artificial and the natural is crucial to debates about patenting, in particular the patenting of genes and other parts of organisms. The reason is that discoveries of natural phenomena are not patentable, while inventions of artificial phenomena are (see Sterckx 2000).

Although philosophers of experiment cannot be expected to solve all of those broader social and normative problems, they may be legitimately asked to contribute to the debate on possible approaches and solutions. In this respect, the philosophy of scientific experimentation could profit from its kinship to the philosophy of technology, which has always shown a keen sensitivity to the interconnectedness between technological and social or normative issues.

Hans Radder

References

Baird, Davis (2004), *Thing Knowledge: A Philosophy of Scientific Instruments*. Berkeley and Los Angeles: University of California Press.

Bernard, Claude ([1865] 1957), *An Introduction to the Study of Experimental Medicine*. New York: Dover Publications.

Bhaskar, Roy (1978), *A Realist Theory of Science*. Hassocks, UK: Harvester Press.

Brown, James R. (1991), *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge.

Buchwald, Jed Z. (ed.) (1995), *Scientific Practice: Theories and Stories of Doing Physics*. Chicago: University of Chicago Press.

Campbell, Donald T., and Julian C. Stanley (1963), *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally College Publishing.

Collins, H. M. (1985), *Changing Order: Replication and Induction in Scientific Practice*. London: Sage.

Dingler, Hugo (1928), *Das Experiment. Sein Wesen und seine Geschichte*. Munich: Verlag Ernest Reinhardt.

Feenberg, Andrew (1999), *Questioning Technology*. London: Routledge.

Franklin, Allan (1986), *The Neglect of Experiment*. Cambridge: Cambridge University Press.

Gooding, David, Trevor Pinch, and Simon Schaffer (eds.) (1989), *The Uses of Experiment*. Cambridge: Cambridge University Press.

Habermas, Jürgen ([1968] 1978), *Knowledge and Human Interests*, 2nd ed. London: Heinemann.

Hacking, Ian (1983), *Representing and Intervening*. Cambridge: Cambridge University Press.

Heidelberger, Michael (2003), "Theory-Ladenness and Scientific Instruments in Experimentation," in Hans Radder (ed.), *The Philosophy of Scientific Experimentation*. Pittsburgh: University of Pittsburgh Press, 138–151.

Heidelberger, Michael, and Friedrich Steinle (eds.) (1998), *Experimental Essays—Versuche zum Experiment*. Baden-Baden: Nomos Verlagsgesellschaft.

Hon, Giora (2003), "The Idols of Experiment: Transcending the 'Etc. List,'" in Hans Radder (ed.), *The Philosophy of Scientific Experimentation*. Pittsburgh: University of Pittsburgh Press, 174–197.

Janich, Peter (1996), *Konstruktivismus und Naturerkenntnis*. Frankfurt am Main: Suhrkamp.

Kuhn, Thomas S. (1977), "Mathematical versus Experimental Traditions in the Development of Physical Science," in *The Essential Tension*. Chicago: University of Chicago Press, 31–65.

Latour, Bruno, and Steve Woolgar (1979), *Laboratory Life: The Social Construction of Scientific Facts*. London: Sage.

Lelas, Srðan (2000), *Science and Modernity: Toward an Integral Theory of Science*. Dordrecht, Netherlands: Kluwer.

Mayo, Deborah G. (1996), *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Morrison, Margaret (1990), "Theory, Intervention and Realism," *Synthese* 82: 1–22.

Radder, Hans (1996), *In and About the World*. Albany: State University of New York Press.

——— (ed.) (2003), *The Philosophy of Scientific Experimentation*. Pittsburgh: University of Pittsburgh Press.

Shapin, Steven, and Simon Schaffer (1985), *Leviathan and the Air-Pump: Hobbes, Boyle and Experimental Life*. Princeton, NJ: Princeton University Press.

Steinle, Friedrich (1998), "Exploratives vs. theoriebestimmtes Experimentieren: Ampères erste Arbeiten zum Elektromagnetismus," in Michael Heidelberger and Friedrich Steinle (eds.), *Experimental Essays—Versuche zum Experiment*. Baden-Baden: Nomos Verlagsgesellschaft, 272–297.

Sterckx, Sigrid (ed.) (2000), *Biotechnology, Patents and Morality*, 2nd ed. Aldershot, UK: Ashgate.

Tiles, Mary, and Hans Oberdiek (1995), *Living in a Technological Culture*. London: Routledge.

See also **Bridgman, Percy; Empiricism; Epistemology; Kuhn, Thomas; Logical Empiricism; Observation; Perception; Popper, Karl Raimund; Psychology, Philosophy of; Scientific Change; Scientific Realism; Social Constructionism; Social Sciences, Philosophy of the; Statistics, Philosophy of; Theories; Visual Representation**

EXPLANATION

One of the most important aims of science is to provide explanations of natural phenomena. Consequently, philosophers have devoted much attention to the nature of scientific explanation. The twentieth century's most influential model of scientific explanation is known as the covering-law model, which has been articulated most fully in the work of Carl Hempel (1965). According to this model, an explanation is an argument whose conclusion is a statement of some fact to be explained (the explanandum) and whose premises (the explanans) comprise a set of statements that include at least one natural law and collectively provide either inductive or deductive support for the explanandum. The covering-law model suggests that the explanandum is explained by rendering it *nomically expectable*. The covering-law model is part of the legacy of logical empiricism. While this model is rejected by most contemporary philosophers of science, the majority of the literature in the last 50 years has been concerned with either defending or correcting problems with the covering-law model.

Both exponents and critics of the covering-law model have emphasized explanation in the physical sciences. Philosophers concerned with explanation in the biological and social sciences have sometimes argued that within these domains different kinds of explanations—notably *functional explanations* and *reductive explanations*—play some special role. Another major issue in the contemporary literature concerns the nature of these kinds of explanation and their relation to the covering-law model.

The Covering-Law Model of Explanation

A covering-law explanation is any explanation in which the explanandum is the conclusion of an argument whose premises contain at least one natural law. Hempel recognizes three varieties of covering-law explanation:

1. *Deductive-nomological* (D-N): These are deductive arguments whose premises include universal (deterministic) laws along with statements of particular conditions.
2. *Inductive-statistical* (I-S): These are inductive arguments whose premises include statistical laws.
3. *Deductive-statistical* (D-S): These are deductive arguments in which statistical laws are entailed by more comprehensive statistical laws.

Explanations may be further subdivided according to the logical character of the explanandum statement. The explanandum statement may be either a singular statement, a universal statement, or a statistical generalization.

The Deductive-Nomological Model

As an example of a D-N explanation, consider why a partially submerged oar appears to bend at the point where it enters the water. The phenomenon is a consequence of refraction, and a derivation of the observed angle can be given using Snell's law together with the indices of refraction of air and water. Note that the derivation will rely both upon general law statements (Snell's law together with statements of the indices of refraction for air and water) and particular statements (in particular, the angle of incidence of the oar with the water). Schematically, one can represent the D-N explanation as an argument:

$$\begin{array}{l} \left.\begin{array}{l} L_1, L_2 \ldots L_k \\ C_1, C_2 \ldots C_n \end{array}\right\} \textit{Explanans} \\ \hline \left. E \right\} \textit{Explanandum} \end{array}$$

where L_1, L_2 ... L_k are laws, C_1, C_2 ... C_n are statements of antecedent particular conditions, and E is a statement of the explanandum. In their original statement of the D-N model, Hempel and Oppenheim (1948) stipulated that for an argument to be a D-N explanation, it must meet three logical conditions of adequacy:

1. The explanandum must be a logical consequence of the explanans.
2. The explanans must include at least one law.
3. The explanans must have empirical content, in the sense that (at least some of) its component statements must be susceptible to empirical test.

Hempel and Oppenheim take the second requirement to be a logical condition, because they hold that the distinction between laws and nonlaws is essentially syntactic (see Hempel, Carl). Following a Humean analysis, they take laws simply to be universal generalizations of unrestricted scope, with no designations of particular objects and containing only purely qualitative predicates. They do not require that the explanation contain any singular statements, because they wish to allow D-N explanations of laws, in which case no singular statements should be required. A fourth condition, not mentioned, but clearly in the spirit of their model, is that the laws are essential to the explanation in the sense that omitting them from the explanans will make the argument invalid.

To the three logical conditions, Hempel and Oppenheim add an empirical condition of adequacy, which is that the statements in the explanans be true. It is useful to distinguish between a potential explanation, which meets the logical criteria, and an actual explanation, which meets both the logical and empirical criteria. Thus, for instance, Descartes offered a potential explanation of Snell's law, but it is not an actual explanation because certain laws of the corpuscular theory of light on which his derivation was based are in fact false.

An important consequence of the D-N model, and more generally of the thesis that explanations are arguments, is that explanations are logically indistinguishable from predictions. This consequence, often called the *structural identity thesis*, suggests that every explanation is a potential prediction and every prediction a potential explanation. According to Hempel, the distinction between explanation and prediction is essentially pragmatic. An explanatory argument can serve as a prediction when an explanandum statement is not antecedently known but the explanans statements are.

While many scientific explanations seem to fit the D-N model, critics have presented a number of counterexamples showing that the D-N model is either too restrictive, in the sense that its requirements rule out genuine explanations, or too permissive, in the sense that there are pseudo-explanations that meet the logical and empirical requirements of the model.

The claim that the D-N requirements are too restrictive has chiefly been made on the grounds that some genuine explanations do not involve any laws. Scriven (1962) considers the example of how to explain an ink stain on his carpet. If asked for such an explanation, Scriven might truthfully claim that the carpet was stained when he hit his writing table with his knee, overturning an ink bottle on the table, causing the ink to drip onto the carpet. According to Scriven, the assertion of these (singular) facts constitutes a complete explanation of the explanandum event. He does not know, and it is impossible to supply, universal laws that together with particular conditions entail the explanandum statement.

Hempel (1965) responds that such an explanation is not complete but is rather an enthymeme. One can have confidence that the (partial) explanation is a good one only if one believes that there exists some as yet unknown law or laws that together with the particulars cited in the explanans would entail the explanandum. Scriven points out that people quite often, as in this case, have causal knowledge without having knowledge of laws, but Hempel argues that this practical ability is not inconsistent with the claim that the existence of a causal connection implies the existence of a law. Scriven's putative counterexamples highlight the fact that scientific explanations typically cite causes of the explanandum event. Whether the D-N model is too restrictive depends, then, on whether all causally related events are instances of lawful regularities.

The D-N model is also subject to several well-known counterexamples that apparently show that the D-N requirements are too permissive. Four important ones are as follows:

1. It is possible to predict the length of a shadow cast by a flagpole using some elementary trigonometry plus measurements of the height of the flagpole and the angle of the sun. This calculation provides a D-N explanation of the length of the flagpole's shadow. However, it is equally possible to use the length of the shadow and the angle of sun to calculate the height of the flagpole. Such a calculation satisfies the requirements of the D-N model, but clearly the length of the shadow cannot be used to explain the height of the flagpole. This example, due to Bromberger (1966), calls into question the structural identity thesis, because the length of the shadow can be used to predict the height of the flagpole, but it cannot explain it.
2. Consider the following argument:

   Whenever the barometer drops, a storm occurs.
   The barometer drops.
   A storm occurs.

   Supposing that the first premise is a law, this argument meets the requirement for a D-N explanation, and yet, while the falling barometer may serve to predict the storm, it cannot explain it.
3. Suppose Smith ingests a lethal dose of a slow-acting poison that kills everyone who takes it within 24 hours. Suppose that immediately after ingesting the poison, Smith steps into the street and is run over by a bus. Given that Smith has ingested the poison, one can predict Smith's death, but the fact that Smith died is not explained by that fact.
4. Joe Jones (a male), though sexually active, regularly takes birth control pills. If the pill is 100% effective, then the following is a D-N explanation:

   Whenever a person takes birth control pills, that person avoids pregnancy.
   Joe Jones regularly takes birth control pills.
   Joe Jones avoids pregnancy.

While the argument is sound, Jones's failure to get pregnant is not explained by his use of birth control pills, but by his gender.

Each of these counterexamples offers an argument that purportedly meets the logical stipulations on D-N explanations but is not genuinely explanatory. In each case, however, the argument is predictive. Thus the counterexamples raise doubts both for the adequacy of the D-N model and for the correctness of the structural identity thesis. In each of these cases, the account of why the explanation is spurious has to do with the failure of the putative explanation to cite the causes of the explanandum. The explanatory asymmetry in the flagpole case arises from the asymmetry of cause and effect. It is possible both to explain and to predict effects from causes, but one can predict but not explain causes from effects. In the second case, the law connecting storms to falling barometers does not display a direct causal relationship but is explained by the operation of a common cause. Both the barometer's drop and the storm's occurrence are caused by falling atmospheric pressure. One effect of a common cause may be used to predict the other, but it cannot explain it. The third counterexample is a case of causal preemption. The ingestion of the poison initiates a causal process that will lead to Smith's death, but the process is preempted by the bus, which has the same effect. Again, the moral is that the D-N argument allows for prediction but not explanation. The fourth is a kind of overdetermination. The law cited is predictive, not explanatorily relevant.

The last major difficulty with the D-N model is that it is not possible to assess the correctness of a D-N explanation without an adequate understanding of what constitutes a natural law (see Laws of Nature). Hempel understood laws to be universal generalizations of unrestricted scope. For instance, it is a law that no object travels faster than the speed of light. Such laws can be represented in first-order predicate logic by a universal generalization. The problem with such a characterization is that not all universal generalizations express laws. Many universal generalizations are only accidentally true. For instance, it may be the case that all of the students in my classroom are under 20 years of age, but this universal generalization would be only accidentally true. It does not support counterfactuals, in the sense that it does not imply the claim that if a student were in my class, she would be under 20 years of age. The idea that a law is a universal generalization that supports counterfactuals is one that has considerable plausibility, but it is difficult to explicate the semantics of counterfactuals within a framework acceptable to empiricists.

Various philosophers, including Hempel (1965) and Nagel (1961), have tried instead to stipulate extra syntactic and semantic conditions that would distinguish lawful from accidental generalizations. In particular they have suggested that laws are exceptionless universal generalizations that include no reference to particulars and involve purely qualitative predicates. One consequence of this definition is that many general claims that scientists call laws would not be considered such. For instance, Kepler's and Mendel's laws would not be genuine laws, since they refer implicitly or explicitly to particulars of Earth's solar system and life on Earth. In fact, it may well be the case that no science except physics has any laws in the required sense. Hempel may not have been concerned with this, since he assumed that in principle, laws of the special sciences could be derived from more general physical laws in combination with statements about particulars. This assumption has been widely challenged by critics of reduction. Perhaps more telling is Goodman's critique of the concept of a purely qualitative predicate. Goodman's (1956) new riddle of induction seems to imply that there is no empirically respectable way to characterize purely qualitative predicates.

Another way to approach this problem, advocated by Woodward (2000), is to replace the appeal to laws in D-N explanations with appeals to invariant generalizations. In Woodward's view, explanations can be made by subsuming explananda under generalizations of varying degrees of invariance. True laws, which are strictly invariant, are a rare but limiting case. While Woodward's approach is broadly in the spirit of Hempel's covering-law models, Woodward's analysis of invariance requires appeals to counterfactuals of a kind inconsistent with stricter versions of empiricism.

Deductive-Nomological and Deductive-Statistical Explanations of General Laws

In the examples considered so far, the explananda have been singular events or states of affairs. Hempel and Oppenheim (1948) also envisioned using the D-N model to analyze explanations of laws and regularities. A D-N explanation of a law is simply a derivation of that law according to the D-N model. If the explanandum is a statistical law, Hempel called the explanatory argument a D-S explanation, but D-S explanations are essentially a species of D-N explanation.

The idea that less fundamental laws can be explained by more fundamental ones is appealing. The case of Newton's explanation of Kepler's laws in terms of his laws of motion and gravitation seems to fit this mold. There are, however, problems with Hempel and Oppenheim's explication. Perhaps the most important difficulty is illustrated by the following example: Consider an argument whose single premise is the conjunction of the law of universal gravitation (UG) with Snell's law and whose conclusion is UG alone. This apparently meets the formal requirements for a D-N explanation of UG, and the conjunctive law is "more general" than UG in the sense that both UG and Snell's law are derivable from it. Nonetheless, this is clearly not a genuine explanation of UG. The example shows that it is unclear how to use derivability relations to distinguish more and less fundamental laws.

The Inductive-Statistical Model

In "Aspects of Scientific Explanation," Hempel (1965) extends the covering-law model to statistical explanations with his I-S model. As an example of an I-S explanation, Hempel considers the treatment of streptococcus with penicillin. Suppose that a patient Jones is suffering from streptococcus, takes penicillin, and subsequently recovers from the infection. The following inductive argument serves as an I-S explanation of Jones's recovery:

A patient with streptococcus who takes penicillin has a high probability of recovery.
Jones has streptococcus and takes penicillin (with high probability)
Jones recovers.

The formal requirements for an I-S explanation are identical to those of a D-N explanation, except that (1) the explanans must contain a statistical law and (2) the relationship between the premises and the conclusion of the explanatory argument is one of inductive strength rather than deductive validity. The simplest I-S explanations will have the following form:

$$\begin{array}{l} P(G|H) = r \\ \underline{Hj \qquad\qquad} [r] \\ Gj \end{array}$$

Hempel understood the probability used in the statistical law as a relative frequency, while the bracketed probability was understood as an inductive (logical) probability. The use of these probabilities raises questions about I-S explanation. First, the concept of inductive probability is very difficult to explicate, and it has so far proved impossible to establish a definitive measure of inductive probability (see Probability). Second, it is unlikely that the statistical probability concept Hempel uses in his characterization of statistical laws is adequate for distinguishing statistical laws from accidental statistical associations. The problems are in many respects analogous to those of distinguishing lawful from accidental universal generalizations, but it is in other respects worse (cf. Dupré and Cartwright 1988).

A more immediate difficulty is what Hempel calls the problem of the ambiguity of I-S explanation. The problem is that it may be possible to formulate two inductively strong arguments with true premises that support opposite conclusions. Suppose, for instance, that the strain of streptococcus with which Jones is infected is known to be penicillin resistant. One then has the I-S explanation:

A patient with penicillin-resistant streptococcus who takes penicillin has a low probability of recovery.
Jones has penicillin-resistant streptococcus and takes penicillin (with high probability)
Jones does not recover.

The premises of both this and the previous argument are true, and yet the arguments support opposite conclusions. This situation does not arise with D-N explanations because of an important difference between inductive and deductive arguments. While deductively valid arguments can never be made invalid by addition of further premises, inductively strong arguments can be weakened by additional premises. Here, the original argument explaining Jones's recovery is weakened by the additional information that the strain with which Jones was infected is penicillin resistant.

This general problem with inductive inference motivated Rudolf Carnap (1950, 211) to stipulate that correct measures of inductive support of a hypothesis can be made only in light of total evidence. Applying this requirement naively to the I-S model would suggest that the appropriate explanans for any I-S explanation is the entire knowledge base K. Hempel (1965, 399–400) suggested a refined version of Carnap's principle, which he called the *principle of maximal specificity*.

In the simple version of Hempel's model, the explanandum statement is an assertion that a particular individual j is a member of a class G. The statistical law in the explanans is an assertion that the relative frequency of individuals in H who are in G is r. The ambiguity of I-S explanation arises because different choices of H lead to different values of r. Hempel's solution is to demand that the choice of H be maximally specific. If j is a member of a class H and of a class H' that is a proper subclass of H, the explanans should contain the probabilistic law $P(G|H') = r'$, rather than $P(G|H) = r$. For instance, since the class of persons with penicillin-resistant streptococcus infections is a subclass of the class of persons with streptococcus infections, the explanatory argument should be based on the former class. So if Jones has a penicillin-resistant infection, his recovery is not explained by his having taken penicillin.

A major problem with Hempel's proposal concerns what classes may legitimately be taken as reference classes. If one is allowed to take any class as a reference class, one may simply take the reference class H' to be the intersection of H and G. This generates a trivial and nonexplanatory I-S explanation. Hempel does not have a satisfactory formal way out of this difficulty, but clearly his intention is that reference classes are the extensions of observable or theoretical predicates used in characterizing total scientific knowledge K. A given body of knowledge will entail a particular set of reference classes with respect to which statistical probabilities should be measured. For instance, in a given knowledge situation, when one wishes to explain the cause of lung cancer, it is recognized that reference classes should be partitioned by such properties as age, gender, weight, smoking habits, presence of environmental pollutants, etc. The consequence of this proposal, as Hempel recognized, is that the concept of statistical explanation is essentially relativized to a knowledge situation. This means that, in Hempel's account, there is an essential pragmatic element of statistical explanation that is not present in D-N explanation.

While it is possible to construct statistical analogs of the problems facing D-N explanation, the I-S model has certain difficulties that are peculiar to it. The most widely discussed of these concerns Hempel's high-probability requirement. Hempel believed that all I-S explanations must show that the explanandum is to be expected. Recognizing that what counts as high probability is a pragmatic issue, Hempel's requirement rules out the probabilistic explanation of unlikely events. This is in keeping with his defense of the structural identity thesis.

The difficulty with this requirement is illustrated by Scriven's example of syphilis and paresis. Paresis is a form of tertiary syphilis that can be contracted only by persons who have originally contracted syphilis. However, paresis is relatively rare even among syphilitics, so the probability of contracting paresis given that one has syphilis is low. Nonetheless, Scriven argues, it is reasonable to cite a person's syphilis as an explanation for his paresis. About this case, Hempel insists that paresis is not explained by syphilis, arguing that necessary conditions are not generally explanatory and that, given the fact that most syphilitics do not contract paresis, other factors must be cited in the explanation.

The high-probability requirement is also open to the objection that it is sometimes too lax. Suppose, for instance, one seeks to explain teenagers' interest in sex by reference to the TV programs they watch. It is clearly spurious to argue that teenagers' interest in sex is explained by their TV-viewing habits, just because most teenagers who watch TV are interested in sex. Most teenagers appear to be interested in sex quite independently of TV.

What both of these examples suggest is that the central issue in statistical explanation is not whether the probability of the explanandum given the explanans is high, but whether the factors cited in the explanans make a difference to the probability of the explanandum. This is the intuition behind the statistical relevance approach discussed below.

Alternatives to Covering-Law Models

The problems for Hempel's D-N and I-S models described above have led to a number of alternative theories of explanation. Four alternative approaches will be considered: the statistical relevance model (Salmon, Greeno, Jeffrey), the causal/mechanical approach (Railton, Salmon), the pragmatic approach (Bromberger, van Fraassen), and explanatory unification (Friedman, Kitcher).

The Statistical Relevance Model

The fundamental intuition behind Hempel's I-S model is that a statistical explanation explains its explanandum by providing an argument that renders the explanandum probable. Scriven's example of the syphilitic man calls this intuition into question. The man's paresis is explained by the fact that he has syphilis, even though paresis is unusual among syphilitics. The statistical relevance (SR) approach, developed chiefly by Wesley Salmon (Salmon, Greeno, and Jeffrey 1971; Salmon 1984), provides a model of explanation that shows how such factors can be explanatory. In general, a factor C is statistically relevant to a factor B just in case $P(B|C) \neq P(B)$. The intuition behind the SR model is that the presence of B is explained in a particular case by finding factors C that are positively relevant to B. The factors may be explanatory even if the probability of B given C is small.

Suppose that one is interested in explaining why an adolescent female Jenny became pregnant. The explanatory query can be phrased in this way: Why is it that Jenny, who is an adolescent female, is also pregnant? The reference class A, the class of adolescent females, serves as a baseline from which one calculates probabilities (construed as relative frequencies). One then partitions this class by a mutually exclusive and exhaustive set of factors B_i.

In this case, there are only two—B_1 refers to the class of those who have become pregnant and B_2 refers to the class of those who have not. This partition is called the *explanandum partition*. One then partitions the reference class by a set of explanatory factors into classes C_i that are mutually exclusive and exhaustive. In this case, factors might include the parents' education level, family income, ethnic group, religious affiliation, etc. Values for C_i represent various conjunctions of these factors. An SR explanation would show that a set of factors are explanatory by finding the partition C_a to which Jenny belongs and showing that $P(B_1|A \& C_a) > P(B_1|A)$. For instance, one might find that Jenny is a Polish Catholic woman of high school educated parents with a family income of less than $30,000 and that women within this group are more likely to become pregnant than teenage women as a whole. The SR model is a formalization of an approach to statistical explanation familiar from the social sciences. Social scientists interested in explaining the occurrence of a property within a population will partition that population according to some set of potentially relevant factors and collect data to discover in which partitions the property is most frequent.

This procedure is open to a number of objections. One of the most familiar is connected with a statistical phenomenon called *Simpson's paradox* (Cartwright 1979). The problem raised by Simpson's paradox is that a partition C that is positively relevant to B may sometimes be partitioned into subpartitions D, in which each D is negatively relevant to B:

$$P\,(B \mid A \wedge C) > P\,(B \mid A),$$

but for all i,

$$P\,(B \mid A \wedge D_i) < P\,(B \mid A).$$

Cartwright discusses a famous example of this case concerning admissions to graduate school at the University of California, Berkeley. Questions had been raised about whether Berkeley was discriminating against women in admissions, because it turned out that women had a lower admission rate than men. A closer look dispelled this concern. It turned out that on a department-by-department basis, women were admitted at rates equal to or higher than men's. The lower admission rate for women was caused by the fact that women applied disproportionately to departments with lower overall admission rates.

The trick to avoiding spurious inferences is to create reference-class partitions that are so fine-grained that their members are homogeneous with respect to all causally relevant properties. If one can be sure that members of a reference-class partition are *objectively homogeneous* with respect to causally relevant factors, then one can justifiably say that the set of factors defining the partition is positively or negatively relevant to the explanandum. Unfortunately, the concept of an objectively homogeneous reference class is fraught with conceptual and epistemic difficulties (cf. Salmon 1984, Ch. 3).

A second problem with the SR approach is that it admits explanatory factors that are correlated with the explanandum but that are not *causally* relevant. This problem can arise when the explanatory factor is correlated with the explanandum due to a common cause. In the earlier counterexample given, the correlation of the barometer and the storm to the D-N model is a case in point, as the barometer's falling is statistically relevant to the occurrence of the storm. To meet this objection, one can amend the SR account by stipulating that a statistically relevant factor is not explanatory if it can be *screened off* by another factor. A factor D screens off a factor C from B if

$$P\,(B \mid C \,\&\, D \,\&\, A) = P(B \mid D \,\&\, A)$$

but

$$P(B \mid C \,\&\, D \,\&\, A) = P(B \mid C \,\&\, A)$$

.

Applying this to the barometer case, actual change in atmospheric pressure screens off barometer readings, because once the atmospheric pressure is fixed, variations in barometer readings (say, due to barometer malfunctions) become irrelevant. While this is intuitively plausible, there are again conceptual and empirical problems with applying screening-off criteria in such a way as to completely eliminate spurious causes.

Salmon himself ultimately became convinced that it was not possible to solve all of the problems associated with the SR approach. Statistical relevance relations provide an evidential basis for making judgments of causal relevance, but causal (and hence explanatory) relevance must be understood independently of statistical relations. Salmon's mechanistic account of causation, described below, was meant to provide this missing ingredient.

The Causal Mechanical Approach

Peter Railton (1978) first introduced the concept of mechanism into the contemporary literature on explanation. His deductive-nomothetic model of probabilistic explanation (the D-NP model) was meant as an alternative to Hempel's I-S model. Railton was concerned with Hempel's requirement that the explanans of an I-S explanation rendered the explanandum probable or nomically expectable. Railton argued that explanations describe causes, and sometimes the causal sequence of events leading up to the event to be explained may be improbable. According to Railton, while an explanation of some event may include a reference to a law that renders the event nomically expectable, the account must be supplemented by "an account of the mechanism(s) at work" (1978, 208). Railton is vague on just what a mechanism is, indicating only that an "account of the mechanism(s)" is "a more or less complete filling-in of the links in the causal chains" (ibid).

Salmon's work on causal-mechanical explanation, beginning with his seminal *Scientific Explanation and the Causal Structure of the World* (1984), elaborates Railton's earlier account of mechanistic explanation. Though he dubs his theory "mechanistic," his actual analysis is not of the concept of mechanism. Rather, he argues that explanations must refer to what he calls the "causal nexus," which he takes to be a vast network of interacting causal processes. Salmon defines a *process* to be an entity that maintains a persistent structure through space-time, a *causal process* to be a process capable of transmitting changes in its structure, and a *causal interaction* to be an intersection between causal processes in which an alteration of the persistent properties of those processes occurs. In Salmon's original formulation, interactions were defined in terms of a counterfactual criterion of mark transmission. In response to criticisms of this criterion, he has eliminated the reference to counterfactuals (Salmon 1994), relying instead on a definition in which causal interactions involve exchanges of conserved quantities.

Both versions of Salmon's theory seem vulnerable to a criticism raised by Hitchcock (1995), whose concern is that there can be events causally connected to but irrelevant to the explanation of an explanandum event. Salmon's example of Jones and the birth control pills illustrates the point. Jones's ingestion of birth control pills counts as a causal interaction under either the counterfactual or the conserved quantity account. It is part of the causal nexus preceding Jones's failure to get pregnant. How is this interaction to be excluded as explanatorily irrelevant? The obvious answer is that the counterfactual claim that Jones would have gotten pregnant if he had not taken birth control pills is false; but nothing in Salmon's theory appears to require that this claim be true.

Glennan (2002) has argued that this and other difficulties with the mechanistic approach to explanation arise from an incorrect analysis of the concept of mechanism. Salmon and Railton both conceive of mechanisms as constituting a nexus of intersecting causal processes. An alternative view of mechanisms, advocated by Glennan and by Machamer, Darden, and Carver (2000), among others, suggests that mechanisms are complex systems—ensembles of interacting parts. While a causal process in Salmon's sense may involve the operation of a mechanism, the mechanism is not identified with the single instance of this process but is rather the system that reliably underlies processes of a certain type. The spurious explanation of Jones's failure to get pregnant is rejected both because the only reliable mechanism for preventing pregnancy by birth control pills involves the female reproductive system and because there is no reliable mechanism for the production of male pregnancy, so there is no need to explain the failure of the mechanism.

Pragmatic Accounts of Explanation

From a linguistic point of view, an explanation can be viewed as an answer to a *why*-question. Aristotle, in his theory of the four causes, already recognizes that the same *why*-question can be correctly answered in a number of ways, depending upon the beliefs and interests of the questioner. Pragmatic theories of explanation seek to explicate the relationship between the context in which a *why*-question is asked and the kinds of answers that can be appropriately given.

Hempel was certainly aware that explanation had a pragmatic dimension, but at least in the case of D-N explanation, the essential part is semantic. A D-N explanation is a valid argument, and the validity of an argument is independent of pragmatic factors. Pragmatic factors will explain which questions are asked as well as how the argument is presented (e.g., which premises are treated as implied), and not much more.

The pragmatic response to Hempel's account begins with Scriven but has been more fully developed in the work of Bromberger (1966) and, especially, van Fraassen (1980, Ch. 5). Van Fraassen claims that his pragmatic theory of explanation has the resources to resolve the problems of asymmetry and irrelevance that plague the D-N model.

An explanation is an answer to a *why*-question, "Why *P*?," where *P* is some true proposition. *P* is, in Hempel's terminology, the explanandum. Van

Fraassen argues, however, that there is more to the question than P itself. First, when one asks "Why P?," one is implicitly contrasting the state of affairs expressed by P to an alternative set of states of affairs, called the *contrast class*. To take one of van Fraassen's examples, the question "Why did Adam eat the apple?" can be understood variously as asking (1) why *he* ate the apple (as opposed to the serpent or other creatures in the garden), (2) why he *ate* the apple (as opposed to refusing it, or perhaps throwing it at Eve), or (3) why he ate *the apple* (as opposed to some other fruit in the garden). Besides the contrast class, van Fraassen suggests that the context of the question includes a *relevance relation*, which specifies the kinds of answers that are considered relevant to the question. For instance, if one were to ask why primates have opposable thumbs (in contrast to the pattern of fingers in other mammals), one could either be interested in an evolutionary explanation, in which case the relevance relation would relate selectively relevant features of environments of ancestral species to various morphological traits. Or one could be interested in a developmental explanation, in which case the relevance relation would relate combinations of genetic and environmental factors to traits that these combinations cause to develop.

Van Fraassen's (1980) formal account incorporates the following definitions:

1. A question Q is a triple $<P_K, X, R>$, where P_K is the topic (or explanandum), X is the contrast class, which is a set of propositions that includes the topic, and R is a relevance relation of propositions to ordered pairs of topic propositions and contrast classes.
2. The *presupposition* of Q is the conjunction of the claims
   a. that the topic P_K is true,
   b. that every other proposition in the contrast class X is false, and
   c. that there is at least one proposition that bears the relation R to $<P_K, X>$ that is true.
3. A *direct answer* to Q is the conjunction of the presupposition and a proposition A that bears relation R to $<P_K, X>$.
4. A is called the *core* of the answer to Q. (143)

Most commentators agree that van Fraassen's account goes a long way toward elucidating explanatory practices. It shows, for instance, why certain *why*-questions can be rejected (e.g., if the topic is false or if the other elements of the contrast class are not all false) and why a verbally identical *why*-question can admit of different answers (e.g., because of different implied but unstated relevance relations). Van Fraassen, however, claims that his theory is sufficient to solve the explanatory asymmetry problems that plague the D-N model. He supports his claim by means of an amusing parable regarding a tower and its shadow (van Fraassen 1980, 132–134). As with Bromberger's flagpole, one would expect that the length of the shadow can be explained in terms of the height of the tower and the altitude of the sun, but not vice versa. In van Fraassen's parable, however, a wealthy chevalier has constructed a tower of a certain height in order that it should cast a shadow over the spot where he proclaimed his love to a woman he subsequently killed. Thus, in this case, it really would be appropriate to explain the tower's height in terms of the length of the shadow it cast. Van Fraassen's point is that a particular context will fix a particular relevance relation, and this relevance relation will specify the direction of explanation.

While van Fraassen shows that a change in context is sufficient to reverse the direction of explanation, this fact does not by itself show that pragmatic constraints can eliminate spurious explanations generated by the symmetries. What makes the reversal of direction legitimate in the case of the chevalier's tower is that there is an objective relevance relation connecting the chevalier's mental states to his actions (including building the tower). But, as Kitcher and Salmon (1987) point out, van Fraassen's theory places no substantive constraints on the choice of relevance relations. They show how to construct gerrymandered relevance relations meeting van Fraassen's formal criteria but giving rise to spurious explanations like Bromberger's flagpole. One can grant van Fraassen's point that there can be other legitimate relevance relations that give rise to different answers while still maintaining that a central task of a theory of explanation is to describe the kinds of relevance relations that are objectively legitimate. The various explanatory accounts, including Hempel's D-N and I-S models, the statistical relevance model, and Salmon's causal theory, can be seen as attempts to accomplish this task.

Explanatory Unification

The explanatory unification approach, introduced by Friedman (1974) and developed by Kitcher (1981 and 1989), represents another attempt to remedy the inadequacies in the covering-law approach to explanation. While distinctly different from Hempel's version of the covering-law model, explanatory unification is probably more in its spirit than are any of the other models discussed in this essay. Explanations are arguments. What the explanatory unification model adds is the requirement that the arguments used to explain explananda be instances of unifying explanatory patterns. The counterexamples to covering laws (and the D-N model in particular) are thought to be ruled out on the grounds that the spurious explanatory arguments are not unifying.

To understand what is meant by a unifying explanatory pattern, it is useful to consider some examples of unification. Perhaps the greatest of Newton's achievements was to unify celestial and terrestrial mechanics. In practice, what this achievement amounted to was the discovery that the motion of celestial and terrestrial bodies could be explained by deriving the trajectory of that motion from a common set of laws. For instance, the same explanatory pattern can be used both to derive the trajectory of a satellite around the Earth and to derive the trajectory of a projectile like a ballistic missile. As a second example, consider Mendelian explanations of the distribution of traits in successive generations of populations. Mendelian genetics is unifying because it allows for a diverse set of facts about distributions of traits in populations of different species to be explained in terms of common patterns like dominance and recessiveness.

Explanation, according to Kitcher, begins with a set of accepted beliefs *K*. Given this set, the problem is to identify a set of argument patterns, called the "explanatory store," *E*(*K*), from which explananda are derived. The explanatory unification model suggests that *E*(*K*) is the set of argument patterns that maximally unify *K*. Most of Kitcher's work is devoted to spelling out what counts as an argument pattern. For Kitcher, an argument pattern consists of three parts: (1) a set of *schematic sentences* containing dummy letters for some nonlogical vocabulary; (2) a set of *filling instructions* regarding the sorts of terms that can be substituted for the dummy letters, and (3) a *classification* of instructions regarding which sentences are to be regarded as premises and of the rules of inference that may be used to derive conclusions from those premises. Kitcher's argument patterns are similar in some respects to metalinguistic argument schemata familiar from formal logic. The essential difference is that the filling instructions are semantic rather than syntactic. The terms that replace a particular dummy letter need not have precisely the same logical form, but they must belong to a similar semantic category. For instance, a filling instruction for a Newtonian pattern of explanation might specify that the term to replace a dummy variable refer to a body or to a position within a Cartesian coordinate system.

It might seem that the unification approach is open to counterexamples of spurious explanation very similar to those that confront the D-N model. For instance, the explanatory store for all of science could contain just one argument pattern $\frac{P}{P}$ because every statement in *K* is derivable from itself. Kitcher's answer to this objection is that the unifying power of the explanatory store is judged not just on the numbers of argument patterns in it, but on the *stringency* of those patterns. A stringent argument pattern is one whose schematic sentences and filling instructions limit the number of possible ways in which the pattern can be instantiated. The argument pattern above is clearly not stringent. While stringency requirements are a plausible solution to the problem of spurious unification, a major problem for the explanatory unification approach is to find an adequate account of stringency, as well as of the way in which to assess the trade-off between the size of the explanatory store and the stringency of its argument patterns.

Critics of the unification approach have also raised more general concerns. For one thing, it should be noted that what counts as a good explanation is relativized to the current set of accepted beliefs *K*. As those beliefs change, so will the explanations. While it is certainly the case that the arguments *accepted* as explanatory will change with changes in *K*, many philosophers will argue that the true explanation of some fact should remain constant over time and that as beliefs change, formerly accepted explanations are regarded as spurious. A second puzzling feature of the unification approach is what can be called the *nonlocality* of explanation. According to the unification approach, whether something counts as a correct explanation depends upon whether the argument pattern used to explain it is also useful elsewhere. This means that one cannot judge the adequacy of an explanation of a particular event simply by reference to claims about other local events. Such a requirement is at odds with causal approaches to explanation, which suggest that explanations consist in describing events and processes causally relevant to the explanandum event. Whether similar causal patterns actually occur elsewhere is immaterial. Kitcher's (1989) reply to this sort of objection is that the processes identified as causal are just those that can serve in maximally unifying explanatory patterns. Advocates of the causal approach respond that Kitcher has confused ontological and epistemological issues. While considerations like simplicity and scope enter into epistemological judgments about the correctness of theories, what makes an explanation correct is that it describes an actual causal process.

Reductive Explanation

Reduction can be understood as a kind of explanation and can be analyzed using the general models of explanation so far discussed (see Reductionism). Historically, most discussion of reduction has focused on three general areas. The first concerns the attempts of logical empiricists to provide a reduction of theoretical terms to observational terms. The second, generally called successional reduction, involves the study of the relationship between succeeding theories of the same domain, such as the relationship between Newtonian mechanics and the general theory of relativity. The third, generally called interlevel reduction, involves the study of the relationship between theories of different levels of organization, such as the relationship between neurobiological and psychological theories. The focus here will be on the third, which is most clearly a kind of scientific explanation.

The classical model of theoretical reduction is due to Nagel (1961). According to this model, a theory is understood as a collection of laws and other statements. Reduction is accomplished by discovering a set of bridge principles, which are universal biconditional statements that identify terms of the reduced theory with terms of the reducing theory. For instance, if one were to attempt a reduction of classical to molecular genetics, the bridge principle would say "For all x, x is a gene if, and only if, $\phi(x)$," where $\phi(x)$ is some (presumably complicated) formula of molecular biology. In a successful reduction, the laws of the reduced theory should be derivable from the laws (and perhaps other statements) of the reducing theory. A reduction of this sort would satisfy the conditions for a D-N explanation of a general law.

Nagel's account has been criticized by Fodor, Putnam, Wimsatt, and Kitcher, among others (see Sarkar 1992 for a review). Perhaps the most influential of these criticisms is Fodor's (1974) *multiple realizability argument*. Suppose one attempts to give a reduction of a higher-level science like economics to physics. Economic laws (if there are any) will describe relations between theoretical terms like money, debt, interest, etc. What Fodor points out is that the property of, for instance, being money is in fact a functional property. What makes something money is that it plays a certain causal role in an economic system. Many things, from gold to dollar bills to digitized numbers on magnetic media, can perform this causal role, and these things—the realizations of money—have very little in common in terms of their physical properties. It would, consequently, be difficult or impossible to formulate bridge principles. At best one would identify economic predicates with a large disjunction of physical predicates. Even if this were possible, the disjunction would not form a physical kind and the bridge principle would not be lawlike.

The purport of arguments like Fodor's is to establish the explanatory autonomy of special sciences. For instance, it would seem to justify the view that psychological theory can be used to explain psychological events without reference to the physical substrate of psychological agents. However, critics of Fodor and the "antireductionist consensus" point to the various ways in which lower-level theories can increase understanding of a higher level of phenomena. They argue that the explanatory significance of interlevel relations suggests that the problem is not with reductionism per se, but with Nagel's model of reduction. An example of this approach can be found in the work of Kim (2000), who suggests that the properties in higher-level sciences are functional. In Kim's account, a reductive explanation consists in the specification of the mechanism that realizes this function. It is true that a given function may have different realizers in different contexts, but it is nonetheless explanatory to consider how a function is realized in a particular case. Moreover, it is often the case that many or all actual instances of a function may be realized in the same general way. So even if, for instance, having a pain is a functional property, that property is realized by similar mechanisms among all human beings, and to some degree among many other species. If Kim's account is correct, then reductive explanation is possible, but it is closely connected both to functional explanation and to mechanical explanation.

Functional and Teleological Explanation

The ideas of functional and teleological explanation originated with Aristotle's concept of the final cause, the reason or purpose of the existence of a thing. Although the metaphysical assumption that everything in nature has a final cause has generally been rejected since the seventeenth century, teleological explanation is still considered legitimate in areas like biology, where systems are products of design or selection processes. Thus, for instance, if one explains the structure of the human hand in terms of its adaptive function, one is providing a legitimate teleological explanation (see Function).

Hempel (1959) considered how explanations of this kind could be integrated into the framework of the D-N model. He saw the task of functional explanation as that of explaining the presence of a certain part within a complex system by showing that it contributed to that system's functioning. For instance, one could explain the presence of the heart in the human body by showing that its pumping of blood contributes to the proper functioning of the body. If, however, one attempts to frame this explanation as a D-N argument for a particular human body (say Joe's), one gets something like this:

> Joe's body functions properly.
> Pumping blood is an essential activity in the proper functioning of a body, and a heart is a body component that functions as a blood pump.
> Thus, Joe's body has a heart.

Hempel, however, points out that the fact that a component plays an indispensable role in the functioning of a system does not allow one to infer that that component must be present. Because functions are multiply realizable, the same role could be played by something different than a heart (e.g., an artificial heart). The point is even clearer in the case of opposable thumbs. Although opposable thumbs play a certain role in human bodies that is clearly adaptive, one could not *predict* the development of opposable thumbs, because other morphologically distinct traits could perform the same function. Hempel concluded that functional arguments are, at best, very weak explanations, because all that can legitimately be inferred from premises like those in the argument above is that one of an indefinitely large range of realizers will be present in the system.

Hempel's conviction that functional explanations are weak is a consequence of his attempt to fit them within the D-N model, and in particular of his belief in the structural identity of explanation and prediction. Other philosophers, especially Cummins (1975), have argued that these assumptions mistake the distinctive character of functional explanation. According to Cummins, the point of functional explanation is to show not that the presence of a certain component in a system was predictable, but rather how that component contributes to the functioning of the system in which it is contained. Functional analysis involves identifying a certain capacity of a system and showing how more basic capacities of the system or its components give the system that capacity. For instance, functional analysis shows how the heart, arteries, veins, lungs, etc., give the human body its capacity to transport oxygen and other products to its various areas. As Craver (2001) has pointed out, when these more basic capacities are associated with components of a system and when the organization of these components is specified, functional analysis leads to a pattern of explanation very similar to mechanistic explanation in the complex-systems sense. Whatever the success of functional analysis as an explanatory strategy, many philosophers of science still believe that there is a kind of functional explanation in which the adaptive value of a trait explains the presence of a trait.

STUART GLENNAN

References

Bromberger, Sylvain (1966), "Why Questions," in Robert Colodny (ed.), *Mind and Cosmos*. Pittsburgh: University of Pittsburgh Press, 86–111.

Carnap, Rudolf (1950), *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Cartwright, Nancy (1979), "Causal Laws and Effective Strategies," *Noûs* 13: 419–437.

Craver, Carl (2001), "Role Functions, Mechanisms and Hierarchy," *Philosophy of Science* 68: 53–74.

Cummins, Robert (1975), "Functional Analysis," *Journal of Philosophy* 72: 741–765.

Dupré, John, and Nancy Cartwright (1988), "Probability and Causality: Why Hume and Indeterminism Don't Mix," *Noûs* 22: 521–536.

Fodor, Jerry (1974), "Special Sciences, or the Disunity of Sciences as a Working Hypothesis," *Synthese* 28: 97–115.

Friedman, Michael (1974), "Explanation and Scientific Understanding," *Journal of Philosophy* 71: 5–19.

Glennan, Stuart (2002), "Rethinking Mechanistic Explanation," *Philosophy of Science* 69: S342–S353.

Goodman, Nelson (1956), *Fact, Fiction and Forecast*. Indianapolis: Bobbs-Merrill.

Hempel, Carl (1965), "Aspects of Scientific Explanation," in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, 331–496.

——— (1959), "The Logic of Functional Analysis," in Llewellyn Gross (ed.), *Symposium on Social Theory*. New York: Harper & Row.

Hempel, Carl, and Paul Oppenheim (1948), "Studies in the Logic of Explanation," *Philosophy of Science* 15: 135–175.

Hitchcock, Christopher (1995), "Discussion: Salmon on Explanatory Relevance," *Philosophy of Science* 62: 305–320.

Kim, Jaegwon (2000), *Mind in a Physical World: An Essay on the Mind–Body Problem and Mental Causation*. Cambridge, MA: Bradford Books.

Kitcher, Phillip (1981), "Explanatory Unification," *Philosophy of Science* 48: 507–531.

——— (1989), "Explanatory Unification and the Causal Structure of the World," in Philip Kitcher and Wesley Salmon (eds.), *Scientific Explanation: Minnesota Studies in the Philosophy of Science*, vol. 13. Minneapolis: University of Minnesota Press, 410–505.

Kitcher, Phillip, and Wesley Salmon, "Van Fraassen on Explanation," *Journal of Philosophy* 84: 315–330.

Machamer, P., Darden, L., and Carver, C. (2000), "Thinking about Mechanisms," *Philosophy of Science* 67: 1–25.
Nagel, Ernest (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace and World.
Railton, Peter (1978), "A Deductive-Nomological Model of Probabilistic Explanation," *Philosophy of Science* 45: 206–226.
——— (1981), "Probability, Explanation, and Information," *Synthese* 48: 233–256.
Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
——— (1994), "Causality without Counterfactuals," *Philosophy of Science* 61: 297–312.
Salmon, Wesley, Richard Greeno, and Richard Jeffrey (1971), *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.
Sarkar, Sahotra (1992), "Models of Reduction and Categories of Reductionism," *Synthese* 91: 167–194.
Scriven, Michael (1962), "Explanations, Predictions and Laws," in Herbert Feigl and Grover Maxwell (eds.), *Scientific Explanation, Space and Time: Minnesota Studies in the Philosophy of Science*, vol. 3. Minneapolis: University of Minnesota Press.
van Fraassen, Bas (1980), *The Scientific Image*. Oxford: Clarendon Press.
Woodward, James (2000), "Explanation and Invariance in the Special Science," *British Journal for the Philosophy of Science* 51: 197–254.

See also **Carnap, Rudolf; Causality; Function; Hempel, Carl Gustav; Laws of Nature; Mechanism; Prediction; Reductionism; Scientific Models; Theories**

EXPLICATION

Explication is a form of conceptual clarification developed by the philosopher Rudolf Carnap. The purpose of explication is to diminish scientific and philosophical vagueness and confusion. Although the development and advocacy of Carnap's notion of explication was largely due to Carnap himself, many of the ideas and methods presented in his explication project have been incorporated into contemporary analytic philosophy and linguistics (see Carnap, Rudolf).

Carnap on Explication

Carnap's most detailed exposition of his notion of explication appeared in the first chapter of his *Logical Foundations of Probability* (Carnap 1950). There he described explication as the following procedure:

> [*E*]*xplication* consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second. We call the given concept (or the term used for it) the *explicandum*, and the exact concept proposed to take the place of the first (or the term proposed for it) the *explicatum*. The explicandum may belong to everyday language or to a previous stage in the development of scientific language. The explicatum must be given by explicit rules for its use, for example, by a definition which incorporates it into a well-constructed system of scientific either logicomathematical or empirical concepts. (Carnap 1950, 3)

(A similar description appears in Carnap 1947, 7–9). Explication thus involves the replacement of an inexact concept by another, more exact one. Since the explicandum in an explication is replaced rather than elucidated or elaborated upon, explication is distinct from lexical definition, and more closely related to stipulative definition.

Explication is also distinct from the analysis of a concept, where 'analysis' is understood either as the breaking down of a concept into its constituent parts (as in Kant 1965, 48) or as the substitution of an ordinary concept with a formally more precise one for the purposes of clarifying the ordinary concept's ontological commitments (as in Russell 1956). Both of these notions of analysis appear to require that the analysans preserve the meaning of the analysandum in such a way that the former can be viewed as the definiens for the latter (cf. Orilia and Varzi 1998, 107). Carnap imposed no such constraint on explication. While he thought that the explicatum should be similar to the explicandum, he did not require that it function as a definiens for it, and in fact explicitly allowed for the possibility that some loss of the meaning of the explicandum could occur in explication (Carnap 1950, 7). Furthermore, in analysis the existence of distinct and nonequivalent analysanda for the same concept would almost certainly signal an ambiguity in that concept. But in explicating a concept, Carnap thought it possible and at times even desirable to have distinct and nonequivalent explicata for it, even if the concept were unambiguous. This is discussed further below.

Carnap (1950) identifies a simple example of explication in the replacement of the "prescientific" concept 'fish' by the concept *piscis* within a systematic zoology (5–6). The concept 'fish' is vague and broad. It arguably includes, for instance, tadpoles, seals, whales ("Wal*fische*" in German) and possibly other aquatic animals that are not cold-blooded or that do not have gills throughout life. The concept *piscis*, on the other hand, was stipulated to denote just those aquatic animals having the characteristics of being cold-blooded and having gills throughout life. This stipulation was introduced, Carnap thinks, because it was more *fruitful*, within zoology, to classify these animals together. For example, the concept has proved more fruitful with respect to its appearance in laws or useful generalizations: More true and informative generalizations involve *piscis* than *fish* (in the older sense of 'fish'). Within a systematic zoology, *piscis* would function as the explicatum of fish, the explicandum.

Carnap suggested (1950, 3) that his notion of explication was informed by Immanuel Kant's notion of an explicative judgment, in which the concept of the predicate is analyzed within the subject (see Kant 1965, 48), and by Edmund Husserl's notion of an *Explikat*, or the distinct, articulated outcome of an analysis (cf. Husserl 1973, 112ff). These connections are superficial, however, for Carnap's notion of an explication was embedded within the project of a systematic axiomatization of knowledge and discourse that has no correlate in Kant or Husserl. From the perspective of axiomatics and language construction, Carnap's explication project has a rather greater affinity with Leibniz's notion of a "universal language," which was a proposal for a constructed language that would precisely specify and define key philosophical and scientific terms. Unlike Leibniz, however, Carnap would have rejected the existence of a single "correct" language. A more contemporary and direct influence on Carnap's explication project was David Hilbert's pioneering work in formal axiom systems, which Carnap made extensive use of in his later work (see Hilbert, David).

While the explicatum is a new concept that replaces the explicandum, the explicandum nonetheless guides the choice of its replacement, in that Carnap (1950) makes it a requirement on explication that the explicatum be "similar" to the explicandum in the sense that the former can be used in most cases in which the latter is used, although he emphasizes that a "close similarity" is not required (7). Similarity is the first of four conditions that Carnap places on an explicatum (7–8), *viz*.:

1. The explicatum should be *similar* to the explicandum.
2. The explicatum should be given an *exact* specification within a rule-governed system of scientific concepts.
3. The explicatum should be a *fruitful* concept, and in particular allow for the formulation of many universal statements.
4. The explicatum should be as *simple* as possible. (This condition Carnap makes subsidiary to the first three [8]).

Carnap attempted explications of a variety of concepts throughout his later works. Examples of explicanda/explicata pairs proposed by him include: denotation/extension, meaning/intension, logical truth/*L*-truth, logical implication/*L*-implication, empirical truth/semantic truth (all given in Carnap 1947), verification/confirmation, inductive inference/logical probability, and estimation/degree of confirmation (Carnap 1950). As these examples illustrate, Carnap thought that explications could be performed not just on the concepts of empirical science, but on concepts from philosophy or from formal sciences such as set theory.

Consistent with the constraints guiding the process of explication, Carnap sometimes proposed different explicata for the same explicandum. For example, Carnap (1950) offered two distinct explicata for the explicandum "probability" (23f.). For instance, probability is explicated as the degree of confirmation of a hypothesis *H* with respect to an evidence statement. According to another explication, probability is the relative frequency (in the long run) of one property of events or things with respect to another. Carnap recognized that both conceptions of probability were important (25), and he did not regard the existence of different and even incompatible explicata for the same explicandum as an inconsistency or defect by itself. He did, however, regard it as essential that distinct probability concepts be recognized as such, and saw his explication of probability as removing various confusions that had been generated by a failure to clearly identify distinct probability concepts (Carnap 1950, 35).

The exactness condition on explication in particular calls for further discussion, since it leads to Carnap's philosophical framework for the activity of explication. This framework in turn helps to illuminate the remaining conditions and gives some indication of why there are not more than these four.

Exactness in an explication is, for Carnap, ideally accomplished through the *axiomatization* of an area of knowledge. It was Carnap's lifelong conviction that progress within philosophy and science was hampered by the vagueness and imprecision of concepts and theories, which he believed was often manifested in the form of "sterile and useless" metaphysical disputes (Carnap 1963a, 44–5). The construction of axiomatic systems with precisely specified rules was his proposed solution. If a set of basic axioms can be stipulated for some domain of knowledge, and if such axioms can be conjoined with clear definitions of key concepts and with rules governing inferential relations and relations of justification or confirmation, disputes arising from vagueness or imprecision could, Carnap thought, be eliminated.

Consider a simple example axiom system for a theory of the thermal expansion of iron rods (Carnap 1938, 199f). One might lay down a series of syntactical rules that specify sets of typed signs and permissible concatenations of those signs, and then "translate" certain fundamental empirical laws into that language by including unary predicates such as *Sol* and *Fe* and function symbols like *te(x,t), lg(x, t)*, and *th(x)*. A pair of axioms in such a (quantified) logical language might appear as:

$$\text{A1: } (\forall x)(\forall t)(\forall l)(\forall T)[\text{IF } (((((Solx \text{ and } \mathrm{lg}(x,t_1)=l_1) \text{ and } \mathrm{lg}(x,t_2)=l_2)) \text{ and } te(x,t_1)=T_1)) \text{ and } te(x,t_2)=T_2) \text{ and } th(x)=\beta) \text{ THEN } (l_2 = l_1 \times (1+\beta(T_2-T_1)))].$$

$$\text{A2: } (\forall x)(\text{IF } (Solx \text{ and } Fex) \text{ THEN } th(x) = 0.000012).$$

Here *x, β*, and subscripted formulae are real number variables. From this basis semantical rules may be introduced. These rules assign to signs of the primitive, or "ground," type a class of material objects. Such rules would further assign to the predicates *Sol* and *Fe* the properties of being solid and ferrous, respectively, and assign to the functions *te, lg*, and *th* the values of temperature in degrees centigrade (T), length in centimeters, and coefficient of thermal expansion, respectively, for bodies x at times t. Under this interpretation, axiom 1 (A1) is the quantitative law of thermal expansion, and axiom 2 (A2) gives the coefficient of thermal expansion (in the appropriate units) for iron.

Carnap (1938) illustrates how even simple interpreted axiom systems like this can, when conjoined with a mathematical calculus, be used to derive predictions about the changes in length that an iron body will undergo when heated (201–2). Hence, unlike a set of axioms for a formal science such as logic, an interpreted set of axioms for an area of empirical science contains statements that have empirical content, and so allows the derivation of "factual" statements whose truth can be empirically determined, as well as of theorems. This raises questions concerning how the truth of such axioms is to be understood. For instance, is A2 akin to a stipulative definition? If so, what explains the fact that it has merely contingent empirical applicability and seems to have been discovered rather than stipulated? If on the other hand A2 is just a formal expression of an experimentally determined result, what does it mean to regard it as an axiom, as opposed to a true inductive generalization? To the extent that axioms for Carnap are to be treated as akin to stipulative definitions, one can see how philosophers such as Willard Van Quine found it natural to question Carnap's distinction between axioms and other truths of an empirical theory (cf. Quine 1953 and 1966) (see Quine, Willard Van). Yet, Carnap was not oblivious to such concerns as these, as a look at his research into axiomatics reveals.

Explication and Axiomatics

The nature of the relationship of axiom systems to empirical reality was something that Carnap worked throughout his career to clarify. His position changed over time in response to developments in axiomatics in the first half of the twentieth century. One of his major concerns was the status of the concepts implicitly defined by an axiom system. Like his teacher Frege, Carnap regarded implicitly defined concepts, such as 'point' or 'line' in axiomatic geometry, as problematic on the grounds that the law of the excluded middle does not typically hold for them (Carnap 1927, 364–366). It is now evident from his unpublished work on axiomatics that in the late 1920s Carnap (1927) thought that the specification of a consistent set of axioms for a complete and decidable theory T would guarantee the categoricity of T, and conversely (364–365). He (wrongly) believed himself to have a proof of this result, which he called the *Gabelbarkeitssatz* (the results of this unpublished work are presented in Awodey and Carus 2001). Carnap believed that the *Gabelbarkeitssatz* allowed him to "ground" an axiom system *A* in empirical reality in the following sense: Given a demonstration of the categoricity of the theory generated by *A*, it would follow by Carnap's proof that *A* was decidable. This in turn means that the concepts implicitly defined by *A* would be such that the law of the excluded middle held for them. On the basis of this result, Carnap set to work on a general theory of axiomatics.

In 1930, however, this project was abandoned after Carnap became persuaded by Alfred Tarski and Kurt Gödel that the *Gabelbarkeitssatz* was incorrect. Tarski persuaded Carnap to distinguish more clearly between statements framed in the formal language used *in* the axiom system and statements framed in the language used to talk *about* the axiom system (cf. Carnap 1963a, 53–54). A confusion of this distinction arguably lies at the basis of the defective *Gabelbarkeitssatz* proof (Awodey and Carus 2001, 159). And Gödel's incompleteness theorems showed Carnap that there exist categorical axiom systems (such as the Peano axioms formulated in second-order logic) that are not decidable.

Carnap's response to these developments seems to have been to abandon his earlier concerns about implicit definitions and liberalize the constraints on philosophically suitable axiom systems. No longer able to specify a formal feature internal to a system of axioms that would guarantee an application for concepts defined by those axioms, Carnap in 1934 introduced his principle of tolerance, according to which there are no "morals" for logic to obey beyond the constraints of fruitfulness and exactness in specifying a language and axioms formulated in it (Carnap 1937, 51–52) (see Conventionalism). The principle of tolerance was to guide Carnap throughout the remainder of his career, and it illustrates his considered attitude toward the formal systems in terms of which explications are to be performed.

After 1934, the principle of tolerance informed Carnap's treatment of the truth of axioms with empirical content (such as A2). When such axioms were interpreted in such a way that their nonlogical descriptive signs had empirical designata, Carnap came to simply regard them as true only insofar as their truth had been established inductively by observation and experiment (Carnap 1930, 203; see also 1963a, 60). What then renders such statements axioms? His answer relied upon the flexibility enabled by the principle of tolerance. If, as Carnap thought, there are no "facts of the matter" in logic, that is, if there is no single true or correct logic (or language system) that must be accepted, but rather a plurality of logics and language systems, each of which may be engineered to suit particular purposes, then nothing prevents the construction of systems in which certain statements are stipulated to have a privileged role. If making something like A2 an axiom leads to greater simplicity and fruitfulness in the application of mechanics, then it may be made into an axiom, despite the fact that it is regarded as true in virtue of empirical data (and not, say, a priori intuitions). The conceptual framework that one operates with—the language and formalizations of knowledge that are constructed in it—is under human control.

Thus understood as a kind of *linguistic engineering*, explication was justified in part by the very flexibility and plurality of languages and logics that Carnap—having failed to uphold a doctrine of a single "correct" logic—now believed to exist. Nothing prohibits a philosophically minded scientist from modifying a part of language by axiomatizing and precisely defining certain concepts within those axiomatizations to serve as explicata for less clearly defined ordinary concepts, provided that it is useful to do so. Over time, some explicatum might completely displace its explicandum, even in ordinary, nonsystematic linguistic usage. The replacement in everyday parlance of the vague concept *germ* by more precise and scientifically delimited concepts such as *bacterium* and *virus* might provide an example of this tendency.

Although few philosophers, and still fewer scientists, make explicit mention or use of Carnap's explication project, the general idea of providing a formal systematization of a body of knowledge and explicating concepts within it has formed a significant component of research in contemporary linguistics, mathematics, and the philosophy of language. As suggested above, Carnap's own explication projects focused largely on probability and induction (Carnap 1950) and the philosophy of language (Carnap 1947). Both the nature of explication itself and the results of these individual projects have raised philosophical questions, some of which are examined below.

Strawson's Objections

Carnap's treatment of explication as a form of *philosophical* clarification was criticized by P. F. Strawson in an exchange with Carnap (Schilpp 1963). Strawson's critiques, and Carnap's reply to them, help to illuminate aspects of Carnapian explication, as well as to suggest potential weaknesses.

Strawson and Carnap seem to agree that Carnap's method of explication is intended as a way of clarifying philosophical concepts. Strawson thinks of Carnap's method as the construction of a formal system employing concepts that are precisely defined and then comparing the concepts within the constructed system with those that were to be explicated or clarified. Strawson objects that this method is neither the only nor the best way to clear up philosophical perplexities and confusions concerning concepts of ordinary language. He thinks that most philosophical perplexities arise within ordinary discourse, to which certain basic concepts are *essential*. Carnap's approach of constructing a formal system to clarify problems arising within ordinary discourse is "utterly irrelevant" for what is needed:

> It seems *prima facie* evident that to offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant—is a sheer misunderstanding, like offering a textbook on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart. (Strawson 1963, 505)

Strawson's charge here thus appears to be that Carnap's explications will miss analyzing those concepts that he thinks are essential for ordinary discourse, and thereby fail to resolve the philosophical problems that they may give rise to. If our ordinary concepts lead to puzzles, then it would seem to be *these* concepts that one ought to investigate and clarify, and not some other, more or less homologous ones. Carnap responds to this by drawing on his considerably more "tolerant" conception of language, according to which all ordinary concepts are "dispensable" and hence can be replaced when the need (e.g., philosophical perplexity) arises. Carnap thus provides a different analogy:

> A natural language is like a crude, primitive pocketknife, very useful for a hundred different purposes. But for certain specific purposes, special tools are more efficient, e.g., chisels, cutting-machines, and finally the microtome. If we find that the pocketknife is too crude for a given purpose and creates defective products, we shall try to discover the cause of the failure, and then either use the knife more skillfully, or replace it for this special purpose by a more suitable tool, or even invent a new one. The naturalist's thesis is like saying that by using a special tool we evade the problem of the correct use of the crude tool. But would anyone criticize the bacteriologist for using a microtome, and assert that he is evading the problem of correctly using a pocketknife? (Carnap 1963b, 939)

The "linguistic naturalist"—Carnap's term for advocates of Strawson's position—could presumably agree with Carnap that no one is to be criticized for employing concepts or tools as needed. But Carnap's analogy allows for the obvious rejoinders that (1) pocketknives are not replaceable by microtomes for most ordinary uses and (2) someone who was having trouble using a pocketknife in an ordinary circumstance would not be helped in the least by being shown the workings of a microtome. So it is not obvious that Carnap's analogy adequately answers Strawson's charge of the irrelevance of explication for unraveling perplexity involving ordinary notions.

In addition, it may be misleading to treat an entire language as analogous to a tool, as Carnap does here. Just as tools can be used in tandem with each other, so can concepts. Just as it would be a terrible impediment to restrict oneself to a single tool for everything, so too would it be disastrous to get by with a single concept. However, if one thinks of individual concepts as akin to tools rather than whole languages, Strawson's worries about the irrelevance of some tools to the workings of other tools return to salience.

Carnap provides another example, involving the concepts of warmth and temperature, to help clarify the role of explication. Two different people, or one person in different circumstances, might describe the same thing as warm and as not warm. This might lead to questions such as whether warmth is a feature that things can have independently of perceivers. Carnap writes:

> In order to solve this puzzle, we have first to distinguish between the following two concepts: (1) "the thing *x* feels warm to the person *y*" and (2) "the thing *x* is warm", and then to clarify the relation between them. The method and terminology used for this clarification depends upon the specific purpose we may have in mind. First it is indeed possible to clarify the distinction in a simple way in ordinary language. But if we require a more thorough clarification, we must search for explications of the two concepts. The explication of concept (1) may be given in an improved version of the ordinary language concerning perceptions and the like. If a still more exact explication is desired, we may go to the scientific language of psychology. The explication of concept (2) must use an objective language, which may be a carefully selected qualitative part of the ordinary language. If we wish the explicatum to be more precise, then we use the quantitative term "temperature" either as a term of the developed ordinary language, or as a scientific term of the language of physics. (Carnap 1963b, 934)

Although he does not say so explicitly, it seems that Carnap may here be identifying ever greater precision with ever "more thorough clarification." If he is, then Strawson's worries persist. Substituting more and more precise concepts need not yield any clarification of issues pertaining to other (even if less precise) concepts, any more than understanding the workings of a microtome will tell us how to skillfully use a pocketknife. At the very least, Carnap needs to show how clarity is supposed to emerge in such cases. In particular, recall the question of whether warmth is an objective feature of objects, independent of perceivers. Why should we assume as a matter of course that this question becomes uninteresting or obviously answered by pointing out that temperature is a reasonably precisely defined feature of things? Later in his reply to Strawson, Carnap (1963b) adds:

> The process of the acquisition of knowledge begins with common sense knowledge; gradually the methods become more refined and systematic, and thus more scientific. . . . Suppose the statement "it will probably be very hot tomorrow at noon" is made for the purpose of communicating a future state to be expected, perhaps with regard to practical consequences. The use of the explicatum "temperature" instead of "very hot" in the above statement makes it possible to fulfill the same purpose in a more efficient way: "the temperature tomorrow at noon will probably be about so and so much". (934, 936)

Carnap thus thinks of improvements in precision as signs of scientific progress and clarification. But just how invoking temperature has made anything "more efficient" in this context, he does not explain.

Perhaps a better example for Carnap (1963b) is one that he mentions later on in his response to Strawson (939): the solution of Zeno's paradoxes. Zeno, the ancient Eleatic philosopher, raised some well-known perplexities about how motion is possible. For example, the traversal of any finite distance or spatial interval involves crossing infinitely many subintervals and thereby completing an incompletable (infinite) process. But since completing the incompletable is impossible, motion is impossible. Carnap takes it that Zeno's paradoxes of motion are definitively resolved by appeal to technical advances in mathematics involving limits, the real numbers, and other notions. To the extent that this is correct (an interesting issue that cannot be addressed here), it would appear that the construction of a new relatively formal system of concepts has provided a solution or "conceptual clarification" of paradoxes involving ordinary concepts such as motion from one location to another. Strawson does not discuss this example in his paper, but a possible response that he might give leads to another strand of his argument.

Strawson grants that the construction of formal systems of precisely defined concepts, when coupled with the comparison of such constructed concepts to ordinary ones that lead to perplexities, can yield illuminating results. In order to illuminate the ordinary landscape, one might introduce an artificial one with which to compare it, and this sort of comparison and contrast is arguably what helps to resolve Zeno's paradoxes. However, Strawson argues, such applications *presuppose* the sort of analysis of the relations between ordinary concepts of the sort that he takes to be primary and essential. A sketch of the "ordinary" landscape is required before it can be compared with another.

To the extent that commonsense concepts, troubling as they are, are *essential* to ordinary language and action, then it may be that Carnap's "explication" does not help with certain problems involving such concepts that will necessarily arise. To the extent that one thinks that commonsense concepts are dispensable in favor of "more precise" reconstructed concepts, then Strawson's concerns will seem uninteresting. One will naturally focus instead on the construction of new and improved schemes, rather than muddle about with confused ones.

In his concluding remarks in response to Strawson, Carnap (1963b) writes as though there were a well-defined notion of success that could be used to evaluate his program as well as Strawson's, so that one can, as it were, inductively decide which one is better according to the standard:

> We all agree that it is important that good analytic work on philosophical problems be performed. Everyone may do this according to the method which seems the most promising to him. The future will show which of the two methods, or which of the many varieties of each, or which combinations of both, furnishes the best results. (940)

But however laudatory Carnap's tolerance might be in other settings, it is not so clearly appropriate here. Either the philosophical goal is conceptual clarification of ordinary concepts, in which case Strawson's argument (that the examination of those concepts is the most essential component) seems sound; or the goal is the construction of conceptual systems that optimize our capacity to predict and control "experience," in which case Carnap's "constructive" process of "precisification" seems appropriate.

Quine's Objections

Carnap's explication project received criticism from another philosopher, Willard Van Quine. Quine argued at length that there could be no philosophically useful definition of "analyticity" and that the definitions that had been provided for this notion, including Carnap's, suffered from defects such as circularity and an empty extension (see Analyticity). Yet Carnap's explication project seems to require that some statements, such as the axioms and "meaning postulates" of a formalized system in which explications are conducted, have a privileged role, and this role seems to make such statements functionally very similar to the analytic statements that Quine rejected. Indeed, Carnap himself emphasized that the difference between the logical and mathematical formulae of an (interpreted) axiomatized theory, on the one hand, and the physical propositions of that theory, on the other hand, was essential both to the theory and to the clarification that the theory might provide (see e.g., Carnap 1938, 202). Quine found this distinction unintelligible or at best unhelpful.

Quine's criticism of analyticity can be seen to pose a challenge to Carnap's explication project on two fronts. One front concerns the question of whether Carnap was correct in invoking and assigning philosophical importance to a distinction between statements that are true by stipulation and those that are true in virtue of matters of fact. Here the difference between Quine and Carnap needs to be carefully identified. Quine did not deny the possibility that certain statements, such as meaning postulates, could be stipulated to be true (cf. Quine 1953, 34). And as has been noted, Carnap did not deny that a statement that is functioning as an axiom, such as A2 above, could not also be regarded as an empirical generalization. What appeared to separate the two positions was rather each philosopher's appraisal of the significance of elevating empirical generalizations to the status of axioms within the systematization of an area of discourse. Carnap regarded this elevation as an essential component of explication, and thereby of philosophical clarification, while Quine regarded it as a useless and singularly unhelpful bit of stipulation.

Resolving this dispute is difficult, as it was for the original disputants themselves (see, for instance, the correspondence that Quine and Carnap exchanged over the issue in Creath 1990). It is noteworthy that even if Quine's attack on analyticity is judged a failure, its very presence exposes a second front on which Carnap's explication project may be criticized. For Carnap's attempts to defend his use of the term "analytic" against Quine's objections itself involved recourse to explication. Thus in replying to Quine, Carnap wrote that "it is not clear whether [Quine] is asking about the elucidation explicandum, 'analytic,' or about an explicatum. If he means the latter, then it is given in the rules of a semantical system" (Carnap, in Creath 1990, 430).

The problem here is that an appeal to "rules of a semantical system" to clarify "analytic" was exactly the kind of thing that Quine found objectionable. As Quine put it, "the explanation 'true according to the semantical rules of *L*' is unavailing; for the relative term 'semantical rule of ' is as much in need of clarification, at least, as 'analytic for'" (Quine 1953, 34). It is thus understandable that Carnap's claim that philosophical disputes about "analytic" would be resolved by explication failed to placate Quine. Indeed, it appears that Carnap's desire to resolve Quine's concerns about the analytic/synthetic by means of explications risked begging the question in Carnap's favor. On the other hand, there is an argument to be made that Quine's own position at this point is problematic (see Analyticity).

There is some irony in the fact that the very project that Carnap had hoped would lead to the resolution of apparently intractable and seemingly interminable philosophical disputes appeared to be constitutionally incapable of being applied (in a non-question-begging way) to the very dispute with Quine in which Carnap found himself increasingly enmeshed. So at the very least, Quine's objections may expose the presence of philosophical problems that arguably fall outside the purview of explication. In a further irony, Quine and Strawson, who were at loggerheads over the analyticity issue, found themselves allied in their rejection of Carnap's explication project, for both regarded this project as inadequate to the task of philosophical and conceptual clarification, although they did so for very different reasons.

Eric Loomis
Cory Juhl

References

Awodey, S., and A. W. Carus (2001), "Carnap, Completeness, and Categoricity," *Erkenntnis* 54: 145–172.

Carnap, Rudolf (1927), "Eigentliche und Uneigentliche Begriffe," *Symposion* 1: 355–374.

——— (1930), "Bericht über Untersuchungen zur allgemeinen Axiomatik," *Erkenntnis* 1: 303–307.

——— (1938), "Foundations of Logic and Mathematics," in Otto Neurath, Rudolf Carnap, and Charles Morris (eds.), *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press.

——— (1963a), "Intellectual Autobiography," in Paul A. Schilpp (ed.), *The Philosophy of Rudolph Carnap*. La Salle, IL: Open Court, 3–85.

——— (1963b), "P. F. Strawson on Linguistic Naturalism," in Paul A. Schilpp (ed.), *The Philosophy of Rudolph Carnap*. La Salle, IL: Open Court, 933–940.

——— (1950), *Logical Foundations of Probability*. Chicago: University of Chicago Press.

——— (1937), *The Logical Syntax of Language*. London: Routledge and Kegan Paul.

——— (1947), *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Creath, R. (ed.) (1990), *Dear Carnap, Dear Van: The Quine-Carnap correspondence and Other Material*. Berkeley and Los Angeles: University of California Press.

Husserl, Edmund (1973), *Experience and Judgment: Investigations in the Genealogy of Logic*. Translated by James Churchill and Karl Ameriks. Evanston, IL: Northwestern University Press.

Kant, Immanuel (1965), *Critique of Pure Reason*. Translated by Norman Kemp Smith. New York: St. Martin's Press.

Orilia, F., and A. C. Varzi (1998), "A Note on Analysis and Circular Definitions," *Grazer Philosophische Studien* 54: 107–115.

Quine, Willard Van (1966), "Carnap and Logical Truth," in *The Ways of Paradox and Other Essays*. New York: Random House.

——— (1953), "Two Dogmas of Empiricism," in *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 21–46.

Russell, Bertrand (1956), "On Denoting," in R. Marsh (ed.), *Logic and Knowledge*. London: Allen and Unwin, 39–55.

Schilpp, Paul (ed.) (1963), *The Philosophy of Rudolph Carnap*. La Salle, IL: Open Court.

Strawson, P. F. (1963), "Carnap's Views on Constructed Systems versus Natural Languages in Analytic Philosophy," in Paul A. Schilpp (ed.), *The Philosophy of Rudolph Carnap*. La Salle, IL: Open Court, 503–518.

See also **Analyticity; Carnap, Rudolf; Quine, Willard Van**

F

FALSIFICATION

See **Cognitive Significance; Demarcation, Problem of; Popper, Karl Raimund**

FEMINIST PHILOSOPHY OF SCIENCE

This article will encompass philosophical analyses of science undertaken from feminist perspectives. The tradition is part of the larger field of feminist science studies, which includes feminist internal science critique, feminist work in the history of science, and feminist engagements in social studies of science.

Feminist philosophy of science is a dynamic research tradition, loosely delineated by its origins, research questions, and history. Its origins include the emergence in the 1970s of feminist scholarship, in which academics, including scientists, brought the analytic category of gender to bear on research questions, methods, and theories in their fields. In philosophy, feminists analyzed relationships between gender, on the one hand, and theories of ethics, social and political theory, metaphysics, and epistemology, on the other. Although the details varied by area, historical period, and specific theory, feminists found that assumptions about gender informed many philosophical theories, including those about knowledge and science. Some associated men and women with what were argued to be opposing characteristics or categories—respectively, for example, mind and body, reason and emotion, objectivity and subjectivity, culture and nature, and activity and passivity—and took the first characteristic of each pair to be rightfully dominant in relation to or superior to the second. Many incorporated symbolic gender associations in which valued traits or characteristics (e.g., rationality) were closely aligned with traits associated with (stereotypical) masculinity. And many took men, or more accurately some subset of men, as their only or primary subject.

During this period, feminists in a number of sciences found that androcentrism, or "male centeredness," as well as sexism, informed research questions, methods, and hypotheses in their fields. They identified and criticized the emphasis on male behavior and activities in psychology, the social sciences, biobehavioral science, animal sociology, and various fields in the biological sciences. They criticized biological explanations for what were alleged to be differences in behavior, cognitive abilities, and temperament between the sexes. They also explored symbolic gender associations in general views about science, including how traits closely aligned with it, such as detachment and objectivity, were also strongly aligned with (stereotypical) masculinity. Finally, they detailed informal barriers women continued to confront in entering the sciences or succeeding in them. Both bodies of research would contribute to the emergence of feminist philosophy of science in the 1980s.

Developments in the philosophy of science in the 1970s have also influenced the emphases and methods of this tradition. Challenges in the preceding decades to what Suppe (1972) called "the received view" led not only to the abandonment of specific traditional positions, but also to changes in how those in the discipline viewed it. The challenges included arguments against the plausibility of an extrascientific "foundation" for science, and for the theory-ladenness of observation, underdetermination, and versions of holism (see Theories). Together with increased interest in the history of science and the details of scientific practice, both in part a response to the work of Kuhn, these arguments contributed to the emergence of more contextualist approaches in mainstream philosophy of science (see Kuhn, Thomas), to finely focused studies of the special sciences, of the role of community-specific standards in research, and of the role of other contingent and contextual (or external) factors in scientific practice. Like their colleagues, feminist philosophers have explored the implications of these developments, including how they might provide insights into the relationships between gender and science, and their implications for interpretive notions such as social constructivism and realism (see Scientific Realism; Social Constructionism).

Research in feminist philosophy of science has also developed apace with work in other traditions in science studies, including anthropology and sociology of science. Some feminists have found resources in one or more of these traditions (e.g., Barad 1996). Others maintain that there are important differences between their own methods and goals and those characterizing one or more of them (e.g., Harding 1986). In particular, although feminists are interested in understanding the role of social factors in scientific theorizing, many reject the epistemic relativism espoused in early work in the sociology and anthropology of science. Increasingly, however, work in these traditions recognizes the role of the material world in scientific practice, and further engagements between them and feminist philosophy of science are likely.

The traditions and developments noted do not receive equal treatment in this article, but each has influenced the trajectories of feminist philosophy of science.

Research Questions

Feminist philosophy of science encompasses a variety of methods and emphases, and these continue to evolve. But several questions, albeit differently formulated and pursued in the last two decades, inform much of the work undertaken in it. There is the general question:

What are the relationships between social relations (e.g., gender, race, class, culture) and methods (directions and/or content) of the sciences?

Feminist philosophers of science have explored several kinds of relationship between the sciences and their social contexts. As noted earlier, they study the ways in which assumptions concerning gender contribute to scientific questions, methods, hypotheses, and other aspects of scientific practice. In addition, there is increasing (though some would argue still insufficient) attention to the ways in which race, class, and other social relations impact the directions or content of science (e.g., Harding 1991; Schiebinger 2004; Weasel 2004). Conversely, feminists analyze the impact of scientific hypotheses and technologies on cultures as a whole and/or on specific groups, including the recurrent interest in establishing differences in abilities or behavior along the politically salient axes of gender and race (e.g., Schiebinger 2004; Weasel 2004). They also explore the nature and consequences of divisions in cognitive authority between women and men, scientists and laypersons, specialties within science, White persons and persons of color, and Western science and knowledge of other cultures (e.g., Addelson 1983 and 2003; Harding 1991).

These investigations differ in several respects from those that characterized mid-twentieth-century philosophy of science. For one thing, many feminists do not assume that the social or cultural identities of scientists, or divisions in cognitive authority and labor along the lines of gender or other social relations, are of no epistemological consequence. For another, many do not assume that all relationships between social interests and research compromise the science in question. These, they argue, are empirical issues to be investigated on a case-by-case basis (e.g., Anderson 2004; Harding 1986; Keller 1985; Longino 1990 and 1996; Nelson 1990 and 1996; Wylie 2004). Finally, many feminist analyses are empirically based, focusing on specific episodes in the history of science or contemporary research. The emphasis on case studies, also common in naturalized philosophy of science and philosophy of the special sciences, is in part a function of the developments in the philosophy of science noted above. But these methodological approaches and the empirical hypotheses underlying them also trace their roots to the internal science critiques leveled by feminist scientists in the 1970s and 1980s, which are summarized in the next section.

The role of epistemic values in scientific practice, such as simplicity, generality of scope, and conservatism, are of long-standing interest to philosophers of science, including feminists. Among feminists and others, there is now substantial interest in the question, What role(s) do *nonepistemic* values have in scientific practice and what role(s) should such values have?

Again reflecting a break with traditional approaches in the philosophy of science, feminists are among those who explore how nonepistemic values might have a positive role in science and how, whether positively or negatively, they can influence the directions or content of well-regarded research yet remain unrecognized (e.g., Longino 1990; Potter 1989). These two lines of empirical investigation yield another question: *If* relationships are found between social relations and science, and/or between science and so-called nonepistemic values, what are the *normative implications* for scientific practice and for the philosophy of science?

In the 1970s and 1980s, many feminist scientists took it to be an obvious implication of the research they analyzed that science was a more human and culturally bound activity than previously acknowledged in much mainstream philosophy of science, although recognized by previous work in other traditions (e.g., neo-Marxism). So stated, the empirical content of this view is quite vague and its normative implications are unclear. In the intervening years, feminist philosophers of science and scientists have worked to understand both.

The questions just outlined are not unique to feminist philosophy of science, so one might well ask, What work is done by *feminist* in "feminist philosophy of science"? In this article, 'feminist' locates a dynamic research tradition in relation to its history and a now extensive tradition of feminist science studies (cf. Alcoff and Potter 1993). This essay also follows Longino in understanding feminist philosophy of science as a *way of doing* the philosophy of science, not as a specific theory about science. As Longino puts this point, it is a way of engaging in the philosophy of science that reflects a commitment to not let gender "be disappeared," of studying its relationships to science and, as needed, working to change them (Longino 1994). Again, reflecting developments in a range of disciplines and approaches, many feminists have come to recognize that gender is an insufficient variable for understanding women's experiences, including those of women scientists, or for understanding the impact that the sciences have on women. Accordingly, recent work in feminist philosophy of science analyzes the role of other social relations, such as race and culture, in addressing the questions earlier outlined (e.g., the essays in Nelson and Wylie 2004) and makes increasing use of the work done in postcolonial science studies (e.g., Harding 1996 and Schiebinger 2004).

Its origins, core questions, and internal history delineate feminist philosophy of science as a recognizably distinct tradition. But the following discussion also reveals its strong relationships to the broader tradition of the philosophy of science. Feminists have appealed to and built on a number of contextualist approaches and positions, particularly neo-empiricism and naturalism, in the broader discipline. The contrasts earlier emphasized distinguish feminist approaches from mid-twentieth-century philosophy of science and from some still quite traditional approaches in philosophical epistemology.

Feminist Internal Science Critiques: Critical and Constructive

As noted earlier, feminist philosophy of science emerged in part in response to the analyses undertaken by feminist scientists. A continuing focus of feminist internal science critique is the relative underrepresentation of women and minorities in the sciences, and the informal barriers to full participation in the sciences that members of these groups have faced. It first arose as an issue of fairness. But feminist scientists would soon investigate the epistemic consequences of inequities in access to and opportunities in the sciences—that is, their potential consequences for the directions and content of science. It is to such consequences, and the questions they raised for traditional understandings of science, that this section is devoted.

In the 1970s and 1980s, feminists in the social sciences identified several levels of androcentrism in the research questions, methods, and theories of their fields—perhaps of most significance, in research not concerned to identify or explain sex or gender differences. They criticized methodological approaches to and accounts of social life that emphasized men's activities as defining the so-called public sphere and "culture" and associated women with the "private" sphere and reproductive activities, in turn treated as "natural" and without need of explanation. They also criticized the lack of attention to issues of concern to women, including gender discrimination in the workplace and violence against women. These issues, they argued, were of epistemic consequence. For one thing, such accounts of social life were at least incomplete because they ignored the productive and diverse nature of women's activities in specific cultural contexts, as well as phenomena such as rape and domestic violence. For another, they argued, the association of men with culture and production, and of women with nature and reproduction, obscured basic relationships between the domains so dichotomized. Analyses detailing these problems and proposing constructive alternatives were offered in economics (e.g., Hartmann 1981), sociology (e.g., Smith 1987), history (e.g., Kelly-Gadol 1976), anthropology (e.g., Rosaldo and Lamphere 1974), and human evolution (e.g., Tanner and Zihlman 1976). Finally, many argued that the fact that most scientists were White males was somehow implicated in the androcentrism and other biases they were identifying, although few maintained that the bulk of the problems were purposeful.

These critiques and alternatives were paralleled in other sciences. In psychology, feminists criticized models of psychological development and maturity based solely on research involving boys and men, and hypotheses that women's trajectory was "truncated" when it did not fit such models. They maintained not only that the models were likely to be empirically inadequate, but that the alternative trajectories that became visible when the subject pool was enlarged to include women suggested that social factors had more of a role in development than earlier models recognized. They also developed alternative models based on empirical research devoted to women and girls (e.g., Gilligan 1982).

In empirical psychology and developmental biology, feminists argued that laboratory-animal investigations into the effects of prenatal sex hormones on brain development, behavior, and temperament were characterized by circular reasoning and androcentric assumptions. For example, much research began from assumptions linking males with "aggressivity" and "spatial abilities," and females with "receptivity," assumptions that feminist scientists argued unduly influenced the nature of research tests and the interpretation of results (e.g., Bleier 1984). In animal sociology and biobehavioral science, feminists argued that stereotypical gender associations, such as that of males with aggression and dominance hierarchies and of females with passivity and reproduction, shaped organizing principles, observations, and hypotheses. They argued that the models generated not only were incomplete, but distorted relevant phenomena, and offered alternative observations, methods, and hypotheses (e.g., Haraway 1978).

In the 1980s, feminist biologists expanded their critiques beyond research concerned with a biological "origin" for alleged sex differences, to more subtle ways that androcentrism and other cultural assumptions informed the biological sciences. In embryology, they criticized a then exclusive emphasis on androgens and other features of male fetal development in models of "human" fetal development (e.g., Fausto-Sterling 1985). They criticized the imposition of gender connotations and sexual dimorphism on objects that are not sexed, including hormones, the nucleus and cytoplasm, and bacteria (e.g., Biology and Gender Study Group 1988). In evolutionary biology, they criticized the lack of attention to the selection pressures on females (e.g., Hubbard 1982). Feminist biologists (e.g., Bleier 1984) and biophysicist Keller (e.g., Keller 1985) also criticized linear and hierarchical models of biological processes that posited discrete entities and linear trajectories, including models of cellular protein synthesis that posited DNA as the "executive" of the process. Such models, they argued, reflect unwarranted assumptions about the ubiquity of single and dominant causes for complex processes and, in so doing, oversimplify the relevant processes. Some, including Keller, argued that a preference for linear, hierarchical models of biological processes was linked to masculine experience and self-identity issues in cultures fostering strong sex differences in temperament and behavior (Keller 1985). More plausibly, many biologists argued that such models functioned to support empirically inadequate and determinist explanations of alleged psychological and behavioral sex differences (e.g., as offered in human sociobiology in the 1970s and 1980s). As alternatives, they proposed multifactor and nonlinear models of a number of processes, including cellular protein synthesis, fetal brain development, and the relationships between genes and traits (e.g., Bleier 1984; Fausto-Sterling 1985; Keller 1985). These, they argued, make biological determinist explanations of traits and capacities highly implausible. Feminist scientists were not alone in relating linear models to biological determinism. Other scientists concerned with both the empirical adequacy and the political import of such models include Gould (e.g., 1981) and Lewontin (e.g., 1992).

In arguments both critical and constructive, feminist scientists appealed to the epistemic virtues of empirical adequacy, explanatory power, and generality of scope. They also explored social and political issues: how scientific theories can reinforce cultural beliefs and values, and the ways in which such beliefs and values can inform scientific theorizing. And they often identified epistemological questions. For instance, Were the cases involving androcentrism "bad" science and/or idiosyncratic, and thus without implication for "science as usual"? It seemed to many that the answer to this question was no. Social and cultural assumptions had been found to inform mainstream and credible research, much of it not concerned with sex differences, and few thought the androcentrism they were uncovering was conscious. In addition, many feminists recognized a relationship between their own feminist commitments and their ability to recognize problems that many of their colleagues had not (e.g., Bleier 1984; Hubbard 1982; Wylie 1996).

In considering the philosophical import of these internal science critiques, some emphasized social constructivist notions, citing "scientific facts" as human constructions, and scientific theories as "self-fulfilling prophecies" (e.g., Hubbard 1982, 7). But some also recognized the apparent tensions between their dual emphasis on and concern with the empirical and the political. Representative is Keller's (1985) argument that concluding that science is "just politics" would undermine the empirical force of feminist science critiques and that determining "how things are" is crucial in choosing effective courses of action for improving women's lives.

Themes in Feminist Philosophy of Science

The work of feminist philosophers of science is continuous with that of feminist scientists in two ways. Feminist philosophers also explore the role of androcentrism and other contextual factors in the sciences and, reflecting the emphasis in the broader field of philosophy of science, often do so through focused case studies. These include studies of the relative ignorance about female sexuality and anatomy (Tuana 2004), androcentric hypotheses in archaeology and related fields (Wylie 1996), and androcentrism in investigations into relationships between prenatal hormones and sex differences in behavior and/or cognitive abilities (Longino 1990; Nelson 1990). Feminist philosophers have also offered analyses of historical episodes, including those not overtly concerned with gender. For example, Potter has analyzed the role of then current debates concerning gender and Boyle's choice of one formulation of the ideal gas law over another formulation, equally compatible with available data, the metaphysics of which was aligned with liberal political positions concerning gender to which Boyle was opposed (Potter 1989).

Second, feminist philosophers of science often engage epistemological questions initially identified by feminist scientists—for example, What are the ways in which androcentrism can and does inform research and is it only bad science that is so informed? But they also engage the empirical questions cited at the outset of this essay, concerning (1) how precisely social relations and beliefs, such as androcentric and feminist perspectives, *can* come to inform scientific practice and (2) the normative implications for scientific practice and the philosophy of science of the findings of such investigations.

As will become clear, the approaches of feminist philosophers to both the empirical and normative issues often reflect developments in the philosophy of science in the second half of the twentieth century. Many are also keenly aware of the dangers noted by Keller of embracing a thoroughgoing social constructivism. Accordingly, they have sought to develop models of scientific practice and reasoning that encompass the role, suggested by feminist science critiques and developments in the philosophy of science, of both the natural/material world and contextual factors. Feminist models of science, many have argued, must take seriously constructivist insights into the historically and culturally contingent aspects of science "without sacrificing the ability to explain and justify its existence as a reliable, though not foolproof, process" (Alcoff 1989, 122). After all, these philosophers argue, feminist scientists appealed to evidence and to cognitive values such as empirical adequacy in both their critical and constructive engagements with science (e.g., Nelson 1990). Many argued that feminist models of science must also be able to reconceptualize objectivity in ways that disentangle epistemic adequacy from unattainable ideals of value freedom (Harding 1986; Longino 1990; Nelson 1990;

Wylie 2004). So understood, the challenge is to explain "with some precision" how science that is "good on all traditional criteria" can nonetheless be influenced by contextual values such as androcentrism or ethnocentrism (Potter 1989, 132), and how science informed by feminist values may be even better science. Feminist engagements with these issues are extensive, and only a representative sample is mentioned.

Harding's *The Science Question in Feminism* (1986) had a substantial impact on feminist philosophy of science in the 1980s and early 1990s. Harding identified three epistemological frameworks emerging in feminist science studies: feminist empiricism, feminist standpoint theory, and feminist postmodernism. She noted that each represented an effort to revise an earlier, nonfeminist tradition in light of the emergence of feminist science critiques. (The origins of standpoint theory are in Marxism.) Harding urged ambivalence toward the frameworks, predicting their further development in light of one another and feminist science studies. But her analysis was widely understood to suggest that feminist empiricism was the least promising. Harding (1986) argued that unlike the other two frameworks, central tenets of empiricism (her list included individualism and the distinction between contexts of discovery and justification) ruled out any relationship between social movements and scientific progress (24). If correct—that is, if such tenets are inseparable from empiricism, a claim other feminists would dispute (e.g., Longino 1990; Nelson 1990)—then feminist empiricists would be limited to claiming that androcentrism represents a failure to uphold traditional norms, would contribute no new insights into science, and would be unable to cite feminism as enabling feminist science critiques.

In contrast, Harding argued, feminist standpoint theory provides a framework for understanding the emergence of feminist science critiques precisely because it insists on relationships between power and knowledge. As Harding would later formulate this argument, those in dominant positions in political hierarchies (in this case, men) are at an epistemic disadvantage because their specific locations "organize and set limits on what [they] can understand about themselves and the world around them." From such perspectives, "the real relations of humans with each other and with the natural world are not visible" (Harding 1991, 54). Conversely, the activities and experiences of those disadvantaged (in this case, women) can provide a standpoint from which contradictions between reality and the dominant ideology could come to be recognized. This is not to say that such standpoints are inevitable. Given divisions in experiences and labor by gender, feminist standpoints are possible but are achieved through social movements. Harding argued that standpoint theory's hypothesis that some locations allowed for "better" knowledge than others could more reasonably explain both androcentrism in the sciences and feminist scientists' ability to recognize it, and avoid relativism.

In her 1986 analysis, Harding also argued that feminist postmodernist critiques of epistemology indicated that aspects of both feminist empiricism and feminist standpoint theory were problematic. She cited the work of Haraway, Flax, and others as constituting important challenges to "universalizing claims" about the power of reason, science, and the "subject/self" (Harding 1986, 28), which she and they saw as implicit in the other two frameworks.

In the intervening years, feminist philosophers of science have devoted considerable attention to understanding the ways in which scientists and other knowers are "situated" in socially and historically specific contexts, and the epistemic and normative implications of this hypothesis. Most have worked to develop understandings of science that recognize both the epistemic limits of any given location (political, disciplinary, historical, and so forth) *and* the constraints the world imposes. Most work in the tradition has also been committed to a symmetry thesis: that gender analysis is relevant to understanding not just bad science, but *good* science—indeed, even the best. In this respect, their work parallels the *methodological* relativism advocated in the Strong Programme in the Sociology of Knowledge (see Social Constructivism). But, for reasons cited above, many strongly reject the epistemic relativism earlier espoused by its advocates. In what follows, feminists' analyses are grouped according to Harding's three categories in cases in which their authors use them. It will be clear, however, that divisions between feminist empiricism and standpoint theory are increasingly less definitive and that although few feminist philosophers of science have wholeheartedly embraced postmodernism, the latter's arguments against universalizing claims have been influential.

Feminist standpoint theory has developed in response to criticism that early versions presupposed gender essentialism and did not adequately address the differences in women's lives along the axes of race, class, ethnicity, and culture, problems that Harding herself identified in her initial analysis. Some have incorporated divisions by race (e.g., Collins 1990) and class (e.g., Harstock 1983) into their analyses of women's standpoints. Harding and others have also worked to develop the implications of the deep divisions in perspectives and knowledge it suggests and whether, in this and other ways, it entails relativism. Two developments are significant. Harding has developed an argument for how individuals can "reinvent themselves as others," that is, learn and understand the standpoints of those differently situated. And she and others have developed the notion of strong objectivity, which calls not for "nonsituated" perspectives (unattainable according to standpoint theory) but for "reflexivity," for seeking to understand the limits of one's own perspective through active efforts to understand alternatives. Such reflexivity, standpoint theorists argue, can lead to theories that are "less false" (Harding 1991).

Feminist philosophers interested in developing empiricist models challenged Harding's account of empiricism, particularly her arguments that tenets such as those earlier noted are inseparable from empiricism and her account of the limitations of feminist empiricism (see Empiricism). Feminist empiricists have worked to understand and accommodate the evidence for situatedness and for contingent and discipline-specific features of scientific practice suggested by both recent research in the philosophy of science and feminist internal science critiques (e.g., Longino 1990; Nelson 1990). They too reject relativism, and many have adopted or developed holistic models of evidential relations. These assume the general thesis, advanced by Duhem, Hesse, Kuhn, and Quine, that individual hypotheses are constrained by data and convey evidential status on data only as part of larger bodies of theories (see Duhem Thesis; Kuhn, Thomas; Quine, Willard Van Orman). Feminist empiricists have used such models, together with Quine's thesis of underdetermination (see Theories, Underdetermination of), to engage the empirical questions earlier outlined: how androcentric and feminist assumptions can mediate the inferences drawn between data and hypotheses, the interpretation of research results, the hypotheses entertained, and the categories and methods of well-respected research (e.g., Alcoff 1989; Campbell 1998; Longino 1990; Nelson 1990; Potter 1989). In these efforts, they have argued that the bodies of theories within which hypotheses emerge and are accepted are not limited to those of science proper but include social and cultural beliefs, which can operate (often unrecognized) as background assumptions in specific research programs. Feminist empiricists have also explored how underdetermination can enable a role for nonepistemic values in scientific practice (Alcoff 1989; Longino 1990; Nelson 1990). At a more abstract level, some have explored how, in specific cases, nonepistemic values can contribute to the weight scientists attribute to epistemic values that, as Kuhn (1977) argued, cannot be realized simultaneously—for example, simplicity and empirical adequacy (e.g., Anderson 2004; Longino 1996; Ruetsche 2004).

As earlier noted, feminist standpoint theory emphasizes the role of social movements in enabling less distorted "angles of vision" (Wylie 1996). An emphasis on the social nature of science and the normative implications of this feature also characterize work in feminist empiricism. Some use holism, arguments against foundationalism, and the emergence of feminist science critiques to argue against reconstructions and explanations of scientific practice that focus on the reasoning capacities and methods of scientists *qua* individuals. As alternatives, they have developed models that take shared theories, standards, and practices of science communities as their primary focus and that understand the weighting of evidential warrant as a social (and contingent) rather than an individual achievement solely driven by logic and data (e.g., Longino 1990 and 1996; Nelson 1990 and 1996). Such approaches parallel those in so-called mainstream philosophy of science taken by "social empiricists" (e.g., Solomon 2001). Although few advocate scientific realism (Campbell 1998 is an exception), feminist empiricists contend that research assumptions and methods, hypotheses, and the interpretation of data can and should be assessed on the basis of empirical adequacy and other cognitive values.

At the same time, because they recognize that different background assumptions, theories, and interests are at work in contemporary science (of which androcentric and feminist perspectives are but two examples), recent models in feminist empiricism, like others in the philosophy of science, understand the sciences to be more "disunified" than early versions of holism recognized (see Unity and Disunity of Science). Indeed, proposals for ways to enable reflexivity on the part of scientists draw on the perception of such disunities. Some propose more diverse science communities and the development of norms that encourage conceptual discussion and debate, on the grounds that they would lead to more empirically adequate theories (e.g., Longino 1990; Nelson 1990), while others study how disunities have led to advances in specific research programs (e.g., Nelson 1996; Wylie 2004).

Important work in feminist philosophy of science is neither self-identified nor easily categorized in terms of the three frameworks Harding identified.

Barad's and Wylie's work are representative. Barad has developed "agential realism" as a model of the epistemology of science that is both realist and social constructivist (see Scientific Realism; Social Constructionism). She builds from the epistemology she attributes to Bohr to argue that the "phenomena" that are the actual objects of scientific study are simultaneously the products of human construction and of nature. Wylie's work incorporates features of feminist empiricism as well as of feminist standpoint theory (e.g., Wylie 2004). Her model of evidential warrant is like that advocated by feminist empiricists in assuming underdetermination, taking "data" to be the least defeasible aspect, and in seeing the relationship between data and hypotheses as mediated by background assumptions. But Wylie also builds on Hacking's work to argue that the more evidence supporting a hypothesis enjoys a degree of horizontal and/or vertical independence from it, the stronger the hypothesis. In these arguments, Wylie makes use of the standpoint hypothesis that different angles of vision yield different and, in some cases, better hypotheses, and supports her account with case studies from archaeology (Wylie 1996).

Recent work in feminist empiricism is characterized by an understanding that situatedness, holism, and underdetermination entail the contingency of all knowledge claims, including those of science. This view, common also to feminist standpoint theory, is generally taken to call for an understanding of objectivity different from the traditional "the view from nowhere" as well as from "the view from everywhere" associated with relativism. Not unlike the strong objectivity advocated by standpoint theorists, feminist empiricists and some feminist postmodernists have argued that reflexivity is a necessary condition for objectivity in doing science and in the philosophy of science. Representative is Haraway's notion of "partial vision," a metaphor for situatedness and contingency, in which she makes use of both standpoint and postmodern insights:

> Not so perversely, objectivity turns out to be about particular and specific embodiment, and definitely not about the false vision promising transcendence of all limits and responsibility.... Feminist objectivity is about limited location and situated knowledge, not about transcendence and splitting of subject and object. (Haraway 1988, 190)

Directions

A special issue of *Hypatia,* a journal of feminist philosophy, was published in 2004 devoted to feminist science studies (Nelson and Wylie 2004). It reflected many of the themes that have characterized feminist engagements with science and the philosophy of science. A number of its authors were scientists, and the collection as a whole contained numerous cross-disciplinary engagements. The symmetry thesis prominent in the mid- and late 1980s and the interest in carving out a middle ground between oppositional interpretive positions such as scientific realism and social constructivism remained prominent. Notably, however, few contributors felt the need to "defend" feminist science studies from charges of relativism or irrationality, something that feminists did feel the need to do in the 1980s and early 1990s, or to offer full-blown arguments for situatedness, contingency, and constraint. Developments in science studies disciplines, including the philosophy of science, are seen to have made such arguments unnecessary. There were also analyses that advanced the study of relationships among race, gender, culture, and science (Schiebinger 2004; Weasel 2004).

The collection suggested new directions. Some authors extended and revised earlier feminist arguments. For example, some offered more substantive analyses of the nature of noncognitive values and of ways in which they could inform scientific practice than had earlier analyses (Anderson 2004; Ruetsche 2004), and some provided more substance to earlier feminist arguments that science was inherently "social" (Sobstyl 2004). One author (Okruhlik 2004) called on feminist philosophers of science to reassess the Vienna Circle and the usefulness to feminist science studies of the arguments and positions developed by some of its members, such as Neurath (see Neurath, Otto; Vienna Circle). Perhaps also predictive, few contributors identified their methods or approaches as "empiricist," "standpoint," or "postmodernist," although they often incorporated one or more views earlier associated with these approaches. Finally, reflexivity was emphasized by many authors, who discussed what it meant to practice science and/or to engage in the history or philosophy of science as a feminist.

Lynn Hankinson Nelson

References

Addelson, Kathryn Pyne (1983), "The Man of Professional Wisdom," in S. Harding and M. Hintikka (eds.), *Discovering Reality*. Dordrecht, Netherlands: Kluwer, 165–186.

——— (2003), "Naturalizing Quine," in L. H. Nelson and J. Nelson (eds.), *Feminist Interpretations of Willard Van Quine*. University Park: Penn State University Press, 241–268.

Anderson, Elizabeth (2004), "Uses of Value Judgments in Science," *Hypatia* 19: 1–24.

Alcoff, Linda Martin (1989), "Justifying Feminist Social Science," in N. Tuana (ed.), *Feminism and Science*. Bloomington: Indiana University Press, 85–103.

Alcoff, L. M., and Elizabeth Potter (eds.) (1993), *Feminist Epistemologies*. New York and London: Routledge.

Barad, Karen (1996), "Meeting the Universe Halfway: Realism and Social Constructivism without Contradiction," in L. H. Nelson and J. Nelson (eds.), *Feminism, Science, and the Philosophy of Science*. Dordrecht, Netherlands: Kluwer, 161–194.

Biology and Gender Study Group (1988), "The Importance of Feminist Critique for Contemporary Cell Biology," *Hypatia* 3: 61–76.

Bleier, Ruth (1984), *Science and Gender*. New York: Pergamon Press.

Campbell, Richmond (1998), *Illusions of Paradox: A Feminist Epistemology Naturalized*. Ithaca, NY: Cornell University Press.

Collins, Patricia Hill (1990), *Black Feminist Thought*. Boston: Unwin Hyman.

Fausto-Sterling, Anne (1985), *Myths of Gender*. New York: Basic Books.

Gilligan, Carol (1982), *In a Different Voice*. Cambridge, MA: Harvard University Press.

Gould, S. J. (1981), *The Mismeasure of Man*. New York: Norton.

Haraway, Donna (1978), "Animal Sociology and a Natural Economy of the Body Politic, Part I," *Signs* 4: 21–36.

——— (1988), "Situated Knowledges," *Feminist Studies* 14: 575–599.

Harding, Sandra G. (1986), *The Science Question in Feminism*. Ithaca, NY: Cornell University Press.

——— (1991), *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Ithaca, NY: Cornell University Press.

——— (1996), "Multicultural and Global Feminist Philosophies of Science: Resources and Challenges," in L. H. Nelson and J. Nelson (eds.), *Feminism, Science, and the Philosophy of Science*, Dordrecht, Netherlands: Kluwer.

Harstock, Nancy (1983), "The Feminist Standpoint," in S. Harding and M. Hintikka (eds.), *Discovering Reality*. Dordrecht, Netherlands: Kluwer Academic Press, 283–310.

Hartmann, Heidi (1981), "The Family As the Locus of Gender, Class, and Political Struggle," *Signs* 6: 366–394.

Hubbard, Ruth (1982), "Have Only Men Evolved?" in R. Hubbard, M. Henifin, and B. Fried (eds.), *Biological Woman—The Convenient Myth*. Cambridge, MA: Schenkman, 7–36.

Keller, Evelyn Fox (1985), *Reflections on Gender and Science*. New Haven, CT: Yale University Press.

Kelly-Gadol, J. (1976), "The Social Relations of the Sexes: Methodological Implications of Women's History," *Signs* 1: 809–823.

Kuhn, Thomas (1977), *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

Lewontin, R. C. (1992), *Biology as Ideology: The Doctrine of DNA*. New York: HarperPerennial.

Longino, Helen E. (1990), *Science as Social Knowledge*. Princeton, NJ: Princeton University Press.

——— (1994), "In Search of Feminist Epistemology," *The Monist* 77: 472–485.

——— (1996), "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy," in L. H. Nelson and J. Nelson (eds.), *Feminism, Science, and the Philosophy of Science*. Dordrecht, Netherlands: Kluwer, 39–58.

Nelson, Lynn Hankinson (1990), *Who Knows: From Quine to Feminist Empiricism*. Philadelphia, PA: Temple University Press.

——— (1996), "Empiricism without Dogmas," in L. H. Nelson and J. Nelson (eds.), *Feminism, Science, and the Philosophy of Science*. Dordrecht, Netherlands: Kluwer, 95–120.

Nelson, Lynn Hankinson, and Alison Wylie (eds.) (2004), Special Issue of *Hypatia: Feminist Science Studies* 19(1).

Okruhlik, Kathleen (2004), "Logical Empiricism, Feminism, and Neurath's Auxiliary Motive," *Hypatia* 19: 48–72.

Potter, Elizabeth (1989), "Modeling the Gender Politics in Science," in N. Tuana (ed.), *Feminism and Science*. Bloomington: Indiana University Press, 132–146.

Ruetsche, Laura (2004), "Virtue and Contingent History: Possibilities for Feminist Epistemology," *Hypatia* 19: 73–101.

Rosaldo, Michelle Z., and Louise Lamphere (eds.) (1974), *Woman, Culture, and Society*. Stanford, CA: Stanford University Press.

Schiebinger, Londa (2004), "Feminist History of Colonial Science," *Hypatia* 19: 233–254.

Smith, Dorothy (1987), *The Everyday World As Problematic: A Feminist Sociology*. Boston: Northeastern University Press.

Sobstyl, Edrie (2004), "Re-Radicalizing Nelson's Feminist Empiricism," *Hypatia* 19: 119–141.

Solomon, Miriam (2001), *Social Empiricism*. Cambridge, MA: Bradford Books/MIT Press.

Suppe, Frederick (1972), "What's Wrong with the Received View on the Structure of Scientific Theories?" *Philosophy of Science* 39: 1–19.

Tanner, Nancy, and A. Zihlman (1976), "Women in Evolution," *Signs* 1: 585–608.

Tuana, Nancy (ed.) (2004), "Coming to Understand: Orgasm and the Epistemology of Ignorance," *Hypatia* 19: 194–232.

Weasel, Lisa H. (2004), "Feminist Intersections in Science: Race, Gender and Sexuality through the Microscope," *Hypatia* 19: 183–193.

Wylie, Alison (1996), "The Constitution of Archaeological Evidence: Gender Politics and Science," in P. Galison and D.J. Stump (eds.), *The Disunity of Science*. Stanford, CA: Stanford University Press, 311–343.

——— (2004), "Why Standpoint Matters," in R. Figueroa and S. Harding (eds.), *Science and Other Cultures*. New York: Routledge, 26–48.

See also **Empiricism; Instrumentalism; Logical Empiricism; Scientific Realism; Social Constructivism; Unity and Disunity of Science**

PAUL KARL FEYERABEND

(13 January 1924–11 February 1994)

A Viennese émigré, Paul Feyerabend taught philosophy of science wherever his restless nature brought him—especially Berkeley, California; London, Auckland, Berlin, and Zürich. His views on methodology and the politics of science established him as one of the most controversial, eccentric, and outrageous figures in contemporary philosophy. Allegedly an irrational thinker, Feyerabend was in fact a skeptical master and iconoclast about the sciences and their philosophy. He denounced the gap between abstract normative philosophical accounts of science and actual, complex, and context-dependent scientific practice. He argued against the hegemony of any intellectual or ideological vision to promote the advantages of tolerance and pluralism in science as well as in society. His anarchistic theory of knowledge and the willingness to question the supremacy of Western scientific rationality vis-à-vis other "forms of life" made him famous beyond the boundaries of the philosophy of science.

A Philosophical Life Spent "Killing Time" and Scientific Idols

Paul Karl Feyerabend was born in Vienna in 1924. As a young man he was attracted to physics, mathematics, and astronomy (a passionate observer through the telescope he built with his father), as well as to drama, cinema, singing, and opera. Four years after the anschluss of Austria by the Third Reich in 1938, he was drafted into the Nazi work service and later entered the German army.

Posted to battle on the Russian front, he was awarded the Iron Cross. The end of the war saw him recovering from a bullet wound in his spine, which was to leave him crippled. He was granted state funding to study singing and stage management, and also cultivated Italian, harmony, piano, and diction. He then decided to study history and sociology in Vienna, but soon changed to theoretical physics and generally adhered to a positivistic scientism, which regarded science as an empirical activity and the basis of all knowledge.

During the following years Feyerabend received his Ph.D. in Philosophy with a dissertation on "basic statements" supervised by Viktor Kraft, and crossed Karl Popper's path for the first time (see Popper, Karl Raimund). He also met Bertholt Brecht, turning down an offer to work as his production assistant ("one of the greatest mistakes of my life," he would later say, adding, however, that as with Marxism and the army, he would probably not have enjoyed the gregarious group mentality prevalent in Brecht's circle).

In 1952, Feyerabend left for Cambridge, England, hoping to study under Wittgenstein; when the latter died, Feyerabend turned to the London School of Economics, where he was supervised by Popper, and genuinely embraced falsificationism. His adherence to it, however, was fairly unorthodox, combining realism and the view that all (observational) terms are theoretical with the principle of tenacity (the idea that it is rational to keep working on a theory despite empirical anomalies) and theoretical pluralism. A year later, he declined the offer of a job as Popper's assistant and left for Vienna.

In 1955, the University of Bristol, England, granted him his first academic post as lecturer in philosophy of science. During the following years Feyerabend confirmed his decision to cut all ties with what he later called the "Popperian Church," a group of scholars who preached but did not practice the critical attitude that plays a central role in Popper's philosophy. From 1958 to 1990 (the year he tendered his official resignation), Feyerabend was lecturer and then professor at the University of California–Berkeley, spending much time both in the United States (Yale University and Minnesota) and abroad (London, Berlin, Auckland, Brighton, Kassel), wherever his restlessness and growing fame took him. During the 1980s, Feyerabend accepted a chair at the Zürich Polytechnic ("ten wonderful years of half-Berkeley, half-Switzerland"). Struck by a brain tumor, he died on February 11, 1994, in Grenolier, Switzerland.

Explanation, Reduction, and Empiricism (Feyerabend 1962) marks both Feyerabend's departure from a foundationalist conception of experience and his endorsement of some of Wittgenstein's later views. Feyerabend argues against the logical empiricist accounts of explanation, theoretical reduction, and meaning invariance (see Explanation; Logical Empiricism; Reductionism). He also derives the methodological implications of his "contextual theory of meaning" and "incommensurability thesis" based on detailed historical examples. During his frequent visits to the London School of Economics, Feyerabend met Imre Lakatos, who encouraged him to collect the impertinent ideas expounded in his lectures about the nonexistence of scientific method. Lakatos was supposed to reply and defend rationality, but their joint project—provisionally titled *For and Against Method*—was never completed. Lakatos unexpectedly died in 1974, and Feyerabend's part of the project, *Against Method*, was ultimately published as a collection of essays (Feyerabend 1975). The publication of the long correspondence between Feyerabend and Lakatos (Feyerabend and Lakatos 1999) partially filled this gap and fully acknowledged the dialectical exchange of ideas between the two friends that helped sharpen Feyerabend's attack on the rationalist position.

While *Against Method* denounces the dichotomous and enigmatic relation between philosophical theories and scientific practice and advocates the freedom of science from the interference of philosophy, *Science in a Free Society* (Feyerabend 1978) argues for the freedom of all "forms of life" from the interference of science. In this book, Feyerabend complains about the "illiteracy" with which his previous book was received, but also elaborates on the political consequences of his epistemological anarchism, argues for the separation of science and state and for the equal right to survival and access to power of all traditions (including those in conflict with accepted "scientific truths"). Feyerabend's corrosive skepticism is here directed toward the uncontrolled and uncritical, yet all-powerful, authority of "scientific expertise." Feyerabend claims that in order to defend society against science, the latter has to be placed under the supervision of democratic councils of laymen—with the aim of assessing and counterbalancing experts' judgments and decisions.

Feyerabend's attempt to dethrone science from its privileged position within Western culture is also carried on in his later writings collected in *Farewell to Reason* (Feyerabend 1987), a *sui generis* apology for cultural relativism. Reviving John Stuart Mill's argument on the means of cultivating human flourishing, Feyerabend argues that the freedom of a society increases as the restrictions imposed on its traditions are removed. Moreover, societies that contain many traditions side by side and stimulate cultural diversity have a better chance to enhance both the quality of the traditions and the maturity of their citizens. The citizens, in turn, should be prepared to use the standards of the traditions to which they belong to judge and supervise the institutions. In Feyerabend's view, this constitutes the best antidote to cultural and political totalitarianism.

The Refutation of Classical and Logical Empiricism, or How to Be a Good Empiricist

Feyerabend's first iconoclastic enterprise is directed against philosophical empiricism: the view that what is to be believed is what experiences establish, and no more. In fact, Feyerabend's line of attack is broad and applies to any foundationalist epistemology (see Epistemology). A naïve appeal to experience assumes that the meaning of observational terms is unequivocally determined by the procedures of observation such as looking, listening, and the like, and that scientific theories can be grounded in independently meaningful facts thus established. To Feyerabend, this view is at variance with actual scientific practice. Moreover, empiricism in the form theorized by logical empiricist philosophers cannot contribute to the growth of knowledge; on the contrary, it is bound to lead to "a dogmatic petrifaction" of theories and "the establishment of a rigid metaphysics" (Feyerabend 1999a, 82).

Feyerabend's argument moves from the consideration that theories are all-pervading conceptualizations of the world and determine the vocabulary that is used in building up "facts." This is in particular the case with the observation-language reputed to ground scientific theories (see Observation; Theories). Feyerabend's first main thesis is that "the interpretation of an observation-language is determined by the theories we use to explain what we observe, and it changes as soon as those theories change" (1981, 31).

In principle, according to Feyerabend, all observational terms are fully theoretical, and there is no semantic difference between theoretical terms and observational terms. Thus, observational terms are neither certain nor stable but share the hypothetical and changing nature of theoretical terms. The consequences for the relation between theory and experience are radical. Crucially, if meanings of observational terms depend on the universal principles of the theory in which they are used, terms that depend on different universal principles will not share the same meaning. Feyerabend then, anticipating some of Kuhn's ideas, argues that theory testing cannot be a matter of confrontation of theory and (theory-laden) empirical data; rather it is a matter of competition between theories that are in part mutually exclusive, or *incommensurable* (see Incommensurability; Kuhn,Thomas).

Theories are incommensurable when the universal principles used to determine the concepts within one theory "suspend" the universal principles of the other, and thus all its facts and concepts. Classical Newtonian mechanics, for example, is said to be incommensurable with relativistic mechanics on the basis that the latter rejects a universal principle of the former "that shapes, masses, periods are changed only by physical interactions" (Feyerabend 1975, 269–271). Consider, in particular, the concept of 'length.' In classical mechanics, length is a relation that is independent of signal velocity, gravitational fields, and the motion of the observers; whereas in relativistic mechanics the value of length depends on these very concepts. The switch from classical mechanics to relativity entails a change of meaning of spatio-temporal concepts (see Classical Mechanics; Space-Time). Classical length and relativistic length are incommensurable notions, and classical mechanics is not explained by, or "reducible to," Einstein's relativity theory (Feyerabend 1981, 76–81). In general, according to Feyerabend, any attempt to derive the universal principles of an old theory from those of a new one necessarily leads to a change of the meanings in the old theory's terms. And this is why the "theoretical reduction" fostered by the orthodox account of explanation is not viable. Feyerabend's second main thesis is thus that there is not any reduction of a theory to another in actual science, but rather a replacement of one theory and its "ontology" with another (1999a, 86–87).

The question now is raised of "how to be a good empiricist." For Feyerabend a good empiricist is a *critical metaphysician:*

> His first step will be the formulation of fairly general assumptions which are not yet directly connected with observations; this means that his first step will be the invention of a new metaphysics. This metaphysics must then be elaborated in sufficient detail in order to be able to compete [with] the theory to be investigated as regards generality, details of prediction, precision of formulation.... Elimination of all metaphysics, far from increasing the empirical content of the remaining theories, is liable to turn these theories into dogmas. (1999a, 102)

However, it should be noticed that contrary to what many critics have claimed, Feyerabend's incommensurability thesis should not be interpreted as maintaining that competing theories cannot be compared. What his thesis entails is that theories cannot be compared in the ways in which many philosophical accounts of scientific explanation and reduction have thought that such comparisons should occur. To reject these accounts is to raise problems about certain philosophical theories of science; it is not to raise any difficulties for scientific practice itself (1981, xi).

Against (Too Much) Method

Against Method aims at demystifying another philosophical idol: the existence of a strictly binding system of rules for (good) scientific practice. Feyerabend highlights the huge gap between the "real thing" (science) and the various images of science. His therapy for philosophers' schizophrenic detachment from scientific reality is methodological anarchism. The therapy is the result of historical analyses. In particular, careful historical investigation supports the thesis that

> [t]here is not a single rule, however plausible, and however firmly grounded in epistemology, that is not violated some time or another.... Such violations are not accidental events. On the contrary we see they are necessary for progress.... The Copernican Revolution, the rise of modern atomism, the gradual emergence of [the] wave theory of life, occurred because some thinkers either *decided* not to be bound by certain "obvious" methodological rules, or because they *unwittingly broke* them. (1993, 14)

If this is the case, then any attempt to reform science by bringing it closer to the abstract image philosophers have of *the* scientific method is bound to damage science. On the contrary, "the only principle that does not inhibit progress is: *anything goes*"(1993, 5). Anything goes (perhaps paradoxically) is also the only general principle to which the coherent rationalist can be committed if looking for a rule valid in all given historical situations. But at the same time—at least in Feyerabend's intention—it is not introduced to replace one set of general rules by another set, but rather "to convince the reader that all methodologies, even the most obvious ones, have their own limits" (1993, 23). Consider for example the application of a clear, well-defined, and well-regarded rule like the consistency condition. According to it, the new hypotheses should agree with the accepted theories. But for Feyerabend it is not a reasonable condition at all. In fact, instead of being of help in obtaining better theories, it is just a factor for preservation of the old ones. Hypotheses contradicting well-confirmed theories should proliferate and not be restricted, because they help provide (theory-laden) evidence that cannot be obtained in any other way. Consider also the rule that a theory that contradicts experience should be excluded from science. This rule, Feyerabend claims, is violated at every run:

> [T]heories are refuted in every moment of their existence . . . *ad hoc* hypotheses patch up gaps in the proofs and cracks in the connection of facts. And internal contradictions are almost never avoided. We do not have proud cathedrals standing before us, instead we have dilapidated ruins, architectural monstrosities whose precarious existence is laboriously prolonged through ugly patch-work by their constructors. This is scientific reality. (1981, 156)

Scientific reality is always too rich in content, too varied, too many-sided, too lively and subtle to be captured by the simple-minded rules of even the best philosophers or historians. Scientists are not rule-followers but opportunists. In the construction of their conceptual world, they cannot be restricted by the adherence to any epistemological system; rather they rely "now on one trick now on the other" (1993, 1). Galileo Galilei's cunning defense of the heliocentric cosmology is paradigmatic in this respect. According to Feyerabend, not only did Galileo develop a research program in striking contrast to the Aristotelian standards and the accepted observation of the time, he was also prepared to defend it by substituting a "natural" interpretation of motion (motion can be expressed only in terms of observable changes) with an "unnatural" and highly theoretical concept, which introduced into the phenomenon of motion some components (such as circular inertia) that cannot be observed. In this way Galileo was able to "defuse a mine" placed under the Copernican system by explaining away the objection regarding the motion of the Earth. This move was possible because people see a phenomenon and interpret it in what they regard as a natural way according with their beliefs. So it is the *interpretation* of the phenomenon and *not* the phenomenon itself that is in contradiction with a given belief. Galileo then resolved the contradiction between empirical observation and the Copernican view by providing a new and highly abstract observational language and thus a newly constructed empirical basis. This, in turn, was a new theory of interpretation (containing the idea of the relativity of motion and the law of circular inertia) fitting the Copernican system (Feyerabend 1993, 55–85).

Galileo also changed the "sensory core" of observational statements that seemed to contradict Copernicus. He claimed to have removed them with the help of a 'superior and better sense' for astronomical matters, the telescope. However, Feyerabend points out, Galileo had no theoretical reasons to support the conclusion that the telescopic phenomena are more veridical than observations by the unaided eye. Once again, behind the clashes of the senses, there was a clash of theoretical assumptions, explicit or not. Galileo chose the research program that promised him the most exciting discoveries and adopted propaganda strategies in which reason was not enough to defend it against the widely accepted methodological canons:

> We see that Galileo's view of the origin of Copernicanism differs markedly from the more familiar historical accounts. He neither points to new facts which offer inductive support [for] the idea of a moving earth, nor does he mention any observations that would refute the geocentric point of view but be accounted for by Copernicanism. On the contrary, he emphasizes that not only Ptolemy, but Copernicus as well, is refuted by the facts, and he praises Aristarchus and Copernicus for not having given up in the face of such tremendous difficulties. He praises them for having proceeded *counterinductively.* (Feyerabend 1993, 80–81)

That is, Galileo wins the battle against the Ptolemiac system by subverting the most carefully established observational results and challenging the most plausible theoretical principles.

The Value of Theoretical Pluralism

Counterinduction can be beneficial to the advancement of science. Even Feyerabend's anarchism, then, provides some positive prescriptions. In particular, counterinductive hypotheses are valuable because they provide a means of criticizing accepted theories in a manner that goes beyond the comparison of the theories with the "facts." He says that

> the only way of arriving at a useful judgment of what is supposed to be the truth, or the correct procedure, is to become acquainted with the widest possible range of alternatives. . . . The reasons were explained by John Stuart Mill in his immortal essay *On Liberty.* It is not possible to improve upon his arguments. (Feyerabend 1978, 86)

One of the arguments Feyerabend is referring to is that silencing the expression of an opinion robs the human race by reducing the opportunity to ascertain truth. The role of tolerant controversy in grounding knowledge is so important that, according to Mill, "if opponents of all important truth do not exist, it is indispensable to imagine them, and supply them with the strongest arguments which the most skilful devil's advocate can conjure up" (Mill [1859] 1977, 229).

Accordingly, science should be organized to generate the continuous generation of alternatives, to strengthen anomalies, and to stimulate controversies. The legacy of Mill's liberal standpoint is what Feyerabend calls the principle of proliferation: "Invent, and elaborate theories which are inconsistent with the accepted point of view, even if the latter should happen to be highly confirmed and generally accepted" (Feyerabend 1981, 105). Of course, knowledge generated by such a principle is of a peculiar sort: It is not a series of self-consistent theories that converges toward an ideal view; it is not a gradual approach to the truth. Rather,

> It is an ever increased ocean of mutually incompatible alternatives. Each single theory, each fairly-tale, each myth is part of the collection forcing others into greater articulation and all of them contributing, via this process of competition, to the development of our consciousness. (1993, 21)

As a consequence, "experts and laymen, professionals and dilettanti, truth-freaks and liars—they all are invited to participate in the contest and to make their contribution to the enrichment of our culture" (Feyerabend 1993, 21). Democratic participation in scientific matters warrants the advocacy of minority opinions and thus sustains the conditions for scientific development and human flourishing. This last consideration leads to the question of "science versus democracy," which in his later years Feyerabend (1993, 3) regarded as most important: "My main motive is humanitarian, not intellectual .... I want to support people, not 'advance knowledge.'" In particular, provided that there is no abstract canon ensuring success in any given field of enquiry, and that scientific achievements can be judged only *after* the event, Feyerabend claims that scientists are no better off than anybody else in these matters. The public, therefore, not only can take part in scientific decisions, but *should* do so:

> [F]irst, because it is a concerned party; secondly, because such participation is the best education the public can get—a full democratisation of science is not in conflict with science. It is in conflict with a philosophy, often called "Rationalism" that uses a frozen image of science to terrorize people with its practice. (1993, xii)

So the humanitarian motive behind Feyerabend's debunking of science is clear: Scientists should adapt their procedures and goals to the values of the people they are supposed to advise. Feyerabend is not against science so understood—"Such a science is one of the most wonderful inventions of the human mind"—he is "against ideologies that use the name of science for cultural murder" (1993, 4).

Relativism and Beyond

Two more consequences emerge from the thesis that the sciences have no common structure, but local and distinct features. First, "the success of 'science' cannot be used as an argument for treating as yet unresolved problems in a standardized way" (Feyerabend 1993, 2); second, "'non-scientific' procedures cannot be pushed outside by arguments" (ibid). The political implication of this epistemological stand is *democratic relativism*, the view that all traditions have equal rights. Democratic relativism, in turn, denies the right of traditions to impose their "form of life" on others, and therefore recommends the protection of traditions from interference from outside, including the interferences of the tradition of Western scientific rationalism. A new question then arises: How is a citizen to judge the suggestions issuing from the institutions that surround him? It is assumed that the citizen will judge "rationally," that is, in accordance with some scientific standards. However, there are no unambiguous scientific standards. Feyerabend's answer is that in a "free society," a citizen will use the standards of the tradition to which the citizen belongs: "Hopi standards if he is Hopi; fundamentalist Protestant standards if he is Fundamentalist; ancient Jewish standards if he belongs to a group trying to revive ancient Jewish traditions" (Feyerabend 1999a, 220). To those who claim the superiority of Western achievements over other traditions, Feyerabend simply objects that such a claim needs to be backed up by comparative studies:

> The sciences, it is said, are uniformly better than all alternatives—but where is the evidence to support this claim? Where, for example, are the control groups which show the uniform (and not only the occasional) superiority of Western scientific medicine over the medicine of the *Nei Ching*? Or over Hopi medicine? Such control groups need patients that have been treated in the Hopi manner, or in the Chinese manner using Hopi experts and experts in traditional Chinese medicine. (1999a, 221)

In his later years, however, Feyerabend acknowledged that relativism can run into trouble: It reflects on traditions "from afar" in an abstract and unrealistic way. Traditions are not closed units, they are not frozen systems of thought:

> Traditions not only have no well-defined boundaries, but contain ambiguities and methods of change which enable their members to think and act as if no boundaries existed: *potentially every tradition is all traditions*. Relativizing existence to a single "conceptual system" that is then closed off from the rest and presented in unambiguous details mutilates real traditions and creates a chimera. (quoted in Munévar 2000, 76)

The same, of course, applies to the tradition of scientific rationality. If scientific rationality were characterised as a well-defined, unambiguous, and "closed" system of rules, then relativism would be correct. On the contrary, scientific theories are not unified semantic domains with rigid borders; they change, they borrow from others, and they adapt to new situations. And so is the case for scientific procedures and value judgments; they are continually adapted to circumstances in an open-ended historical process. After all, Feyerabend clarifies, incommensurability is a difficulty for philosophers, not for scientists—the latter being "experts in the art of arguing across lines which philosophers regard as insuperable boundaries of the discourse" (Feyerabend 1987, 272).

Posthumous Works and the Legacy of Paulus Empiricus

At the time of his death, Feyerabend was working on *The Conquest of Abundance* (1999b), developing the theme of how different traditions, or forms of life, can learn from each other and can grow out of each other. The target here (as elsewhere) is the hegemony of any intellectual or ideological single vision—in particular the entire tradition of rationalism and its heirs. The subtitle ("A Tale of Abstraction Versus the Richness of Being") hints at the poverty of the "reality" produced by the method of abstraction typical of Western thought, compared with the abundance, richness, and boundless variety of the world around us. In a "Letter to the Reader" (quoted in Hacking 2000), Feyerabend also makes clear how to approach this text and possibly all his work, which he regarded as specially constructed plays to be performed in the theatre of ideas:

> I want you to sense chaos where first you noticed an orderly arrangement of well-behaved things and processes.... This, my dear reader, is the warning I want you to remember from time to time and especially when the story seems to become so definite that it almost turns into a clearly thought-out and precisely structured point of view. (Hacking 2000, 28)

Feyerabend's (1996) fascinating autobiography shows that he often changed his mind on a variety of subjects, but it also proves that he was neither the worst enemy of science, as depicted by some of his commentators, nor the irrationalist philosopher criticized by most of the profession. He was a skeptic about the foundations of knowledge and a cunning rhetorician who knew how to use effectively all the ancient skeptical tropes. (Feyerabend used to entertain Lakatos by signing his letters and postcards as *Paulus Empiricus.*) Skepticism to him was not only a powerful rhetorical device but also highly regarded for its normative implications for the practice and the role of science in a "free society." Feyerabend's iconoclastic enterprise is against neither reason nor science. It is against the idea that there is some unique set of rules (whatever it is) that ought to be followed in order to produce good science (whatever that is). Feyerabend's favorite slogan, *anything goes,* "is a jocular summary of the predicament of the rationalists" (Feyerabend 1978, 188)—thus "anything goes" from the point of view of the rationalist who believes that only *the* scientific method is admissible. On the contrary, there are lots of ways of moving forward, including the different local and contextual methods of various sciences or traditions. (If anything goes, reason sometimes goes as well; thus Feyerabend is not guilty of any inconsistency by employing rational arguments to attack the rationalist positions he opposes.) In this respect, Feyerabend can be seen not as rejecting rationality *tout court*, but rather as urging a conception of rationality wider than that embodied in some existing version of scientific rationalism (Preston 1997, 203). Feyerabend's arguments are generally to be intended as a *reductio* against certain forms of rationalism, rather than positive arguments in favor of irrationalism (Munévar 2000, 63–64). Far from a self-defeating skepticism, Feyerabend presented an impressive challenge to the received view in the philosophy of science. He argued that its elegant but useless epistemological accounts should be substituted by a detailed study of the primary sources in the history of science:

> *This* is the material to be analysed, and *this* is the material from which philosophical problems should arise. And such problems should not at once be blown up into formalistic tumours which grow incessantly by feeding on their own juices but they should be kept in close contact with the process of science even if this means lots of uncertainty and a low level of precision. (1999a, 137)

In this respect, Feyerabend's legacy can hardly be overestimated.

Matteo Motterlini

References

Feyerabend, P. K. (1962), "Explanation, Reduction, and Empiricism," *Minnesota Studies in the Philosophy of Science*, III. Minneapolis: University of Minnesota Press, 28–97.

——— (1975), *Against Method: Outline of an Anarchistic Theory of Knowledge*. London: New Left Books.

——— (1978), *Science in a Free Society*. London: New Left Books.

——— (1981), *Rationalism and Scientific Method: Philosophical Papers, Vol. I.* Cambridge: Cambridge University Press.

——— (1987), *Farewell to Reason*. London: Verso.

——— (1993), *Against Method* (3rd ed.). London: Verso.

——— (1996), *Killing Time: The Autobiography of Paul Feyerabend.* Chicago: University of Chicago Press.

——— (1999a), *Knowledge, Science and Relativism: Philosophical Papers, Vol. III.* Edited by J. Preston. Cambridge: Cambridge University Press.

——— (1999b), *Conquest of Abundance: A Tale of Abstraction Versus the Richness of Being*. Edited by B. Terpstra. Chicago: University of Chicago Press.

Feyerabend, P. K., and I. Lakatos (1999), *For and Against Method. Including Lakatos's Lectures on Scientific Method, and the Lakatos-Feyerabend Correspondence*. Edited by M. Motterlini. Chicago: University of Chicago Press.

Hacking, I. (2000), "'Screw You, I'm Going Home.' Review of *Conquest of Abundance*," *London Review of Books*, June 2000, 28–29.

Mill, J. S. ([1859] 1977), "On Liberty," in J. M. Robson (ed.), *Collected Works of John Stuart Mill, Vol. 18*. Toronto: Toronto University Press.

Munévar, G. (2000), "A *Réhabilitation* of Paul Feyerabend," in J. M. Preston, G. Munévar, and D. Lamb (eds.), *The Worst Enemy of Science? Essays in Memory of Paul Feyerabend*. New York: Oxford University Press, 58–79.

Preston, J. M. (1997), *Feyerabend: Philosophy, Science and Society*. Cambridge: Polity Press.

See also **Empiricism; Incommensurability; Kuhn, Thomas; Logical Empiricism; Social Constructionism; Theories**

FITNESS

Darwin's theory of evolution by natural selection is often summarized in terms first coined by Herbert Spencer as the claim that among competing organisms the fittest survive. If there is variation among the traits of organisms, and if some variant traits confer advantages on the organisms that bear them, that is, enhance their fitness, then those organisms will have a tendency to live to have more offspring, which in turn will bear the advantageous traits. The success of Darwin's theory turns on the meaning of its central explanatory concept, 'fitness.'

What is fitness and how can one tell when a trait enhances fitness, or more to the point, when one organism is fitter than another? Some opponents of the theory of evolution by natural selection have long claimed that by defining fitness in terms of actual rates of reproduction, the proponents of evolutionary theory are unknowingly condemning the principle of the survival of the fittest to triviality: If one defines fitness in terms of actual reproductive rates, one is making the claim that those organisms with higher rates of reproduction leave more offspring. This is obviously an empty, unfalsifiable tautology bereft of explanatory power.

Evolutionary theory requires a definition of fitness that will protect it from the charges of tautology, triviality, unfalsifiability, and explanatory infirmity. If no such definition is forthcoming, then what is required from the theory's advocates is an alternative account of the theory's structure and content and the theory's role in the research program of biology.

Ensemble Properties and Population Biology

Since the modern synthesis, evolution is usually described in terms of "change in gene ratios" (see Population Genetics). Population genetics is the mathematical formalism used to describe the effects of natural selection on changes in gene frequency. The subject matter of population genetics is populations, ensembles of organisms, or genes, and not individual biological organisms or pairs of them. This has suggested to more than one philosopher that the theory of natural selection is better understood solely as a theory about ensembles and not individuals. As Sterelny and Kitcher (1988, 345) put this view, "evolutionary theory, like statistical mechanics, has no use for such a fine grain of description [as the biography of each organism]: the aim is to make clear the central tendencies in the history of evolving populations."

In this view, though the word "fitness" (or *w*) figures in the theory, it is to be understood as exclusively expressing probabilistic reproduction rates for populations. This is generally operationalized in terms of differential success of genotypes (*genotypic fitness*). It has also been suggested that *allelic fitness* is useful, although allelic fitness will not be predictive of evolutionary change in cases such as that of dominance. Although it has been argued that cases such as heterozygote superiority could be handled at the allelic level as a case of frequency-dependent selection, others disagree (for discussion, see Sober and Lewontin 1982; Sterelny and Kitcher 1988). Whether one uses allelic frequencies, genotypic frequencies, or some other kind of census, evolutionary theory is then treated as a set of claims about how the sizes of populations and subpopulations change over time as a function of differing reproductive rates at some initial time, holding environments constant. According to that account, the theory makes no claims about the local adaptation of individual organisms to their particular environment, or for that matter about the local adaptation of populations to their environment. Thus, it does not provide a local causal explanation of these changes in organism or gene frequency (i.e., the account is silent on what specific features of environment *E* and the organism *x* lead to the changes). Explanations for these changes are to be sought elsewhere. This being said, population-level explanations do offer clear benefits: Some traits may be explainable only by stepping away from explanations that focus on the benefit to the individual organism.

Some population structures in some species (e.g., some colonies of social insects) "encourage" sterility of some of its members to increase the numbers of the overall population. As Sober (1984, 135) puts it: "According to the [Darwinian fitness notion], a sterile organism will have 0 fitness, since its chances of reproducing are nil." If the focus is shifted from the organism to its genes, and if these genes are shared by other organisms, one gets a different view of fitness. With the idea of "inclusive fitness," Hamilton (1964) explained that an organism might have a better chance of having its genes represented at a later generation if it forgoes its own reproductive success to help other bearers of its genes increase theirs. Hamilton suggested that inclusive fitness should be understood as the sum of the individual's fitness and the individual's effect on the fitness of other related individuals. Kinship is used as a tracker to "estimate" the degree to which two organisms "share the same genes." The higher the relatedness, the higher the probability that a significant proportion of genes is shared between the benefactor/donor and the recipient. The net effect in terms of the donor's genes being passed on could be greater by helping relatives reproduce than by reproducing itself. Kin selection provides an explanation for traits such as altruism that prima facie do not seem to fit into evolutionary theory.

However, kin selection has recently been somewhat demoted in favor of more general group-selection explanations (Hamilton himself shifted his views to a group-selection explanation following his reading of Price's [1970] argument on covariance). One of the advantages of group-selection explanations is that they do not presuppose that relatedness is necessary to establish traits such as altruism. More complex interactions are possible, and communities of different species can act as units of selection. However, understanding these interactions and identifying these communities demands a more ecological approach.

Solution to a Design Problem

Suppose, following Dennett (1995) and others before him, one characterizes the relation "*x* is fitter than *y*" as follows:

> *x* is fitter than *y* if and only if *x*'s traits enable it to solve the "design problems" set by the environment more fully than *y*'s traits do.

Call this concept *ecological fitness.* The "ecological" definition is fraught with difficulties.

What are these design problems? How many of them are there? Is there any way of measuring the degree to which *x* exceeds *y* in their solution? To begin with, the notion of "design problems" is vague and metaphorical. If treated literally, design problems will all be relative to the overarching objective of leaving more descendants: If to be fitter means that an organism offers a better solution—better fills an ecological niche—how is this differential success measured any other way than by measuring higher offspring numbers? Thus the definition may simply hide the original problem of distinguishing fitness from reproductive rates, instead of solving it.

Second, the number of design problems is equal to the number of distinct environmental features that affect survival and reproduction, and this number is probably uncountable. It is therefore no wonder that many biologists have favored defining "*x* is fitter than *y*" in terms of quantitatively measurable reproductive rates.

The Propensity Interpretation of Fitness

Among philosophers of biology, there has been a wide consensus that the solution to the problem of defining fitness is to be found in treating it as a probabilistic disposition. The most popular, or the "standard," probabilistic propensity account of fitness has been in terms of offspring contribution (note that there are other propensity accounts that aren't "offspring centric"; more on this below). As such, the propensity causally intervenes between the relationship of environments to organisms that cause it and the actual rates of reproduction, which are its effects. Thus an organism can have a probabilistic disposition to have *n* offspring and yet unluckily never actually reproduce (or produce a number of offspring different than *n*).

Comparative fitness differences are dispositions supervening on the complex of relations between the manifest properties of organisms and environments (Rosenberg, 1978) and will give rise to differential reproductive rates. Thus, definitions such as the following were advanced (Brandon 1978; Beatty and Mills 1979):

> *x* is fitter than *y* in *E* **iff** ["if and only if"] *x* has a probabilistic propensity to leave more offspring in *E* greater than *y*'s probabilistic propensity to leave more offspring in *E*.

If fitness is a probabilistic propensity, then the fitter among competing organisms will not always leave more offspring. Fitness differences do not invariably result in reproductive differences, but only with some probability (since the theory allows for drift, this qualification on its claims will be a welcome one). However, assume, as Hume famously argued, that causes must be distinct from their effects and only contingently related to them. Then if the probability of leaving more offspring is an effect, then it will have to be distinct from its cause—the probabilistic propensities that constitute fitness differences. This is a problem that all probabilistic accounts have to face (not just propensities to leave more offspring). One possible solution is to define probability as long-run relative frequency. The idea here is that chances explain long-run frequencies: If *x* has a probabilistic propensity to leave more offspring than *y* in every generation, then the long-run relative frequency of *x*'s having more offspring than *y* in any generation is greater than *y*'s long-run relative frequency.

But there must be at least a theoretical difference between the chance and the frequency if one wishes the former to explain the latter causally. There are philosophers of science who deny that such a distinction between probabilistic propensities in general (not just biological propensities) and long-run relative frequencies is possible (Earman 1986, 149). Others have argued that, as in the case of quantum mechanics, there are such independent chancy probabilistic fitness propensities that generate the long-run relative frequencies. Proponents of the standard account in biology have envisioned two possible explanations for these propensities. One is that probabilistic propensities at the biological levels of phenomena are the result of quantum probabilities "percolating up" (Sober 1984); the second is that there are brute unexplainable probabilistic propensities at the level of organismal fitness differences (Brandon and Carson. 1996). Few doubt that quantum percolation of some kind could have a biological significance. It may well be a source of mutation (cf. Monod 1971, 111–115 for support of this idea). But the claim that it has a significant role in fitness differences is not supported by any independent evidence (cf. Millstein 2000 for discussion). Even if quantum effects could percolate up, they probably would do it so infrequently that they could not help but ground all biological propensities. The claim that there are brute probabilistic propensities at the level of organismal fitness differences is controversial, depending largely on one's acceptance of emergent autonomous propensities at the macro level.

Some qualifications of the standard propensity account will need to be offered. As Gillespie (1977) has shown, the temporal and/or spatial variance in number of offspring may have an important selective effect. In certain scenarios, it might be more beneficial for an organism to have a lower mean number of offspring with a low variance than a higher mean number of offspring with a wider variance (and with equal means, the organism with the higher variance will always be selected against). To take the example from Brandon (1990, 20): If organism *A* has 2 offspring each year, and organism *B* has 1 offspring in odd-numbered years and 3 in even-numbered ones, then, *ceteris paribus*, after 10 generations there will be 512 descendants of *A* and 243 descendants of *B*. The same holds if *A* and *B* are populations and *B*'s offspring vary between 1 and 3 depending on location instead of period. This is not a problem for propensity accounts that reflect averages over many generations, but it would be a problem for accounts that do not in fact have access to such long-term averages.

Accordingly, the definition needs to be changed to accommodate the effects of variance in offspring number per generation. One could get something like this formulation:

> x is fitter than y iff probably x will have more offspring than y, *unless* their average numbers of offspring are equal and the temporal and/or spatial variance in y's offspring numbers is greater than the variance in x's, or the average numbers of x's offspring are lower than y's but the difference in offspring variance is large enough to counterbalance y's greater number of offspring.

It is also the case that in some biologically actual circumstances (e.g., where mean fitnesses are low), increased variance is sometimes selected for (Ekbohm, Fagerstrom, and Agren 1980). As Beatty and Finsen (1987) further showed, the definition will also have to accommodate skewness along with offspring numbers and variance, on pain of falsity. One simple way to do this is to add a *ceteris paribus* clause to the definition. But the question must then be raised as to how many different exceptions to the original definitions need to be accommodated. If the circumstances under which greater offspring numbers do not make for greater fitness are indefinitely many, then this definition will be unsatisfactory.

Some proponents of the propensity definition recognize these difficulties and are prepared to accept that at most a "schematic" definition can be provided. Thus Brandon (1990, 20) defined the "adaptedness," or expected fitness, of an organism O in an environment E as:

$$A^*(O,E) = \sum P(Q_i^{OE})Q_i^{OE} - f(E,\sigma^2)$$

where Q_i^{OE} represents a range of possible offspring numbers in generation *i*, $P(Q_i^{OE})$ is the probabilistic propensity to leave Q_i^{OE} in generation *i,* and, most importantly, $f(E,\sigma^2)$ is "some function of the variance in offspring numbers for a given type, σ^2 and of the pattern of variation" (Brandon 1990, 20), or, in other words, some function or other that is not known in advance of examining the case. Moreover, one will have to add to variance other factors that determine the function, such as Beatty and Finsen's skewnes. Thus, the final term in the definition will have to be expanded to $f(E,\sigma^2,\ldots)$, where the ellipsis indicates the additional statistical factors that sometimes combine with or cancel the variance to determine fitness levels.

But how many such factors are there, and when do they play a nonzero role in fitness? The number of such factors is probably indefinitely large. The reason for this is given by the facts about natural selection as Darwin and his successors uncovered them. The fact about selection that fates this definition to be forever schematic is the "arms-race" strategic character of evolutionary interactions. Since every strategy for enhancing reproductive fitness (including how many offspring to have in a given environment) calls forth a counterstrategy among competing organisms (which may undercut the initial reproductive strategy), the number of conditions covered by the *ceteris paribus* clause is equal to the number of strategies and counterstrategies of reproduction available in an environment. Brandon (1990) writes, "In the above definition of A^* (O, E), the function $f(E,\sigma^2)$ is a dummy function in the sense that the form can be specified only after the details of the selection scenario have been specified" (20). He acknowledges that the function f will differ for different O and E and will have to be expanded to accommodate an indefinite number of further statistical terms beyond variance. Schematically, it will take the form $f(E,\sigma^2,\ldots)$. Again, adapting Brandon's notation, none of the members of the set that express his generic definition of adaptedness, or expected fitness, $[P(Q_i^{OE})Q_i^{OE} - f_1$ $(E,\sigma^2,\ldots)$. $P(Q_i^{OE})$ $Q_i^{OE} - f_1(E,\sigma^2,\ldots), P(Q_i^{OE})$ $Q_i^{OE} - f_2$ $(E,\sigma^2,\ldots), P(Q_i^{OE})$ $Q_i^{OE} - f_3(E,\sigma^2,\ldots),$ $\ldots]$, is in fact a definition of either term. It is the set of operational measurements of the property of comparative fitness.

It is for reasons such as these that the standard propensity account endorses a schematic view that accommodates all the *ceteris paribus* clauses that could appear (from variance, low mean fitnesses, kin selection effects, etc.). But since the number of these clauses may be indefinite, the standard propensity view does not truly offer a definition of fitness. The schematic nature of this propensity interpretation, along with other problems elaborated in Sober (2002) and Beatty and Finsen (1987), will motivate some to consider alternative approaches to the treatment of fitness.

Conclusion: Models, Ecological Fitness, and the Problem of Evolutionary Drift

Any attempt to turn a generic probabilistic schema for fitness into a complete general definition that is both applicable and adequate to the task of vindicating the truth of the theory of natural selection is problematic. These problems suggest to some philosophers that there is a need to rethink the cognitive status of the theory altogether.

Some (mainly Williams 1970; Rosenberg 1985) have argued that if one strives for an axiomatic formalization of evolutionary theory, one will have to understand fitness as being a primitive notion: A definition of fitness is not available from within the theory itself. Only operational definitions of fitness will be available, which will be only provisional characterizations to help guide investigations and not definitions in the strict sense, since fitness can be characterized only "by appeal to the phenomena that it is employed to account for" (Rosenberg 1985, 141). In other words, fitness can be characterized only through its actual causal role (analogously to the concepts of force and charge). Among other reasons, the very abstract nature of this axiomatic account made it unpalatable for most philosophers and biologists: Biologists do not use a "primitive" notion of fitness in their inquiry, but a more empirical notion better reflected in semantic accounts.

Trying to address these pragmatic concerns, others have argued that the theory of natural selection should not be viewed as a body of general laws but as the prescription for a research program (see Brandon 1990, chap. 4). As such, its central claims need not meet standards of testability, and fitness need not be defined in terms that assure the nontriviality, testability, and direct explanatory power of the theory of natural selection. Evolutionary theory remains a scientifically respectable, but nevertheless untestable, organizing principle for biological science.

Thus, in each particular selective scenario, a different specification of the schematic propensity definition figures in the antecedent of a different and highly restricted principle of natural selection that is applicable only in that scenario. The notion that there is a very large family of principles of natural selection, each with a restricted range of application, may be attractive to those biologists uncomfortable with a single principle or law of natural selection, and to those philosophers of science who treat the theory of natural selection as a class of models. In the "semantic" approach to the theory of natural selection (see Theories), each of the substitution instances of the schematic principle of natural selection generated by a particular specification of the propensity definition of fitness is treated as a definition of a different Darwinian system of population change over time (Beatty 1981). The evolutionary biologist's task is to identify which definition is instantiated by various populations in various environments.

The difficulties of the probabilistic propensity definitions of fitness (standard accounts and others) are serious enough to make the notion of ecological fitness worth revisiting. Recall that in this view "A is fitter than *B* in *E* " is defined as "*A*'s traits result in its solving the design problems set by *E* more fully than do *B*'s traits." The terms in which this definition is couched are certainly in as much need of clarification as is "fitness." There does appear to be important biological work that the ecological fitness concept can do that a definition of fitness solely in terms of differential reproductive rates (actual, expected, or dispositional) cannot.

Suppose one measures the fitness differences between population *A* and population *B* to be in the ratio of 7:3 (e.g., $w_A = 1$, $w_B = 0.428$), and suppose further that in some generation the actual offspring ratio is 5:5. There are two alternatives: (1) The fitness measure of 7:3 is right but drift explains the deviation or (2) the fitness measure of 7:3 is incorrect and drift has occurred.

In the absence of information about the initial conditions of the divergence, there is a way empirically to choose between (1) and (2), which requires that there be ecological fitness differences that are detectable. Suppose that fitness differences were matters of probabilistic differential reproductive success. Then the only access to fitness differences would be via population censuses in previous generations (since these form the bases of the probabilities). Suppose that this census did indeed show a 7:3 ratio between *A* and *B* in the recent past. In order to exclude the absence of fitness differences instead of drift as the source of the current generation's 5:5 outcome, one needs to be able to establish that the 7:3 differences in previous populations were not themselves solely the result of drift. But this is the first step in a regression, since the original problem was in discriminating drift from mismeasures of fitness. Of course, the problem does not arise if one has access to fitness differences independently of previous population censuses. And this access is available, at least in principle, if fitness is a matter of differences in the solution of identifiable design problems, that is, if there is such a thing as ecological fitness and it is (fallibly) measured by probabilistic propensities to leave offspring.

If there is access to ecological fitness differences, one can, at least in principle, decide whether the divergence from predicted long-run relative frequencies is a matter of drift or reflects an ignorance either of ecological fitness differences or the unrepresentativeness of the initial conditions of individual births, deaths, and reproductions. But given the epistemic problems related to ecological fitness highlighted earlier, can one truly say that ecological fitness differences are within reach?

The problem of defining fitness remains. Or at any rate, it does if biology cannot live with an imperfect definition of fitness in terms of an overall design-problem solution or an understanding of fitness as a research program for building models.

Frédéric Bouchard

The author acknowledges the helpful input of Alex Rosenberg, Duke University.

References

Beatty, J. (1981), "What's Wrong with the Received View of Evolutionary Theory?" in P. Asquith and R. Giere (eds.), *PSA 1980*, vol. 2. East Lansing, MI: Philosophy of Science Association, 397–426.

Beatty, J., and S. Finsen (1987), "Rethinking the Propensity Interpretation," in M. Ruse (ed.), *What Philosophy of Biology Is*. Dordrecht, Netherlands: Kluwer.

Beatty, J., and S. Mills (1979), "The Propensity Interpretation of Fitness," *Philosophy of Science* 46: 263–288.

Brandon, R. (1990), *Adaptation and Environment*. Princeton, NJ: Princeton University Press.

——— (1978), "Adaptation and Evolutionary Theory," *Studies in the History and Philosophy of Science* 9: 181–206.

Brandon, R., and S. Carson (1996), "The Indeterministic Character of Evolutionary Theory," *Philosophy of Science* 63: 315–337.

Dennett, D. C. (1995), *Darwin's Dangerous Idea*. New York: Simon and Schuster.

Earman, J. (1986), *A Primer on Determinism*. Dordrecht: Reidel.

Ekbohm, G., T. Fagerstrom, and G. Agren (1980), "Natural Selection for Variation in Offspring Numbers: Comments on a Paper by J. H. Gillespie," *American Naturalist* 115 : 445–447.

Gillespie, G. H. (1977), "Natural Selection for Variances in Offspring Numbers: A New Evolutionary Principle," *American Naturalist* 111: 1010–1014.

Hamilton, W. D. (1964), "The Genetic Evolution of Social Behaviour," *Journal of Theoretical Biology* 7: 1–16.

Millstein, R. L. (2000), "Is the Evolutionary Process Deterministic or Indeterministic? An Argument for Agnosticism," paper presented at the biennial meeting of the Philosophy of Science Association, Vancouver, Canada, November.

Monod, J. (1971), *Chance and Necessity*. New York: Alfred A. Knopf.

Price, G. R. (1970), "Selection and Covariance," *Nature* 227: 520–521.

Rosenberg, A. (1978), "The Supervenience of Biological Concepts," *Philosophy of Science* 45: 368–386.

——— (1985), *The Structure of Biological Science*. Cambridge: Cambridge University Press.

Sober, E. (1984), *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, MA: MIT Press.

——— (2002), "The Two Faces of Fitness," in *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*. Cambridge: Cambridge University Press.

Sober, E., and R. Lewontin (1982), "Artefact, Cause, and Genic Selection," *Philosophy of Science* 47: 157–180.

Sterelny, K., and P. Kitcher (1988), "Return of the Gene," *Journal of Philosophy* 85: 339–362.

Williams, M. B. (1970), "Deducing the Consequences of Evolution: A Mathematical Model," *Journal of Theoretical Biology* 29: 343–385.

See also **Evolution; Natural Selection; Population Genetics**

FUNCTION

Current philosophical debate concerning functions is focused around the concern that commitment to the existence of functions in nature rests uneasily on the assumption that scientific descriptions of the world should be wholly *naturalistic,* meaning that they must not involve any irreducible, mentalistic elements. Talk of 'function' is pervasive in many areas of the life sciences, and a cursory glance at how it is employed suggests that functions *are* generally taken to be real properties of the biological world. However, the concept of function carries with it all sorts of teleological and mentalistic connotations, which are seemingly in tension with an objective, scientific view of the world. For example, there is the possibility that ascription of biological function involves reference to *design* and *agent intention.* Biological functions, like artifact functions, incorporate an effect of a structure into a description of its causal capacities, thus introducing a "forward-looking" element. Thus, saying that "the function of the heart is to pump the blood" might be taken to imply that "the heart beats *so as* to pump the blood" or "the heart is designed for blood pumping." The effects that a designer *intended* an *artifact* to *achieve* clearly do play a role in determining its current causal properties. Hence, appealing to an intended effect in order to account for current properties is unproblematic. However, a conscious designer is not the sort of thing that a naturalistic biology wants to fall back on. As Allen and Bekoff (1995, 611) put it, "Successful naturalization of teleological notions in biology requires that one give an account of these notions that does not involve the goals or purposes of a psychological agent."

A further problem that a naturalistic analysis of biological function must contend with is that many ascriptions of function are *normative,* meaning that the function of a structure is not just something that it *does,* but also something that it is *supposed to do.* A function of *X* does not *accidentally* contribute to the realization of some effect but *properly* does so. Similarly, when something fails to fulfil its function, it does not just *do something else* but rather *goes wrong* or *malfunctions.* Assuming naturalism, a tenable account of functions is required to assimilate them into a purely descriptive framework. Hence, if a concept of normative or proper function is to survive naturalization, the normative must somehow be reduced to the descriptive.

Rather than try to accommodate such concerns, why not just dispense altogether with reference to functions in biology? Why not condemn them to the fate of 'vital spirits' and 'pangenic gemmules'? Given the centrality of functional characterizations in the life sciences, this would call for dramatic revisions of contemporary scientific practices. Functions are not only invoked in describing and explaining the workings of biological structures but also play a role in the classification and individuation of many such structures (Neander 1991). The function of a wing is what makes it a wing. Similarly, a heart would not be a heart if it were not for its function; an eye would not be an eye and a leg would not be a leg. Of course, classification according to function is not exclusive. As Amundson and Lauder (1994, 453) note, biological organs can be classified both anatomically and functionally. Nevertheless, bereft of function, classifications of the biological world would be very different and, most would agree, severely impoverished. Hence, the central problem addressed by most philosophical discussions is how to analyze function in a way that keeps biological science natural and at the same time licenses continued employment of teleological language, whose elimination would make biology very difficult indeed.

Naturalizing Function

The parameters of current debates owe much to the classic accounts of Nagel (1961) and Hempel (1965). They both structure their discussions around the need to account for scientific usage of 'function' in nonteleological terms, accepting (a) that irreducible teleologies are scientifically unacceptable and (b) that some appeals to function in the life sciences are legitimate. Nagel takes it as given that functional language in the life sciences does not ultimately appeal to irreducible teleologies:

> We shall ... assume that teleological (or functional) statements in biology normally neither assert nor presuppose in the materials under discussion either manifest or latent purposes, aims, objectives, or goals. . . . [D]espite the prima facie distinctive character of teleological (or functional) explanations, we shall first argue that they can be reformulated, without loss of asserted content, to take the form of nonteleological ones, so that in an important sense teleological and nontelological explanations are equivalent. (402–403)

Hempel (1965) takes a similar line, conceding a historical connection between biological functions and full-blooded teleology but maintaining at the same time that any teleological explanations involving irreducible purposes or entelechies are "pseudo-explanations" (304). However, like Nagel, he assumes that certain uses of 'function' in science are legitimate, having a "definitely empirical core" (ibid). Hence, the object of both analyses is to account for function in nonteleological terms, defending the assumption of a clear distinction between function and teleology through a naturalistic analysis of the latter.

Nagel's analysis proposes that functions are contributions made by parts of a system to that system's ability to carry out characteristic processes and behaviors. As Nagel (1961) puts it, structure *A* has a function in cases where "[e]very system *S* with organization *C* and in environment *E* engages in process *P*; if *S* with organization *C* and in environment *E* does not have *A,* then *S* does not engage in *P*; hence, *S* with organization *C* must have *A*" (403). He observes that although *A* is never *logically* necessary for process *P*, structures such as brains and livers can be said to be "necessary" in a different sense, in that there are no *actual* alternatives in the biological world (404). Thus an account of the function of a systemic constituent can contribute to a robust *explanation* of a system's ability to sustain a certain behavior or process.

Functions, for Nagel (1961), are essentially causal contributions to systemic goals. A successful naturalization of 'function' therefore requires that 'goal' also be naturalized. Nagel analyzes goals in terms of the properties of *directively organized systems*, invoking criteria such as adaptability and plasticity (417–418), but concedes that any demarcation between goal-directed and non-goal-directed systems will be somewhat vague (419). Given a naturalistic account of functions and goals along the lines set out, Nagel suggests that any remaining differences between teleological and nonteleological explanations in science should be attributed to "emphasis and perspective in formulation" rather than content (422).

Hempel's (1965) account is similar to that of Nagel, in maintaining that functions are roles played by parts of a system that contribute to "keeping the given system in proper working order or maintaining it as a going concern" (305). Hempel locates proper scientific usage of 'function' within a procedure called functional analysis, whose aim is to "exhibit the contribution which the behavior pattern makes to the preservation or the development of the individual or group within which it occurs" (305). He constrains the domain of legitimate function assignment to those *constituents of self-regulating systems that contribute to the activity of self-regulation*, thus ruling out numerous possible systemic goals that would be admitted by Nagel's more permissive account. Hempel is less optimistic than Nagel with regard to the explanatory potential of functional analysis, which, he observes, cannot deductively explain the presence of a functional item and has very little predictive power (312–313). However, he does assign it a heuristic role, in "determining the respects and the degrees in which various systems are self-regulating" (330). Hempel also departs from Nagel in adding a historical dimension to his analysis. Functions are not simply *current* contributions made by parts of a system to that system's capacities. The objects of functional analysis are also "standardized" or "repetitive" items (307) that occur in a context of sustained self-regulation, implying a historical or backward-looking aspect to the concept of function (see McLaughlin 2001, part II).

This historical dimension is developed substantially and in a novel way by Wright's (1973) analysis, from which the majority of recent attempts to naturalize biological functions take their lead. Wright seeks to provide a unified account, embracing both natural and artifact functions. The two key requirements of such an account are, according to Wright, that it be able to distinguish a function of X from an accidental consequence of X and be able to capture the sense in which statements of function are intrinsically explanatory, a statement of X's function constituting an explanation of why X exists. Wright proposes a consequence etiology, which is intended to specify necessary and sufficient conditions for an entity X to have a function Z. The 'function of X is Z' means:

- X is there because it does Z.
- Z is a consequence (or result) of X's being there. (161)

The analysis is devised to accommodate both natural and artifact functions, both conscious design and natural selection. A car (X) has the function of transportation (Z) because transportation is not merely one of its effects but also what it was *designed for* and hence why it exists. A kidney (X) has the function of cleansing the blood (Z), as it was *selected for* that role and hence exists *because* of its role in cleansing the blood. Wright's analysis attempts to preserve the forward-looking element of function (by appealing to history in order to show how the effect can explain the presence of cause), the association between function and origin (the function of X is also the reason for X's existence), and the normativity of function (X is supposed to do what X was designed or selected to do) whilst also providing a naturalistic, unitary account, covering all uses of 'function.'

There are a number of possible exceptions to Wright's analysis, many of which were first pointed out by Boorse (1976). For instance, consider a break in a hose (X), which releases chlorine gas (Z). The break causes the gas to be released, but release of the gas keeps the break open, by gassing anybody who gets close enough to seal it, and so the break continues to exist because it releases the gas (72). This maps onto both of Wright's criteria for Z being a function of X. However, it is intuitively clear that this is not the kind of scenario to which one would want to assign a function.

More recent etiological accounts take Wright's basic formulation as their starting point and attempt to insulate it against various counterexamples in order to capture a sense of function as the notion is employed specifically in biology. This quest has resulted in several deviations from the goals of Wright's original analysis. For example, Millikan (1989) and Neander (1991) both depart from the project of conceptual analysis to formulate accounts that are, to varying degrees, stipulative and not intended to reflect every instance of everyday function assignment. Part of the motivation for this is the need to guard against far-fetched counterexamples such as car engines and livers materializing out of nowhere, or more plausible-sounding scenarios such as complex biological structures emerging through chance macromutations and thus owing nothing to the action of natural selection. Even if commonsense intuitions do lean toward the assignment of function in these cases, such intuitions can simply be declared irrelevant if conceptual analysis is not one of its aims.

Coupled with the departure from conceptual analysis is a renunciation of the goal of a unitary account. Recent etiological approaches tend to focus specifically on *biological* function, which they analyze in terms of natural selection and its products, rather than on the basis of some more general etiology common to both conscious design and natural selection. Hence, they treat biological functions as essentially distinct from artifact functions, given that artifact functions are not the result of a process of blind, nonconscious selection:

> It is the/a proper function of an item (*X*) of an organism (*O*) to do that which items of *X*'s type did to contribute to the inclusive fitness of *O*'s ancestors, and which caused the genotype, of which *X* is the phenotypic expression, to be selected by natural selection. (Neander 1991, 176)

Even given this restriction in scope to biological function and natural selection, further questions have been raised as to whether short-term or long-term selection history should be given priority when assigning functions. For example, a structure might originally have been selected for task *A* and have subsequently been selected for task *B*, with *A* fading out of the picture completely. Most accounts maintain that recent selection pressures are more relevant in determining current function (Godfrey-Smith 1994). Buller (1998) adds yet further fine-tuning to etiological approaches by differentiating between "strong" and "weak" etiological theories. These are distinct in that the former explicitly emphasizes selection *for* a trait, whereas the latter stresses only that the trait be a *result* of the reproduction of structures that had the same effect.

Despite the often subtle differences between contemporary etiological theories of function, all aim to provide an analysis that preserves a concept of normative, teleological, and (most importantly) *natural* function that can be legitimately employed to describe the biological world.

Etiological accounts are not the whole story, however, and causal role accounts, which, like the early analyses of Nagel (1961) and Hempel (1965), emphasize the current contribution of a structure to systemic capacities, are still viewed by many as a plausible alternative or supplement to the etiological view. A causal role account proposed by Cummins ([1975] 1984) has been particularly influential. He argues that functions are contributions to the capacities of containing systems, playing a pivotal role in an explanatory strategy that he, like Hempel, refers to as "functional analysis":

> When a capacity of a containing system is appropriately explained by analyzing it into a number of other capacities whose programmed exercise yields a manifestation of the analyzed capacity, the analyzing capacities emerge as functions. (407)

A function of *X* is a capacity of *X* that contributes to the explanation of a more complex capacity of a containing system *Y*. So functions, for Cummins, play a role in a reductive explanatory strategy that is popular in various areas of the biological sciences, requires an essentially ahistorical understanding of causal function, and, as Amundson and Lauder (1994) point out, does not require the incorporation of normativity.

Cummins' account has been criticized for being both too vague and too liberal. It is unclear precisely what is meant by "system" and, without any explicit restrictions concerning acceptable systems, just about anything can have a Cummins function in some context. In relation to some conceivable system (in the Cummins sense), the function of cancer is to cause death and the function of the nose to support spectacles (see e.g., Manning 1997, 70). Hence, the account needs to be constrained in order to isolate a more specific sense of function that can be applied to biological practice without opening the floodgates to permit all manner of bizarre functions. Certain more recent causal role accounts insulate against excessive generality by restricting function assignment to those roles that enhance the *fitness* of biological systems (Bigelow and Pargetter 1987; Mitchell 1995; Walsh 1996). In contrast to etiological approaches, these theories construe function in terms of *current* rather than *historical* propensity to contribute to biological fitness. Bigelow and Pargetter (1987) propose their propensity account as an *alternative* to etiological accounts. However, the two accounts need not be antagonists and can be thought of as working together to accommodate distinct but equally permissible uses of the notion of function. In recent years a consensus has emerged along these lines, which recognizes the legitimacy of etiological functions and Cummins functions in biology, with fitness functions composing a subfamily of Cummins functions. Etiological approaches pin down a distinctive, specific use of the term "function," whilst Cummins' account identifies a broader, more generally applicable use (Godfrey-Smith, 1993). There are still some issues left concerning the nature and extent of any relationships between Cummins functions and etiological functions. Buller (1998, 515) suggests that one has to identify a Cummins function before going on to ask whether that function is also an etiological function. That is, one has to know whether and how *X* contributes to *Z* (Cummins function) before one can determine whether *X* was selected for because it contributes to *Z* (etiological function). Griffiths

(1993) also regards the two senses as intimately connected:

> We can incorporate the etiological approach into the Cumminsesque picture of function ascription. The proper functions of a biological trait are the functions it is ascribed in a functional analysis of the capacity to survive and reproduce (fitness) which has been displayed by animals with that feature. This means that a feature will have a proper function only if it is an adaptation for that function. The trait must have been selected because it performs that function. (412)

However, Godfrey-Smith (1993) places more of an emphasis on disunity and advises philosophers not to "join what science has put asunder" (207), stressing that the two kinds of function are distinct and symptomatic of different patterns of scientific explanation.

In summary, despite a series of minor disagreements amongst its proponents, there is a currently popular consensus in philosophy to the effect that biological functions can be satisfactorily naturalized via some combination of etiological and causal role accounts, with only the former yielding a formulation of normative, teleological, or proper function. The philosophical interest of a naturalized proper function is not restricted to biology, however. The concept also has considerable potential for application in psychology and the philosophy of mind.

Function and Mind

Etiological accounts purport to provide a naturalistic reduction of normative, teleological function, which accounts for normative properties in purely descriptive terms. The resultant formulation of a proper function has proved to be a very powerful tool, whose application is becoming increasingly central to a number of different projects sharing the common goal of naturalizing the mind. For example, one possibility is in accounting for the seemingly intractable normativity of mental states in terms of the more tangible normativity of biological function. This is the aim of currently popular teleological theories of mental representation, such as that of Millikan (1993). The connection between intentional states, such as beliefs and desires, and their objects cannot be adequately characterized in terms of causation alone. The belief that there is a bird rustling in the bushes might well be caused by the foragings of a rat, a brief glimpse of which gives the mistaken impression of a bird. So what is it about the belief that specifies the content 'bird' rather than 'rat and/or bird'? This is the so-called disjunction problem: determining how the content of a belief is fixed as *X* rather than as *X* and/or *Y*, given that it can be present in both cases. The belief about the rustling in the bush is so not just because it is *commonly* caused by birds rather than rats, bees, or buffalo, but because it is *properly* caused only by birds and thus *mistaken* when it is caused by rats. Teleological theories of representation attempt to account for this normativity of intentional states in terms of the normativity of biological function. In its simplest form, their central claim is that a belief *X* is about *Y* rather than *Z* because its proper function is to represent *Y* and not *Z:*

> Just as the characteristic mark of intentionality is that intentional items can be false, unsatisfied, or seemingly 'about' what does not exist, so the characteristic mark of the purposive, of that which has a function, is that it may not in fact fulfill that purpose or serve that function. (Millikan 1993, 23)

Functional characterizations are also employed more generally in the philosophy of mind. Many philosophers hold that *all* mental processes are best characterized in terms of their *biological roles*, as opposed to a physical description of the structures that instantiate those roles. This general position, known as *teleological functionalism* (see e.g., Sober 1985), maintains that proper functions provide a means of describing what various mental processes essentially *do*, without getting sidetracked by accidental/contingent effects of those processes or excessive attention to the anatomy of the biological structures that perform them. For example, an account of the proper function(s) of consciousness will aim to tell us what consciousness *does* and why it came to *be.*

If one goes so far as to adopt the line that mental processes simply *are* what they *do*, then a comprehensive characterization of psychological states and processes in terms of their various etiological functions will amount to a comprehensive account of what the mind *is.* Some recent work in evolutionary psychology has precisely this goal. Evolutionary psychologists such as Cosmides and Tooby (1992) not only employ 'function' to *describe* psychological processes but also maintain that psychological processes are *individuated* or *defined* by their etiological functions. Cosmides and Tooby's aim is to account for the mind in terms of a set of *modules*, which are innate, domain-specific programs selected to deal with the many problems posed by the environment in which humans evolved. Modules are not *first* identified and *then* assigned a function. Instead, their function is *constitutive* of what they are. For

Cosmides and Tooby, etiological functions are utterly indispensable for a scientific understanding of the mind.

Thus, etiological functions play a major part in several projects in naturalistic philosophy of mind and psychology. These projects are structured around the assumption that if we can rid biological teleology of mind, we can then employ it to rid the rest of the world of mind, by reducing various psychological phenomena to their biological functions and thus naturalizing them. So etiological proper functions are increasingly indispensable, not so much for biology but for projects in the philosophy of mind that require a naturalistic formulation of normative, teleological function.

Function in Mind?

Employment of functions in the service of naturalizing the mind presupposes the possibility of objective, mind-independent functions. However, not all agree with Godfrey-Smith's (1993) naturalistic consensus. In his response to Wright (1973), Boorse (1976) argues that all functions are ultimately *contributions to goals* and that only the inclusion of a goal can ultimately serve to distinguish between cases where the term "function" is and is not legitimate. This sort of claim has also been employed to criticize more recent formulations of the etiological theory. For instance, Manning (1997) examines several cases that fit the etiologists' criteria but clearly do not have functions. Both "junk DNA" and "selfish DNA" (which enhances the chances of its own replication but has a detrimental effect on the organism) have the right etiology but neither is assigned a function (74–75). Manning suggests that functions are assigned only when goals are incorporated into one's description of a system or process. Thus, if naturalism requires the elimination of goal directedness, etiological accounts cannot meet the requirements of naturalism without an additional account of goals (80–81).

If functions depend upon goals or other related mentalistic notions, where do these goals come from? Neither Nagel (1961) nor Hempel (1965), both of whom claim that function assignment involves reference to systemic goals, think that the problem of goals is insurmountable for naturalism. However, an alternative response is to concede that goals come from us; we tacitly incorporate them into our conceptions of the biological world, slipping in values, ends, and intentions that have their ultimate source in human agency. Functions, if they do indeed presuppose psychological goals, turn out to be mind-dependent and so cannot be employed to naturalize the mind.

Some recent versions of this kind of view maintain that 'biological function' originates from a metaphor or analogy with human agency and artifice, which results in one's thinking of the biological world in terms of end directedness, normativity, and value. For instance, Matthen (1997) argues that "function attributions seem to be dependent on user, role, mode, use, and the utility that the user realizes from the outcome." Functions are "attributed to natural things by virtue of an analogy with instruments designed for use by or actually used by an agent for a purpose" (31).

Accepting that functions are ultimately dependent on an analogy with human artifice, one might wonder whether they can and should be eliminated from biology. Such a move would be highly contentious, given that function talk is widespread, entrenched, and apparently central to many areas of biological thinking. Matthen (1997, 37) suggests that elimination of function from biology is possible *in principle*, but he does not recommend such a course of action. Ruse (2000) argues that elimination would be undesirable, as even though functions do not correspond to properties of the mind-independent world, they still serve a central role in biological explanations. Functions need not be *out there* in order to be legitimate and useful tools to enhance "science's heuristic power" and "its predictive fertility" (231). Ratcliffe (2001, 45–47) argues that the role played by teleological language in biology renders it impossible to eliminate without eliminating most of biological practice along with it. Functions are not just dispensable metaphors that attach to biological descriptions, but rather constitutive conditions for the possibility of biology, whose absence would render much of that science impossible.

Others claim that certain understandings of biological function could and even *should* be eliminated. Davies (2001) argues that biological proper function cannot be naturalized and ought therefore to be abandoned, leaving only nonnormative causal role functions. For different reasons, Amundson (2000) argues that a "normal" function, construed as a normative ideal of biological performance, should be discarded. He claims that Darwinian evolutionary processes result in a plethora of differences rather than a few "proper" ways of doing things, deviations from which constitute *mal*function or *ab*normality. In place of the contrast between proper function and malfunction, there are a multitude of effective modes of performance, all of which get the job done, if a little differently from each other (34). Amundson (2000) goes on to suggest that the concept of normative biological function is symptomatic of an ideology or prejudice that serves to disadvantage certain people, who are classified as *dis*eased, *dis*abled or *ab*normal and excluded to varying degrees from society on the grounds that society is not obliged to accept or enable those who are biologically *dys*functional: "The disadvantages of people who are assessed as 'abnormal' derive not from biology, but from implicit social judgments about the acceptability of certain kinds of biological variation" (33). So, according to Amundson, so-called normal function is neither part of the biological world nor useful, but an ideological distortion of biology that serves to justify social attitudes that would otherwise appear morally unacceptable, a contention that adds an important social dimension to the functions debate.

Conclusion

To summarize, current debates concerning biological function center around three issues: (a) the number of distinct senses of 'function' at work inside and outside biology, (b) whether or not certain kinds of function are reducible to properties of the objective, mind-independent world, and (c) the extent to which talk of functions is useful or even indispensable in biology.

The prognosis for biological naturalism may well hinge on the outcome of these debates. Crucial to the success of attempts to naturalize the mind through biology is the possibility of a naturalistic reduction of 'normative function,' which etiological accounts claim to supply. But if one concedes, contrary to these accounts, that normative functions are in some sense irreducibly mind dependent, resting on conceptions of goals, values, design, or agency that are not part of the objective, biological world, then employment of function as an ontological component in naturalistic accounts of mind is ruled out. Employing a mind-dependent concept to rid the world of mind would constitute a viciously circular endeavor.

However, all of this is contingent on what commitments one takes to be essentially constitutive of naturalism. For instance, Bedau (1991) argues that although it is not possible to eliminate *value* from the concept of function, this does not imply that values are woven into the biological world by minds or are dependent on an analogy with values that play an integral role in human artifice. Bedau (1991) advocates a "broader naturalism" that acknowledges that "value notions apply to living things, even those which are not human" (655). If naturalism is compatible with values and goals being an intrinsic part of the natural world, then irreducible teleology in nature will threaten neither a naturalistic biology nor naturalistic accounts of psychological processes that rest upon that biology. However, as Manning (1997, 80–81) notes, any such account will "not amount to the full-scale naturalization of functional properties *if* such a naturalization requires the absolute elimination of the analysis and of all teleological, purposive or 'goal-directed' notions." And whatever its merits, "full-scale" naturalization or reduction of teleological properties is indeed the central aim of most so-called naturalistic analyses of function.

Functions are a philosophical problem precisely because of the specter of their incompatibility with a naturalism that refuses to admit teleology or normativity as ground-floor properties of the world and grants them reality only on the condition that they be reducible to more basic properties. Reductive naturalism not only seeks to accommodate function but has increasingly come to depend upon it, as an essential tool in the project of naturalizing other troublesome concepts, such as intentionality or representation. Hence, an awful lot hinges on whether normative, teleological functions can ultimately be cashed out in descriptive, nonteleological terms.

Matthew Ratcliffe

References

Allen, C., and M. Bekoff (1995), "Biological Function, Adaptation and Natural Design," *Philosophy of Science* 62: 609–622.

Amundson, R. (2000), "Against Normal Function," *Studies in History and Philosophy of Biological and Biomedical Sciences* 31: 33–53.

Amundson, R., and G. Lauder (1994), "Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology," *Biology and Philosophy* 9: 443–469.

Bedau, M. (1991), "Can Biological Teleology Be Naturalized?" *Journal of Philosophy* 88: 647–655.

Bigelow, J., and R. Pargetter (1987), "Functions," *Journal of Philosophy* 84: 181–196.

Boorse, C. (1976), "Wright on Functions," *Philosophical Review* 85: 70–86.

Buller, D. J. (1998), "Etiological Theories of Function: A Geographical Survey," *Biology and Philosophy* 13: 505–527.

Cosmides, L., and J. Tooby (1992), "The Psychological Foundations of Culture," in H. Barkow, L. Cosmides, and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press.

Cummins, R. ([1975] 1984), "Functional Analysis," in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 386–407. Originally published in *Journal of Philosophy* 72: 741–765.

Davies, P. S. (2001), *Norms of Nature: Naturalism and the Nature of Functions*. Cambridge, MA: MIT Press.

Godfrey-Smith, P. (1993), "Functions: Consensus with;out Unity," *Pacific Philosophical Quarterly* 74: 196–208.

——— (1994), "A Modern History Theory of Functions," *Noûs* 28: 344–362.

Griffiths, P. (1993), "Functional Analysis and Proper Functions," *British Journal for the Philosophy of Science* 44: 409–422.

Hempel, C. (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Manning, R. (1997), "Biological Function, Selection and Reduction," *British Journal for the Philosophy of Science* 48: 69–82.

Matthen, M. (1997), "Teleology and the Product Analogy," *Australasian Journal of Philosophy* 75: 21–37.

McLaughlin, P (2001), *What Functions Explain: Functional Explanation and Self-Reproducing Systems*. Cambridge: Cambridge University Press.

Millikan, R. (1989), "In Defense of Proper Functions," *Philosophy of Science* 56: 288–302.

——— (1993), *White Queen Psychology and Other Essays For Alice*. Cambridge, MA: MIT Press.

Mitchell, S. D. (1995), "Function, Fitness and Disposition," *Biology and Philosophy* 10: 39–54.

Nagel, E. (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation*. London: Routledge.

Neander, K. (1991), "Functions as Selected Effects: The Conceptual Analyst's Defense," *Philosophy of Science* 58: 168–184.

Ratcliffe, M. (2001), "A Kantian Stance on the Intentional Stance," *Biology and Philosophy* 16: 29–52.

Ruse, M. (2000), "Teleology: Yesterday, Today and Tomorrow?" *Studies in History and Philosophy of Biological and Biomedical Sciences* 31: 213–232.

Sober, E. (1985), "Panglossian Functionalism and the Philosophy of Mind," *Synthese* 64: 165–193.

Walsh, D. M. (1996), "Fitness and Function," *British Journal for the Philosophy of Science* 47: 553–574.

Wright, L. (1973), "Functions," *Philosophical Review* 82: 139–168.

See also **Evolutionary Psychology; Natural Selection; Naturalism; Teleology**

G

GAME THEORY

Game theory is the branch of decision theory that analyzes interdependent decision problems between rational, strategic agents. A rational agent is one who has a consistent set of preferences defined over some set of possible outcomes and who makes choices consistent with these preferences. A strategic agent is one who, given these preferences, reasons about the best course of action to take in order to satisfy them. Interdependent decision problems arise when the outcome for any particular agent depends upon the actions chosen by all of the agents; that is, when the optimal choice for an agent A depends upon the choices made by other agents, and the optimal choice for the other agents depends in turn upon the choice made by A. It is this strategic feature that distinguishes game-theoretic problems from simpler decision problems such as parametric choice under conditions of risk or uncertainty.

The birth of modern game theory is usually attributed to von Neumann and Morgenstern (1944). However, precursors to game-theoretic analyses of strategic problems can be found in Zermelo (1913), Borel ([1921] 1953), and von Neumann ([1928] 1959), as well as in the works of Hobbes and Hume.

A Theory of Utility

One of von Neumann and Morgenstern's primary contributions was their development of a mathematical theory of utility, allowing one to define, for a given agent, an interval utility measure unique up to a strictly increasing affine transformation. The need for such a notion of utility originates in the fact that in game theory, agents often need to make decisions under conditions of risk or uncertainty, and hence one needs a measure of how strong their preferences for a given outcome are.

If an agent's preferences over outcomes satisfy certain basic coherence criteria, it is possible to define a utility function with the property that if one makes choices consistent with one's preferences, one acts *as if* one were choosing to maximize expected utility. The following axioms (from Luce and Raiffa's [1957] classic text *Games and Decisions*) formalize the coherence criteria necessary to satisfy in order to define a von Neumann–Morgenstern utility function. Let $A = \{a_1, \ldots, a_n\}$ denote the set of outcomes, and let $a_j \precsim a_i$ denote that the agent either prefers a_i over a_j or is indifferent between them. A *lottery* $L = (p_1 a_1, \ldots, p_n a_n)$ is simply a randomization over outcomes, where the outcome a_i occurs with probability p_i. A *compound lottery* $Q = (q_1 L_1, \ldots, q_m L_m)$ is a lottery over lotteries, where the chance that the lottery L_i occurs is q_i.

Ordering of Alternatives

For any outcomes a_i, a_j, and a_k either $a_i \precsim a_j$ or $a_j \precsim a_i$ (and possibly both). Moreover, the relation "$\precsim$" is *transitive*; that is, if $a_i \precsim a_j$ and $a_j \precsim a_k$, then $a_i \precsim a_k$.

Reduction of Compound Lotteries

Let $L^i = (p^i_1 a_1, p^i_2 a_2, \ldots, p^i_n a_n)$ be a lottery, for $i = 1, \ldots, m$. Then the agent is indifferent between the compound lottery $(q_1 L^1, q_2 L^2, \ldots, q_m L^m)$ and the simple lottery $(p_1 a_1, \ldots, p_n a_n)$, where $p_i = q_1 p^1_i + q_2 p^2_i + \ldots + q_m p^m_i$.

Continuity

Suppose that $a_n \precsim a_{n-1} \precsim \ldots \precsim a_1$. Then there exists a number u_i such that the agent is *indifferent* between a_i and the lottery $[u_i a_1, 0 \bullet a_2, \ldots, 0. \bullet a_{\mathrm{n}-1}, (1 - u_i) a_n]$, which is denoted $\hat{a}_i$.

Substitutibility

In any lottery, $\hat{a}_i$ is substitutable for a_i.

Transitivity of Lotteries

The preference and indifference relations over lotteries are transitive relations.

Monotonicity

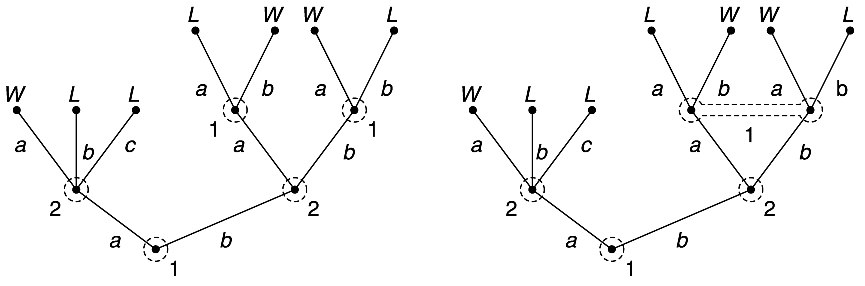
A lottery $(p a_1, (1 - p) a_n)$ is preferred or indifferent to $(p' a_1, (1 - p') a_n)$ if and only if $p \geq p'$.

If an agent's preferences satisfy the above axioms, it is possible to find a number u_i for each outcome a_i such that for any two lotteries L and L' the magnitudes of the expected values $p_1 u_1 + \ldots + p_n u_n$ and $p'_1 u_1 + \ldots + p'_n u_n$ indicate the preference between the lotteries. From this assignment of utilities to the basic alternatives, one can construct a utility function *f* over the set of risky alternatives (the lotteries). Consequently, when an agent makes choices consistent with her preferences, she acts as if she is choosing to maximize personal utility as measured by *f*.

Representations of a Game

Games are most commonly represented in an extensive or a strategic form. One also finds the strategic form referred to as the *normal* form, following von Neumann and Morgenstern, who believed that normally one should reduce the extensive form of a game to the strategic form for the purpose of analysis.

The extensive form uses a game tree to represent the order of play (see Figure 1). Each node in the tree represents a *choice point* for a particular player; the player whose turn it is to move at a particular choice point is indicated by a label attached to the node. All games have a privileged node, the *root* or *initial* node where the game begins. The leaves of the tree, also known as *terminal* nodes, represent endpoints, or outcomes, of the game. Every node in the game tree except for the terminal nodes has at least one edge lying on a path between it and a terminal node; such edges represent choices available to a player at that choice point. In some games, the moves available to a player depend not only on the previous moves of other players, but on the outcome of a chance event like the roll of a die. Such games may be represented by including a fictitious player in the game tree, Chance, whose available moves at a point correspond to the possible outcomes of the random event. A player's choice at a given point is a *move* in the game, and each edge has an attached label naming the move. A path from the root node to a terminal node is one possible *play* of the game. In Figure 1, terminal nodes are labeled with *W* or *L*, meaning that Player 1 wins or loses the game, respectively.



(a) Game of perfect information

(b) Game of imperfect information

Fig. 1. A simple two-player game in extensive form. Terminal nodes labeled "W" and "L" indicate whether Player 1 wins or loses, respectively.

If all players know their exact position in the game tree at every point, the game is said to be one of *perfect information*; all other games are of *imperfect* information. Although players do not always know their exact position in the game tree in a game of imperfect information, they often know that their position is one of a limited number of possible nodes. This subset of nodes is a player's information set. In an extensive form game, a player's *strategy* specifies the choices that the player would make at each of his information sets. A player's information sets are indicated in the game tree by grouping together those nodes among which the player cannot distinguish. Thus, an alternative definition of a game of perfect information is one in which all information sets contain only a single node. Figure 2b illustrates a game of imperfect information in which Player 1 moves first but keeps the move hidden from Player 2. When it is Player 2's turn to move, he does not know whether the choice occurs at the left or the right side of the game tree.

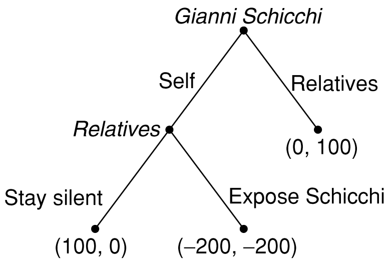
The strategic form of a game is a minimal representation that omits all information about the game except for the relationship between strategies and payoffs. The strategic form of a game consists of a set of players $P = \{1, \ldots, N\}$, a set of pure strategies S_i for each player $i \in P$, and, for each player i, a *payoff function* u_i that maps pure strategy profiles $\sigma = (\sigma_1, \ldots, \sigma_N) \in S_1 \times \ldots \times S_N$ to a real number r. In a two-player game, the strategic form can be represented as a matrix, where each row corresponds to a strategy for Player 1, each column a strategy for Player 2, and each cell the resulting payoffs obtained by Players 1 and 2 when they choose those respective strategies.

In many cases, it proves convenient to allow players to adopt *mixed* strategies, where they choose a pure strategy at random according to some probability distribution defined over the set of pure strategies S_i. The payoff for a mixed strategy $\bar{\sigma}_i$ is defined to be the expected payoff $\sum_\sigma P(\sigma \mid \bar{\sigma}_i) u_i(\sigma)$, where the sum is over all strategy profiles σ and $P(\sigma \mid \bar{\sigma}_i)$ denotes the probability that the strategy profile σ occurs when player i adopts the mixed strategy $\bar{\sigma}_i$.

Although it is often said that which form one uses to represent a game is merely a practical question, on the grounds that any game represented in one form may be represented in the other, this is a topic of some debate. To begin with, it is clear that moving from the extensive form to the strategic form results in a loss of information, for it is possible for two *different* extensive games to have the *same* strategic form. In some cases, this lost information may be relevant to the analysis of the game; if so, it may not always be possible to adequately analyze a game just given its strategic form (see Harper 1988). For example, Figure 2 illustrates the strategic and extensive forms for the decision problem central to Puccini's opera *Gianni Schicchi*, in which the causal dependencies between the players' choices are lost in the normal form, yet seem crucial to the game's analysis.

Noncooperative Games

In a noncooperative game, players independently decide what strategy to adopt in the light of their knowledge of the other players and the payoff matrix. Most of the classical results in game theory have been obtained for noncooperative games, for



(a) Gianni Schicchi, extensive form

	Stay silent, if Schicchi awards fortune to self	Expose Schicchi, if Schicchi awards fortune to self
Relatives	(0, 100)	(0, 100)
Self	(100, 0)	(–200,–200)

(b) Gianni Schicchi, strategic form

Fig. 2. Strategic and extensive form of the Gianni Schicchi. In Puccini's opera *Gianni Schicchi,* the wealthy Buoso Donati dies, and before his will is read, his relatives learn that he has willed a large portion of his fortune to friars. They conspire to have a noted mimic, Gianni Schicchi, impersonate Buoso Donati on his deathbed in order to dictate a new will. Gianni Schicchi agrees but while impersonating Buoso Donati and dictating a new will, he declares his wish to leave a large portion of his fortune to his devoted friend Gianni Schicchi. The relatives contemplate notifying the authorities but decide against it, knowing that the punishment for tampering with a will is banishment and amputation of a hand.

the inability of players to form coalitions and enter into binding agreements make noncooperative games much easier to analyze. Some game theorists (such as Nash) also have a methodological reason for concentrating on noncooperative games: These theorists hold that such games are "more basic" than cooperative games and that the appropriate way to solve a cooperative game is first to transform it into a noncooperative game. However, these views are not universally held (see Osborne and Rubinstein 1994; Binmore 1992).

A *solution* of a game is a specification of the outcomes that may be expected to occur when the game is played by rational agents. Two widely used techniques for solving noncooperative games are *dominance arguments* and *equilibrium analysis.* The goal of each of these approaches is to identify, for each player, a best-response strategy to the anticipated play of all other players. Given a strategy profile σ, the strategy σ_i is a best-response for player i if $u_i(\sigma_{-i}, \sigma_i) \geq u_i(\sigma_{-i}, \sigma_j)$ for all $\sigma_j \in S_i$, where σ_{-i} denotes the set of strategies in the profile σ for the opponents of player i.

A dominance argument rules out certain strategies for play on the grounds that those strategies are inferior to other alternatives, where an inferior strategy is one that is either weakly or strictly dominated: A strategy σ is *weakly dominated* if there exists another strategy σ' such that the payoff from σ' is never worse than the payoff from σ, and there is at least one instance in which the payoff from σ' exceeds that of σ. A strategy σ is *strongly dominated* if there exists another strategy σ^* such that the payoff given by σ^* always exceeds the payoff given by σ.

Iterated elimination of strongly dominated strategies is a procedure for transforming games into a reduced form. One eliminates the strongly dominated strategies for Player 1, transforming the game G to the game G', and then eliminates the strongly dominated strategies for Player 2 from G' to obtain G'', repeating this procedure until no strongly dominated strategies for any player remain. At the end, one obtains a reduced game G^* with the property that every remaining strategy for every player is a best-response to some possible strategy profile. In addition, the resulting game obtained does not depend on the order or the rate at which strongly dominated strategies are removed. This result does not hold for iterated elimination of weakly dominated strategies. The resulting game G^* obtained by iterated elimination of weakly dominated strategies may depend on the order in which strategies are eliminated, as shown in Figure 3. It is never rational to play a strongly dominated strategy, but there are cases where it is not irrational to play a weakly dominated strategy. Although some game theorists freely apply iterated dominance arguments to reduce the complexity of games, others caution against adopting this as a general approach toward their solution (see Binmore 1992).

Although dominance arguments are useful in analyzing a game, the primary tool of analysis in noncooperative game theory is a Nash equilibrium. A strategy profile σ is a Nash equilibrium if each player's strategy is a best-response to the strategies selected by the rest of the players; alternatively, a Nash equilibrium occurs when no player's expected payoff improves by adopting a different strategy unless another player adopts a different strategy as well. More formally, a strategy profile $\sigma = (\sigma_1, \ldots, \sigma_N)$ is a Nash equilibrium if, for $1 \leq i \leq N$, $u_i(\sigma) \geq u_i(\sigma_{-i}, s_i)$ for all $s_i \in S_i$. The wide acceptance of the Nash equilibrium for solving games derives from the fact that it is the only such concept compatible with the rules of the game, the rationality of the players, and the independent selection of strategies all being common knowledge. (For a discussion of common knowledge, see Lewis 1969.)

If players are restricted to pure strategies, not all games have a Nash equilibrium. The game of Matching Pennies, shown in Figure 4, has no Nash equilibrium when the players are restricted to playing either heads or tails. If players may adopt mixed strategies, then it can be shown that all finite games (that is, games in which each player has only finitely many strategies) have at least one Nash equilibrium (Nash 1950).

	μ_1	μ_2
σ_1	(1,1)	(0, 0)
σ_2	(1, 1)	(2, 1)
σ_3	(0, 0)	(2, 1)

Fig. 3. A game in which order matters for the iterated elimination of weakly dominated strategies.

	Heads	Tails
Heads	(1, −1)	(−1, 1)
Tails	(−1, 1)	(1, −1)

Fig. 4. Matching Pennies, a game with no Nash equilibria (in pure strategies).

Refinements of Nash Equilibrium

Although it is generally agreed that a solution to a game must be a strategy profile in a Nash equilibrium, this provides only a necessary, not a sufficient, condition. In general, Nash equilibria lack several desirable properties: They need not be unique, they need not be optimal, and they may allow players to make incredible threats or promises. The game Battle of the Sexes, shown in Figure 5a, has two Nash equilibria, (Boxing, Boxing) and (Ballet, Ballet). The well-known Prisoner's Dilemma, illustrated in Figure 5b, has (Defect, Defect) as its sole Nash equilibria, yet this outcome yields a payoff of 2 to each player, whereas the outcome (Cooperate, Cooperate) yields payoffs of 3. In the game *G* (Binmore 1992), (*rr, LLL*) is a Nash equilibrium, but note that this strategy profile requires that Player 2 commit to playing *L* at node *N*, an irrational move, as Player 2 would thereby lose the game if that node were reached, whereas Player 2 would win by playing *R*. Consequently, a number of refinements and extensions to the concept of a Nash equilibrium have been introduced, two of which are discussed below.

Subgame Perfect Equilibrium

Each node *v* in an extensive game *G* induces a subgame of *G*. A subgame is produced by keeping the node *v*, along with the subtree rooted at *v*, and deleting the rest of the game. If σ is a Nash equilibrium of *G*, it need not be true that σ is a Nash equilibrium for every subgame of *G* as well. Selten (1965) introduced a refinement of the Nash equilibrium concept known as a *subgame perfect* equilibrium, which requires that a strategy profile σ be a Nash equilbrium for every subgame as well. It has been shown that every finite extensive game of perfect information has at least one subgame perfect equilibrium. Since every subgame perfect equilibrium is also a Nash equilibrium, subgame perfection counts as a refinement of the concept of a Nash equilibrium, because it often eliminates Nash equilibria that are unlikely to be adopted by rational players, such as the strategy profile (*rr, LLL*) in the Prisoner's Dilemma game of Figure 5b.

Correlated Equilibrium

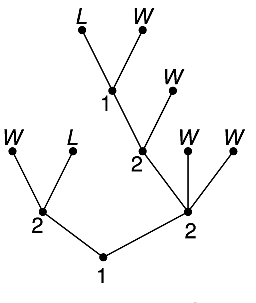
The definition of a Nash equilibrium assumes that the selection of strategies by players occurs independently. Aumann (1974 and 1987) defined a notion of *correlated* equilibrium for noncooperative games. By correlating on shared information about the state of the world (although the information need not be the same for all the players), it is possible for players to arrive at an equilibrium that is self-enforcing in the sense that no player would have reason to deviate from equilibrium play. The fact that correlated equilibria are self-enforcing is significant because it means that adhering to a correlated equilibrium does not require the existence of a binding agreement among the players. In many cases, adopting a strategy profile in correlated equilibrium rewards each player with a higher expected payoff than she could receive in the absence of correlation. For example, consider the game of Battle of the Sexes from Figure 5a and suppose that the players have shared information about the result of a toss of a fair coin. If both players (independently) adopt the strategy of going to a boxing match whenever the coin turns up heads and going to

	Boxing	Ballet
Boxing	(2, 1)	(0, 0)
Ballet	(0, 0)	(1, 2)

(a) Battle of the sexes

	Cooperate	Defect
Cooperate	(3, 3)	(1, 4)
Defect	(4, 1)	(2, 2)

(b) The prisoner's dilemma



(c) The game *G*

Fig. 5. Games with multiple or suboptimal Nash equilibria.

the ballet whenever the coin turns up tails, each player has an expected payoff of $\frac{3}{2}$, a significant improvement upon their expected payoffs in the absence of correlating their strategies. It has been proven that the set of correlated equilibria always contains the set of Nash equilibria and hence is an extension of the concept of a Nash equilibrium.

Cooperative Games

In a cooperative game, players can enter into binding agreements in which they are committed to playing certain strategies. Whereas strategy profiles in noncooperative games need to be self-enforcing (e.g., a Nash equilibrium) in order to be plausible outcomes of play, in cooperative games the binding agreement can be used to bring about any possible outcome. Of the many possible outcomes, how should one be selected?

Nash (1950 and 1953) proposed the following approach to analyzing cooperative games: Although players *may* enter into a binding agreement, they need not. If they choose not to, then there is a noncooperative game in which each player can, adopting the appropriate mixed strategy, be assured of a certain minimum expected payoff; call this outcome the *disagreement point.* The original cooperative game can thus be conceived as a bargaining problem in which players seek to improve their situation by moving away from the disagreement point to a new, more desirable point conferring greater utility. Exactly which point is selected depends upon the particular arbitration scheme used. An arbitration scheme can be thought of as a function mapping the set of possible outcomes to a single outcome—the solution offered by the arbitrator. A cooperative game, then, can be conceived as an extensive form of a noncooperative game where the early stages of the game involve the selection of the disagreement point and the arbitration scheme. This approach, of reducing cooperative games to noncooperative games, is known as the *Nash program.*

Nash argued that a reasonable arbitration scheme for a bargaining problem should satisfy the following four conditions:

- **Pareto optimality:** It is not possible to increase any player's utility without decreasing another player's utility.
- **Independence of irrelevant alternatives:** The selection of the outcome of the bargaining problem should not depend upon alternatives which were not chosen. (One should be aware that Nash's proposed solution is not universally accepted. This axiom is generally viewed as the most controversial.)
- **Symmetry:** If the set of outcomes is symmetric, then the solution point awards the same payoff to all players.
- **Invariance:** Since utility functions are unique only up to a strictly increasing affine transformation, no player should be able to affect the solution point by rescaling his or her utility function.

The fact that there exists a unique outcome satisfying these four conditions was proved by Nash (1950) for the two-person case.

Solution concepts differing from the one suggested by Nash have been defended by Kalai and Smorodinsky (1975), Braithwaite (1954), and Gauthier (1986). The Kalai-Smorodinsky solution has a natural geometric construction that illustrates the underlying intuitions. Define the "Utopia point" as the outcome awarding each player the maximum amount of utility possible for the game under consideration. In all cases of interest, the Utopia point lies outside the set of feasible solutions. Draw a line l connecting the disagreement point to the Utopia point. The point of intersection between l and the Pareto frontier is the Kalai-Smorodinsky solution. That is, the Kalai-Smorodinsky solution is the point arrived at when each player makes "appropriate" relative concessions from the Utopia point. The solution point identified by the Kalai-Smorodinsky solution is often not the same point as that identified by the Nash axioms.

Evolutionary Game Theory

Evolutionary game theory originated as an application of game theory to biology, arising from the realization that frequency-dependent fitness introduces a strategic aspect into evolution. Evolutionary game theory has since become an object of interest to economists in part because the rationality assumptions underlying it are more appropriate for modeling strategic deliberation by real humans, who are only boundedly rational, as opposed to the perfectly rational agents modeled by traditional game theory. In addition, evolutionary game theory provides a way of modeling the dynamics of strategic interaction in a way not possible with the traditional theory of games. Recall that the only way to model the temporal aspect of a game is to use the extensive form of representation. However, methods of analyzing extensive games typically proceed by envisioning that players select

	Rock	Paper	Scissors
Rock	(1, 1)	(0, 2)	(2, 0)
Paper	(2, 0)	(1, 1)	(0, 2)
Scissors	(0, 2)	(2, 0)	(1, 1)

Fig. 6. The game of Rock–Paper–Scissors.

a strategy at the beginning of the game that specifies their course of action at each choice point, which really does not model the dynamical aspect of the game.

The primary equilibrium concept in evolutionary game theory is that of an evolutionarily stable strategy (see Maynard Smith 1982). A strategy is *evolutionarily stable* if when almost every member of the population follows it, no individual who adopts a novel strategy can successfully invade. If σ is evolutionarily stable, the fitness of an individual following σ must be greater than the fitness of an individual following μ (otherwise the individual following μ would be able to invade, and so σ would not be evolutionarily stable). Let $F(s_1, s_2)$ denote the change in fitness for an individual who plays the strategy s_1 against an opponent playing the strategy s_2. Then σ is evolutionarily stable if and only if:

$$F(\sigma, \sigma) > F(\mu, \sigma)$$

or

$$F(\sigma, \sigma) = F(\mu, \sigma) \text{ and } F(\sigma, \mu) > F(\mu, \mu).$$

If a strategy is evolutionarily stable, it must be a best reply against itself, for, if not, a mutant strategy would be able to invade. This means that all evolutionarily stable strategies are Nash equilbria when played against themselves. However, not all games have evolutionarily stable strategies, and not all Nash equilibria are evolutionarily stable. The game of Rock–Paper–Scissors, shown in Figure 6, has a unique Nash equilibrium in mixed strategies where each individual plays Rock, Paper, or Scissors with probability $\frac{1}{3}$, but no evolutionarily stable strategy.

J. McKenzie Alexander

References

Aumann, R. (1974), "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics* 1: 67–96.

——— (1987), "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrika* 55: 1–18.

Binmore, K. (1992), *Fun and Games*. Lexington, MA: D. C. Heath and Company.

Borel, É. ([1921] 1953), "The Theory of Play and Integral Equations with Skew Symmetric Kernels" (translated by L. J. Savage), *Econometrika* 21: 97–100. Originally published as "La théorie du jeu et les équations, intégrales à noyau symmétrique gache," *Comptes Renduc de l'Académie des Sciences* 173: 1304–1308.

Braithwaite, R. B. (1954), *Theory of Games As a Tool for the Moral Philosopher*. Cambridge: Cambridge University Press.

Gauthier, David (1986), *Morals by Agreement*. Oxford: Oxford University Press.

Harper, W. (1988). "Causal Decision Theory and Game Theory: A Classic Argument for Equilibrium Solutions, a Defense of Weak Equilibria, and a New Problem for the Normal Form Representation," in W. Harper and B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics II*. Dordrecht, Netherlands: Kluwer.

Kalai, E., and M. Smorodinsky (1975), "Other Solutions to Nash's Bargaining Problem," *Econometrika* 43: 513–518.

Kreps, D. M. (1990), *Game Theory and Economic Modelling*. Oxford: Oxford University Press.

Lewis, David (1969), *Convention: A Philosophical Study*. Oxford: Basil Blackwell.

Luce, R. D., and H. Raiffa (1957), *Games and Decisions.* New York: Wiley.

Maynard Smith, J. (1982), *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

Myerson, R. B. (1991), *Game Theory: Analysis of Conflict.* Cambridge, MA: Harvard University Press.

Nash, J. (1950), "Equilibrium Points in *N*-person Games," *Proceedings of the National Academy of Sciences of the United States of America* 36: 48–49.

——— (1953), "Two-Person Cooperative Games," *Econometrica* 21: 128–140.

Osborne, M. J., and A. Rubinstein (1994), *A Course in Game Theory.* Cambridge, MA: MIT Press.

Samuelson, L. (1998), *Evolutionary Games and Equilibrium Selection.* Cambridge, MA: MIT Press.

Selten, R. (1965), "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetragheit," *Zeitschrift für die gesampte Staatswissenschaft* 121: 301–324.

von Neumann, John ([1928] 1959), "On the Theory of Games of Strategy" (translated by Sonya Bargmann), in A. W. Tucker and R. D. Luce (eds.), *Contributions to the Theory of Games,* vol. 4 (*Annals of Mathematics Studies* 40). Princeton, NJ: Princeton University Press, 13–43. Originally published as "Zur Theorie der Geselleschaftspiele," *Mathematische Annalen* 100: 295–320.

von Neumann, John, and Oskar Morgenstern (1944), *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

Zermelo, E. (1913), "Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels," in *Proceedings of the Fifth International Congress of Mathematicians* 2: 501–504.

GENETIC INFORMATION

See **Biological Information; Molecular Biology**

GENETICS

Classical Genetics

Genetics was the name given in 1906 by William Bateson (1861–1926) to the emerging branch of biology "devoted to the elucidation of the phenomena of heredity and variation" (Bateson 1928, 1943). The founding opus of genetics is Gregor Mendel's (1822–1884) *Versuche über Planzenhybriden* [Experiments on Plant Hybrids] (Mendel 1866), read at the meetings of the Naturalist Society of Brno (Moravia) on 8 February and 8 March 1865. In 1900, while elaborating his theory of *Intracellulare Pangenesis* into the *Mutationstheorie* on the origin of species by discontinuous, rather than continuous variations (of Darwinian theory), Hugo de Vries (1848–1935) modified his model along the lines of Mendel's hypothesis of particulate inheritance of thirty-five years earlier. However, instead of Mendel's abstract notion of factors for characters, experimentally demonstrated by seven carefully selected discrete traits, de Vries introduced the preformationist notion of organisms composed of "unit characters," for each of which pangenes existed. Cell nuclei, including those of the gametes, contained the full gamut of pangenes, thus providing for continuity of intergenerational inheritance, whereas development was due to differential activation of specific pangenes farmed out to the cytoplasm of the cells of various organs. Thus, whereas de Vries adopted Mendel's insight of dealing with inheritance in terms of *factors* for *discrete traits*, he accorded these abstract factors properties of material entities, introducing into genetic theory a dialectical confrontation that has been part of it ever since.

Mendel, who was educated in the physical sciences, apparently believed that laws of nature were expressible as mathematical statements. His experiments on hybridization, mainly in garden peas, were carefully designed to establish numerical laws for the *inheritance* of selected individual traits, irrespective of the *nature* of these traits. His paper is a masterpiece of didactic presentation of experimental results as support for his theory of inheritance. Mendel posited discrete and independent factors for each trait. A maternal and a paternal factor for any given trait may combine in hybrids without losing their identity and segregate again in the gametes of the hybrids, to combine according to the laws of probability in progeny of further generations (*law of segregation*). Factors of different traits segregate independently of each other (*law of independent segregation*). Plants that produce only one kind of factor for a trait are homozygous for that trait. Those that produce two kinds of factors for a trait are heterozygotes for that trait. Heterozygotes for a trait are often indistinguishable in appearance from one of the homozygotes; the factors for that trait are considered dominant, whereas those of the trait that does not show in heterozygotes are recessive. The alternatives or complementary appearances that a unit can obtain (red or white flowers, A, B, or O blood type) are its *allelomorphs*, or alleles.

Wilhelm Johannsen (1857–1947) studied seed dimensions of bean plants. By repeated inbreeding, "pure lines" were obtained, in which selection was ineffective, since practically all variation among the progeny was due to environmental fluctuations. In 1909, he concluded that the visible or phenotypic variation of a character in a mixed population of individuals is composed of heritable, or genotypic, variance and varies due to nonhereditary fluctuations, or environmental effects. By extending the distinction between the genotypic and phenotypic components of variation to the changes during individual development and to variation in morphological and physiological traits of individuals, Johannsen extended the notion of the predestined factors of unit characters, or the preformationist link between hereditary factors and characters. Following Mendelian theory, Johannsen termed the genotypic component of a distinct character its gene. Genes are *invariant entities of inheritance* and development, which are present in the gametes and the zygotes, through which "a property of the developing organism is or may be conditioned or codetermined" (Johannsen 1911). For Johannsen the concept of the gene was merely an abstraction, an *intervening variable* that purely "summarized" characters, a quantity obtained by a specified manipulation of the values of empirical variables (see Falk 1986):

> The segregation of one sort of "gene" may have influence upon the whole organization. Hence the talk of "genes for any particular character" ought to be omitted . . . . It should be a principle of Mendelian workers to minimize the number of different genes as much as possible. (Johannsen 1911, 147)

Once the notion of unit characters became redundant, biologists could conceive the Mendelian theory of heredity of distinct factors or genes as providing necessary but not sufficient conditions for traits of living organisms. The discriminative trait became merely the phenotypic "marker" of a gene. Evidence from cytological observations indicated that the cell nucleus, or more precisely its chromosomes, provided the material basis of inheritance. Chromosomes maintained continuity between cell generations, and the specific functional role of each chromosome was revealed by the dysfunction of any embryo lacking a full set of them.

Sexually reproducing organisms contain a maternal and a paternal set of chromosomes in each of their cells. Before cell division, or *mitosis*, the chromosomes are duplicated, and a precise division of the duplicated nuclear content to the daughter cells is orchestrated. Before the production of the reproductive cells, or gametes, two nuclear divisions follow only one chromosome duplication. During these coupled divisions, or *meiosis*, a complex process of chromosome pairing and exchange of parts takes place. As a result, corresponding segments of paternal and maternal chromosomes segregate to different gametes, so that each gamete contains a single but full set of chromosomes, although no set is either paternal or maternal. Edmund B. Wilson (1856–1938) and his students showed that chromosome pairs segregate independently at meiosis and suggested that segregation of chromosomes at meiosis and their recombined association at fertilization is what is expected of the material bearers of Mendelian factors.

In 1910, Thomas H. Morgan (1866–1945) observed genes that segregated according to the pattern of the sex chromosomes of his new experimental organism, the fruit fly *Drosophila melanogaster*: Of the pair of sex chromosomes (*X* chromosomes) present in females, only one—of maternal origin—is found in males (who have a paternal *Y* chromosome instead of a second *X* chromosome). Morgan adopted an instrumental approach to genes as entities detected by function while accepting that, materially, these were entities localizable to chromosomes. Thus, rather than being abstract intervening variables, genes were envisioned by Morgan as *hypothetical constructs* to which existence properties were added that were not explicitly defined by the empirical relations. With this dialectical approach, Morgan maintained an epigenetic view of many-to-many relationships between genes and characters, in which there were "manifold effects of each gene" and "each character is the product of many genes" (Morgan 1917).

However, not all observed deviations from independent segregation of traits were explicable in terms of the same gene affecting several traits (*pleiotropy*). Such correlation was considered to be due to material dependence between different genes:

> If the materials that represent these factors are contained in the chromosomes, and if those factors that "couple" be near together in a linear series, then when the parental pairs (in the heterozygote) conjugate, like regions will stand opposed. There is good evidence that during [meiosis] homologous chromosomes twist around each other, but when the chromosomes separate (split), the split is in a single plane . . . . In consequence, we find coupling in certain characters and little or no evidence at all of coupling in other characters; the difference depending on the linear distance apart of the chromosomal materials that represent the factors. (Morgan 1911)

Morgan's distinction between multiple effects of genes and physical dependence between genes was elaborated by him and his students into the theory of *genetic linkage*, according to which genes are located at specific *loci* along the chromosomes. Linked genes on a given chromosome may recombine at a rate that depends on the relative distance of the loci along the chromosome. Alfred H. Sturtevant (1891–1971) provided in 1913 the first linkage (or recombination) map of a chromosome of *Drosophila melanogaster.* These maps, however, were merely of abstract intervening variables of experimental linkage data, that is, linear representations of the deviations from independent segregation of gene pairs in Mendelian hybridization experiments. Notwithstanding, Morgan's students, notably Calvin B. Bridges (1889–1938) and Hermann J. Muller (1890–1968), provided increasing evidence for the genes being discrete material entities arranged along the chromosomes: "[B]esides the ordinary proteins, carbohydrates, lipoids, and extractives, of their several types, there are present within the cell *thousands* of distinct substances—the 'genes'; these genes exist as ultramicroscopic particles" (Muller 1922, 32).

The necessary prerequisite properties of these atoms of heredity are:

- self-replication, or autocatalysis;
- involvement in physiological and developmental processes of organisms, or heterocatalysis; and
- a form of catalysis that upon a change in the structure of the gene "may become correspondingly changed, in such a way as to leave [the gene] still *auto*catalytic" (Muller 1922, 34).

The third property derives from Muller's insistence that given "inheritance *of* change," evolution would automatically follow (35). It was through this "general feature of gene construction," which was indispensable for matter evolving by a Darwinian process of trial and error, that Muller set out to investigate the genes. Such a pivotal image of the gene also led, however, to the genocentric notions that overwhelmed future discourse far beyond the image's heuristic value, although Muller himself emphasized that whatever the genes may be or do, they make sense only in the context of the living cell and its environment:

> Each of these effects, which we call a "character" of the organism, is the product of a highly complex, intricate, and delicately balanced system of reactions, caused by the interaction of countless genes, and every organic structure and activity is [liable to become altered] when the balance of the reaction system is disturbed by an alteration in the nature or the relative quantities of any of the component genes of the system. (Muller 1922, 33)

Muller developed quantitative methods of mutagenesis to investigate the physical properties of the genes, and in 1927 showed that x-rays may induce mutations and chromosomal aberrations. These studies also allowed correlations to be established between genes and their location on chromosomes. The discovery of giant "polytenic" (multistranded) chromosomes in the cells of some insect larvae finally allowed the detailed physical maps of genes on chromosomes. This confirmed the collinearity of the linkage maps and the cytological maps of chromosomes (Figure 1).

However, the dialectics of the theory of the genes did not conceive of them as necessarily particulate atomic entities of heredity. Richard Goldschmidt (1878–1958) conceived of whole chromosomes as integrative functional entities. Changes, such as breaks and rearrangements in the chromosomal continuity, cause functional deviations that may operationally be localizable as mutations in discrete genes. Induced changes in the arrangement of chromosomes that did affect function (*position effects*) supported this notion. L. J. Stadler (1896–1954) emphasized as late as 1953 that the operational tests to support the existence of genes could not prove their indivisibility (Stadler 1954). Doubts about genes as the physical entities of heredity grew when estimates of their size changed under different conditions of mutagenesis. Also, intensive experiments revealed that recombination could occur between what were considered to be alleles, alternative mutants of the same gene. Such a possibility to separate by recombination what turned out to be similar yet different adjacent functional entities, or *pseudo-alleles*, was first believed to be a property of complex loci, but it eventually allowed the experimental analysis of the gene. By 1953, when the structural organization of the hereditary material became clear, the indivisible nature of the abstract and cytological gene entities had already been replaced by a gene analyzable by intragenic recombination in organisms from Drosophila to the mold *Aspergillus nidulans.* Contrary to Muller's project to study the properties of the gene indirectly because "[a] gene can not effectively be ground in a mortar, or distilled in a retort" (Muller 1922, 36), it was eventually the physicochemical analysis of molecules that resolved the puzzle (Watson and Crick 1953a and 1953b). It did not escape the notice of Watson (1928–) or Crick (1916–2004) that their initial paper, "Genetical Implications of the Structure of Deoxyribose Nucleic Acid [DNA]," addressed exactly the three properties that Muller expected of genes as the atoms of inheritance.

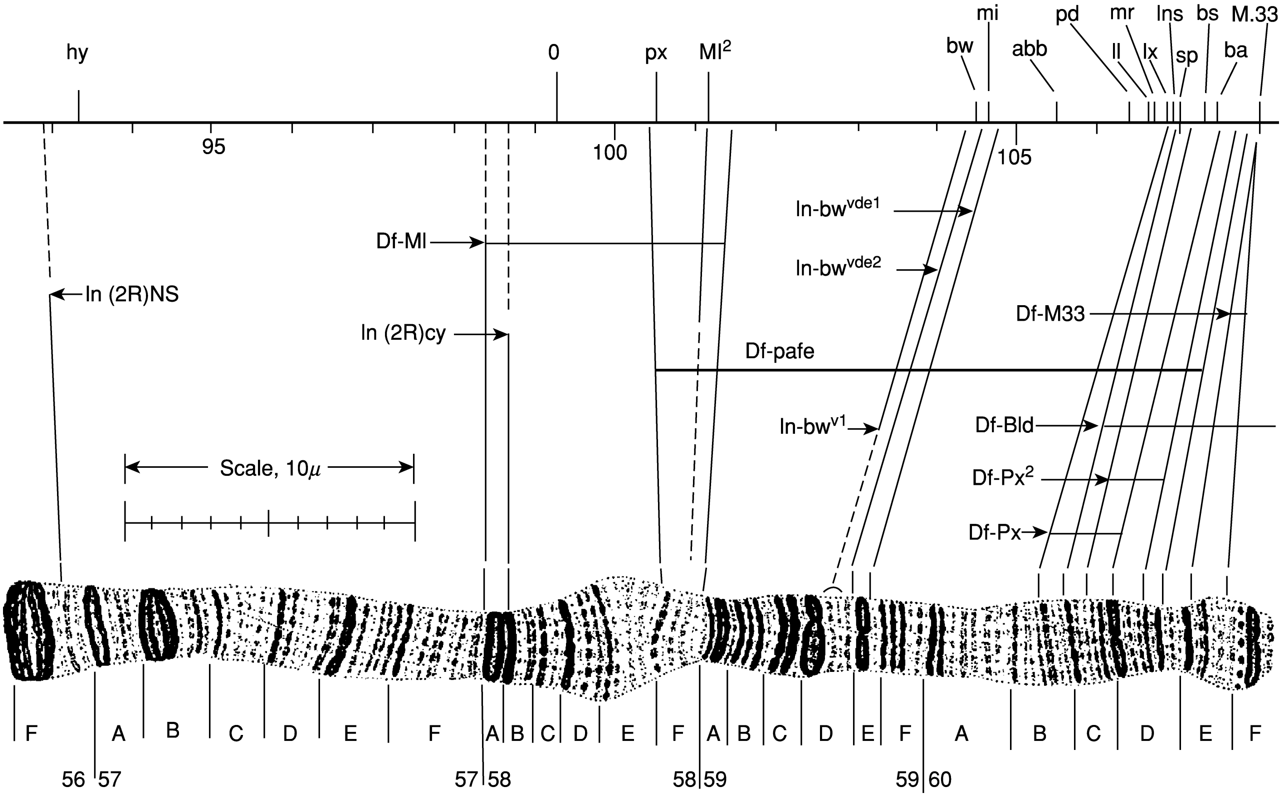


Fig. 1. *Drosophila* linkage map and polytenic chromosome map, aligned.

Thus, genetic analysis of abstract, intervening variables, which evolved in parallel with a phenomenological approach engaged with hypothetical constructs, finally landed at the physicochemical definition of molecular genetic matter. However, the Watson-Crick model of the molecular structure of DNA did not indicate any secondary organization into discrete entities, and experimental data showed chromosomes to be straightforward, continuous DNA sequences. Thus, the model of the molecular structure of DNA did not resolve the confrontation of genes as abstract entities versus that of genes as material atoms of heredity, and the dynamics of this dialectic still play a major role in genetic research. Population geneticists and breeders still may refer to genes as variables in a frequency distribution space, although the complexity of the organization of the genetic material is acknowledged. Likewise, reference to phenomenological entities as "genes for" diseases or behavioral properties are made frequently, often irrespective of available information on their detailed molecular structure.

Formal Genetics

In the early 1940s, the modern synthesis suggested that evolution should be expressed in terms of changes in gene frequencies in populations of interbreeding individuals (see Evolution). The basic law of population genetics was formulated in 1908 independently by G. H. Hardy (1877–1947) and by WilhelmWeinberg (1862–1937) (see Population Genetics). The Hardy-Weinberg law posits that in an infinitely large, randomly mating population, where p is the frequency of allele A_1 and q is the frequency of allele A_2 at a given locus ($p + q = 1$), the frequencies of the genotypes at that locus will be p^2 A_1A_1, $2pq$ A_1A_2, and q^2 A_2A_2, as long as no other forces, like mutation, selection, or migration, affect this population. In other words, within one generation of random mating in an infinite population, in which no outer forces act on the alleles of the gene under consideration, equilibrium in genotype frequencies will be established. Population genetics is essentially the study of theoretical and empirical factors that may cause deviations from Hardy-Weinberg equilibrium.

An important project, especially for animal and plant breeders, was the extension of Mendelian inheritance to quantitative traits, or traits with continuous variation. The effect of numerous genes, each affecting the trait only slightly and more or less to the same extent, would give a binomial distribution of genotypes in a population, which, considering environmental fluctuations, would dissolve into a normal distribution of phenotypes. According to the instrumental reductionist approach, as many genes are allocated to the quantitative trait as are necessary to explain its distribution (Sarkar 1998). The formal analysis of quantitative traits by R. A. Fisher (1890–1962) (Fisher 1918) reconciled the biometricians' interpretation of inheritance with Mendelism. This could be used to construct efficient breeding designs. Even today, identification of quantitative trait loci (QTLs), which are identified as variables at the abstract or phenomenological level, may provide anchors for the search of the functional equivalents at the molecular segments.

Mendelian genetics was also extended to deliberations on biological impacts on human societies. As early as the end of the nineteenth century, Francis Galton (1822–1911) introduced the notion of eugenics, the application of the insights of the science of inheritance and evolution to humans, in order to prevent an anticipated "deterioration of the species." Human populations may be exposed to the effects of mutation and selection like those of any other organism. Social revolutions and advances in health services allegedly caused relaxation of selection, which had to be countered. In the first decades of the twentieth century, the eugenics movement got widespread support from geneticists, who saw it as part of their moral and social obligation to face the consequences of their scientific insights. However, persons who wished to use eugenic arguments to promote social and political aims increasingly usurped the eugenics movement. Eugenics became an important discriminatory tool in the hands of social and political conservatives as well as reformers. At the level of genetic theory, eugenic thinking has suffered too much from oversimplified reductionism, underestimating the extent of environment/genotype interactions and their flexibility, as well as the interactions of the *genome*, the complete collection of genetic information. (For details, see Paul 1995.)

Material Genetics

Although most geneticists accepted the chromosomal theory of heredity from early on, for many years few efforts were made to investigate the chemical aspects of chromosome structure or function. Following Troland, Muller believed that genes acted like enzymes. Suggesting that the newly discovered bacterial viruses might be "naked genes," he hoped that geneticists would "be able to grind genes in a mortar and cook them in a beaker after all" (Muller 1922, 48). The physician Archibald Garrod (1857–1936) recognized as early as the 1910s that gene mutations caused dysfunction or malfunction of relevant enzymes in the normal metabolic pathways of the organism, and accordingly interpreted some diseases as "inborn errors of metabolism." This idea was elaborated by Beadle (1903–1989) and Tatum (1909–1975) into the "one gene–one enzyme" concept, according to which each gene is responsible for one specific enzyme. Beadle and Tatum (1941) studied the growth capacity of the bread mold *Neurospora crassa* on well-defined media from which specific nutrients could be omitted *ad lib.* The one gene–one enzyme concept provided a major framework for genetics, although it was soon overhauled when it turned out that more than one gene may be involved in an enzyme or that genes may code for structural, nonenzymatic proteins.

Seymour Benzer's (1921–) high-resolution recombination analysis of the *r*II gene of the bacterial virus, the bacteriophage T4, indicated that the gene as a functional unit, or *cistron*, might be presented as a linear recombination map of mutated sites. When he had "run the map into the ground," it was possible "to translate linkage distances, as derived from genetic recombination experiments, into molecular units"(Benzer 1955). Similar, though less extensive, recombination experiments with bacterial genes proved that the information along cistrons was collinear with that of the sequence of amino acids in the polypeptides corresponding to those cistrons. The primary information for polypeptide structure is materially coded in the DNA molecule, as revealed by the linear recombination map of the functional units.

Although Friedrich Miescher (1844–1895) had identified nucleic acids by the end of the nineteenth century, and increasingly overwhelming evidence for their involvement in heredity had accumulated since, the role of the carrier of the genetic specificity was persistently ascribed to the protein component of the chromosomal nucleoproteins. The structure of nucleic acids was believed to be a monotonous, repetitive polymer inadequate for encoding complex hereditary specifications. What was crucial was the recruitment of bacteria and viruses as experimental systems for the elucidation of this problem. This was possible only after demonstrating that prokaryotic microorganisms, that is, organisms lacking discernible cell nuclei, obey the same rules of random mutations and selection that were accepted for Darwinian evolution in eukaryotes, those organisms with well-defined cell nuclei (Luria and Delbrück 1943). Still, even in the 1950s, the convincing evidence for the role of DNA was derived from elegant experiments with bacterial viruses rather than from the more scrupulous, straightforward chemical work with bacteria (Avery, MacLeod, and McCarty 1944).

The DNA model of Watson and Crick is that of a double helix constructed of two antipolar strands. Each strand is a string of nucelotides composed of deoxyribose (S) phosphoric acid (P) and a nitrogen base. There are four nucleotides in DNA: adenine (A), guanine (G), thymine (T), and cytosine (C). The strands are held together by weak hydrogen bonds between complementary bases. As a rule, A pairs with T, and G pairs with C. There are no structural limitations on the sequence of nucleotides along the helices (Figure 2).

The Watson-Crick model of the molecular structure of DNA was enthusiastically adopted largely because of the elegance with which it purported to resolve the three properties that Muller assigned to the entities of genetic material. Self-replication was shown to be accomplished by each strand becoming a template for a complementary new strand (*semiconservative replication* [Meselson and Stahl 1958]). The lack of constraints on the sequence of nucleotides along the strands allowed for the endless variability needed for coding genetic specificity, now conceptualized as genetic information. Finally, the consistent structure of the backbone of the helical strands allowed exchange of one base pair for another or rearrangement of whole sequences, causing changes in coding without affecting the self-replication or coding capacity of the molecule, that is, mutations. However, major physicochemical problems, such as the unwinding of the two strands at replication, or the directional specificity of replication enzymes facing the opposite polarity of the two strands, were resolved only years after the model was firmly established. (This process of connecting formal genetics with material genetics has been often controversially interpreted by philosophers of biology as a case of reduction) (Sarkar 1998) (see Reductionism).

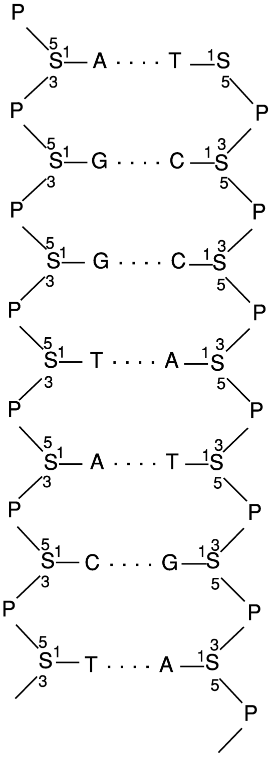


Fig. 2. The Watson-Crick model of double-stranded DNA.

Molecular Genetics

Enzymes and many cell-structure components are proteins. Proteins are polypeptides composed of specific sequences of an array of (usually) twenty amino acids. Protein function relies on the molecules' folding into three-dimensional structures that depend on the sequence of the amino acids in the polypeptide (and on the cellular environment). The Watson-Crick model posits that the information for the sequence of amino acids in polypeptides is encoded in the sequence of the DNA base pairs (see Molecular Biology). In 1958, Crick formulated the central dogma of molecular biology (Crick 1958), according to which the flow of genetic information is unidirectional, from the nucleic acids to proteins but never from proteins to nucleic acids, proving "beyond any doubt but in a totally new way the complete independence of genetic information from events occurring outside or even inside the cell" (Judson 1979, 217). The dogma posited further that the information in the DNA is first transcribed into intermediary polynucleotide molecules, from which it is translated into amino acid sequences at the sites of protein synthesis (see Biological Information). The basic details of cellular-information reading have been elucidated mainly in prokaryotic systems. The intermediaries are molecules of ribonucleic acid (RNA), termed *messenger RNAs* (mRNA). RNA is, as a rule, a single-stranded polymer of nucleotides composed of a ribose (instead of the deoxyribose of DNA) and phosphate in the backbone and of four kinds of bases attached to the ribose residue: adenine (A), guanine (G), uracil (U), and cytosine (C). Transcription is mediated by RNA-polymerase complexes that bind at sites upstream of the sequences to be transcribed and is accomplished by nucleotide complementarity of DNA and RNA:

- G with C
- C with G
- T with A
- A with U.

Translation occurs on special cytoplasmic organelles, the ribosomes, and is catalyzed by them. To code for all twenty amino acids, sequences of three bases of nucleic acids are needed. The code was found to be redundant and comma free. Of the 64 possible triplets, or codons, 61 are sense codons, 1–6 of which code for each of the twenty amino acids; the remaining 3 are termination signals, or nonsense codons (Figure 3). On the ribosomes the code is read sequentially from the mRNA, one triplet after another, from a given starting point (Crick et al. 1961). Each codon is translated into its corresponding amino acid by a specific molecule of transfer RNA (tRNA). Amino acids are enzymatically attached to their specific tRNA, and when the specified tRNA anticodon sequence pairs with its complementary codon in the mRNA on the ribosome, the amino acid is transferred from the tRNA to the nascent polypeptide chain.

Regulation of the genes' activity occurs at transcription as well as posttranscription levels. The elaboration by Jacob (1920–) and Monod (1910–1975) of the negative feedback regulation mechanism of transcription of an (adaptive) enzyme, β-galactosidase, in the bacterium *Escherichia coli* became paradigmatic for genetic regulation (Jacob and Monod 1961). Transcription initiation is controlled by the attachment of an RNA polymerase at the *promoter* site. Numerous transcription factors must combine with the polymerase for its proper function. Various intracellular metabolites (or extracellular *ligands* that attach to cellular *receptors*) affect the formation of different transcription factor complexes, which allow the polymerase to initiate transcription at specific sequences, thus serving as cues that regulate transcription. Regulation may occur by negative-feedback as well as by positive-feedback mechanisms. Although the details of the transcription from DNA and the translation to polypeptides were elaborated in prokaryotes, the essential features were found to hold for all living cells. Furthermore, the same genetic code (with some minor but important exceptions) holds throughout the living world. This strongly endorses the Darwinian model of evolution from an early common ancestor.

However, the expectation of the early molecular geneticists that what was true for *E. coli* was also true for the elephant was exaggerated: Major cellular systems of eukaryotes and prokaryotes diverge significantly. Cells of eukaryotes usually contain orders of magnitude more DNA per nucleus than do prokaryotic cells, in spite of the fact that their basic cell maintenance functions are not much more numerous or complex. Britten and Kohne (1968) found that the nuclear DNA of most mouse cells contains highly repetitive sequences (some of these up to a million times). Such redundancy suggests that these sequences are not involved directly in coding or regulatory functions. In many eukaryotic genomes, not more than 10 percent of the DNA appears to be "meaningful." The observation that often the evolutionarily older taxonomic groups are those especially rich in this repetitive, so-called junk DNA has been described as the C-value paradox. Did birds and mammals evolve more efficient cellular household mechanisms for functions such as packing and unpacking of DNA, instead of the "primitive"

UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine
UUC		UCC		UAC		UGC	
UUA	Leucine	UCA		UAA	Stop†	UGA	Stop
UUG		UCG		UAG		UGG	Tryptophan
CUU		CCU	Proline	CAU	Histidine	CGU	Arginine
CUC		CCC		CAC		CGC	
CUA		CCA		CAA	Glutamine	CGA	
CUG		CCG		CAG		CGG	
AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine
AUC		ACC		AAG		AGC	
AUA		ACA		AAA	Lysine	AGA	Arginine
AUG	Methionine*	ACG		AAG		AGG	
GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine
GUC		GCC		GAC		GGC	
GUA		GCA		GAA	Glutamic acid	GGA	
GUG		GCG		GAG		GGG	

Fig. 3. The genetic code.

mechanisms that need lots of DNA in lungfish and amphibians?

Another unexpected experimental finding, inconsistent with the concept of the gene as a coherent entity of information, was that most eukaryotic coding sequences are interrupted by numerous noncoding sequences (introns). Introns vary in length and sometimes comprise sequences many times longer than the coding exons, the continuity of which they interrupt. The RNA that is transcribed from such DNA sequences is processed by splicing out the introns before the sequence of continuous exons forms a translatable mRNA. Splicing of the introns provides the cells with another level of regulatory control. This includes alternative splicing, whereby numerous different alternative assemblies of sequential exons may be spliced from a given transcription product. Alternative splicing of the same RNA stretch thus effectively codes different mRNA, hence different polypeptides.

Usually only one DNA strand, the "sense" strand, is transcribed into RNA from a given DNA sequence. Sometimes, however, coding regions overlap: Both RNAs could be transcribed from (partly) overlapping sequences of the same DNA strand or from opposite strands, one being the sense strand for one transcript, the other being the sense strand for the other transcript. *RNA editing* by enzymatically changing single nucleotides or whole stretches is another device to increase the repertoire of polypeptides translated from a given sequence of DNA. Thousands of polypeptides were experimentally shown to be referable to given DNA sequences, and more than ten thousand have been predicted for some, defying the one gene–one enzyme notion. No reduction of the classical notion of a gene to such molecular concepts seems possible (see Reductionism). Genes as material entities become merely generic terms for DNA stretches that code some information, whether structural or regulatory (see e.g., Beurton, Falk, and Rheinberger 2000).

Genetics in Context

Arguably, no new major concepts beyond that of the Watson-Crick model of DNA and Crick's central dogma have been formulated by molecular genetics because none are needed (see, however, Molecular Biology). Empirical molecular biology provided the insights that allowed phenomena of genetics to be expressed in physicochemical and physiological terms. Although no reduction of formal genetics to molecular genetics may be possible, biochemical and biophysical details replaced one by one the old concepts of abstract and phenomenological genetics (see Reductionism).

At the beginning of the 1970s, genetic research underwent a profound methodological turn with the introduction of controlled *in vitro* splicing of DNA sequences from any source and the use of appropriate vectors to insert such engineered sequences into host cells, irrespective of the donor's relationship to the host. The possible ethical and social repercussions of this development are obvious, and the scientists involved were the first to take notice (see Krimsky 1982). Genetic engineering completely revolutionized research not only in genetics and its classic sister disciplines, developmental and evolutionary biology, but also in more remote disciplines of the life sciences, such as physiology and neurobiology. With the beginning of the twenty-first century, when sequencing of the complete genome of organisms has become routine, genetic research is undergoing another major conceptual breakthrough with *genomics* and *proteomics,* focusing on the integrated study of whole genomes and on structural and functional interactions instead of the classical Mendelian concentration on one factor at a time.

Genetics has extended far beyond problems of heredity and variation of individuals. The elucidation of the principles of gene regulation allowed a molecular extension of embryological *Entwicklungsmechanik*, or as it is now called, "developmental biology," whose major mode is positive regulation of initiation of transcription. Any protein that is needed for the initiation of transcription but that is not itself part of RNA polymerase is defined as a transcription factor. Transcription factors are provided under tissue-specific control to activate a promoter or a set of promoters that contain a common target sequence upstream of the transcription initiation points. Initiation at a promoter involves a large number of factors. Some recognize the specific target sequences and, once bound to DNA, bind by protein–protein interactions to other components of the transcription apparatus. A generic promoter usually functions at a low efficiency. *Enhancer* sequences that are a major target for tissue-specific or temporal regulation are located often at considerable distance from the start point.

Genetic analysis heavily relies on deducing the normal from the deviant, whether natural or induced. A major feature of most eukaryotes is the defined life span of the organism, a property that extends to the individual somatic cells, whose growth and division is highly regulated. Genetic instability is thought to transform normal cells into cancerous ones. As a rule, growth of transformed cancer cells is less restricted or dependent on external cues than is growth of normal cells, and cancer cells appear to be immortal. Usually, multiple genetic changes are necessary to create a cancer, and the virulence of a cancer may increase as the result of progressive series of changes. One group of genes in which mutations cause transformation is the oncogenes, which have cellular counterparts: the *proto-oncogenes*, which are involved in normal regulated cell function. The generation of an oncogene is by a mutation that inappropriately activates the regulated proto-oncogene. Another tumorigenic factor is loss of function of suppressor genes that usually impose constraints on cell cycle or cell growth. Finally, tumors may result from defects in genetic checkpoint systems that should prevent further damage in cells that went astray by inducing repair mechanisms or by initiating programmed cell death, or apoptosis.

As noted, the near-perfect universality of the genetic code and of the machinery of transcription and translation strongly support the Darwinian hypothesis of evolution of life from a common primeval ancestor by a long process of trial and error of random mutations and natural selection. However, the enormous amount of genetic variability at the level of proteins and DNA suggests that not all of it could be driven or maintained by natural selection. Theoretical considerations indicate that much of this variation is due to random fluctuations of adaptively neutral or near-neutral mutants (Kimura 1968). Physical association with loci that are selected for or against may also affect genetic variability in neighboring stretches of DNA.

In vivo and *in vitro* juxtaposition of DNA sequences from different organisms and the examination of their homologies turned DNA manipulation into a central tool of evolutionary analysis. The insertion of foreign DNA sequences into cells, with or without the knockout replacement of the indigenous sequences, indicated the functional conservation of sequences. Surprisingly many genes were found to be *orthologous*, consisting of homologous, highly conserved DNA sequences in different species; the conservation is even more impressive at the level of the corresponding amino acids (the greater identity of amino acid is due to the redundancy of the genetic code). Likewise many sequences *within* the same genome were found to be homologous (*paralogous*), indicating intensive intragenomic duplication of sequences during evolution; the duplicates were usually modified and mobilized for new related or unrelated functions. A classic example is the human gene that codes the globin component of myoglobin, a structural protein of the muscle fibers, which is paralogous to the respective genes that code embryonic, fetal, and adult components of the blood hemoglobin. Orthologous genes for globin are found throughout the organic world, including in some plant species. Studies have revealed not only the evolutionary path of genes and proteins, but also the developmental constraints on evolution. *Pax*6 is a *homeotic* "master gene" for eye formation in mice as well as in *Drosophila*. A mutation in it may alter a *Drosophila* eye into a homologous structure. The sequence of many master genes contains a special domain, such as the *homeobox* in homeotic genes, which codes for an amino acid domain involved in DNA binding of transcription factors. Such homeoboxes are highly conserved in the evolution of developmental master genes, and orthologous copies are found throughout the animal kingdom, often with many paralogous copies in each.

Such detective work of relationships led to major reevaluations of accepted patterns of the evolution of species. Bacteria were split into two kingdoms. That of the *Archea*, most of the present members of which inhabit niches of high salt concentration and/or extreme temperatures, seem to be nearer relatives of the ancestors of eukaryotes than are the more common *Eubacteria*. Concomitantly, it was surmised that intensive lateral gene transfer must have been the rule in evolution, even between cells belonging to different kingdoms, especially in early phases of the evolution of life. The breakthrough in the evolution of eukaryotes was apparently facilitated by the incorporation of mitochondria and chloroplasts in their cells by symbiosis with prokaryotes, which turned obligatory. The fact that genes like *Pax*6 have similar functions in such diverse eye structures as those of arthropods, mollusks, and vertebrates suggests that much of what was considered to be convergent evolution should be regarded as divergent evolution.

The upsurge of the study of the whole genome as an entity, which depends on the development of techniques such as simultaneous screening of micro-arrays of many thousands of genes or their products, shifts the attention of genetics to multiple interactions between genes and between proteins. A significant insight from these studies is the extent of homeostatic buffering that interactions of integrated gene functions provide even at the most basic functions of living cells. Changing variables, sometimes over several orders of magnitude, may hardly affect the stability of systems in which they are involved. This could provide a new challenge to theories of evolution and development. However, such developments signify a change not only in the conceptions of genetic control of cellular and organismic development and function, but also at the practical level of their application. Technologies of *transgenic*, "genetically modified" domestic animals and plant crops have already affected various aspects of society. The impacts of gene therapy on humans appear to have to wait somewhat longer.

Philosophers of science tried for many years to establish the continuity of genetic theory by formally reducing classical or Mendelian genetics to molecular genetics (see Sarkar 1998). When this failed, it was concluded that genetic theory incorporates essentially at least two incommensurable concepts, best explicated in the central entity of genetics, the gene. Representative of this is Moss's (2003) conceptual analysis of the gene, which results in "defining and distinguishing two different genes. . . . The preformationistic gene (Gene-P) predicts phenotypes but only on an instrumental basis where immediate medical and/or economic benefits can be had. The gene of epigenesis (Gene-D), by contrast, is a developmental resource that provides possible templates for RNA and protein synthesis but has in itself no determinative relationship to organismal phenotypes"(xiv). Such an analysis, falling back on developments in the analysis of whole genomes and of "forward genetics," which purports to predict the function of sequences directly from their sequence, underestimates the fact that throughout the billions of years of organismic evolution, no novel *structurally discrete* DNA entities evolved. It has been a dialectical, philosophically loose discourse, which allowed functions to instrumentally parse DNA and refer to sequences as genes.

Raphael Falk
Sahotra Sarkar

The authors acknowledge the helpful input of Pamela Lyon.

References

Avery, Oswald T., Colin M. MacLeod, and Maclyn McCarty (1944), "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types," *Journal Experimental Medicine* 79: 137–158.

Beadle, George W., and Edward L. Tatum (1941), "Genetic Control of Biochemical Reaction in *Neurospora,*" *Proceedings of the National Academy of Science, Washington* 27: 499–506.

Benzer, Seymour (1955), "Fine Structure of a Genetic Region in Bacteriophage," *Proceedings of the National Academy of Science, Washington* 41: 344–354.

Beurton, Peter J., Raphael Falk, and Hans-Jörg Rheinberger (2000), *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives.* Cambridge and New York: Cambridge University Press.

Britten, Roy J., and D. E. Kohne (1968), "Repeated Sequences in DNA," *Science* 161: 529–540.

Crick, Francis H. C. (1958), "On Protein Synthesis," in *Symposium of the Society for Experimental Biology: The Biological Replication of Macromolecules*, 138–163. Cambridge: Cambridge University Press.

Crick, Francis, H. C., Leslie Barnett, S. Brenner, and R. J. Watts-Tobin (1961), "General Nature of the Genetic Code for Proteins," *Nature* 192: 1227–1232.

Falk, Raphael (1986), "What Is a Gene?" *Studies in the History and Philosophy of Science* 17: 133–173.

Fisher, Ronald A. (1918), "The Correlation Between Relatives on the Supposition of Mendelian Inheritance," *Transactions of the Royal Society, Edinburgh* 52: 399–433.

Jacob, François, and Jacques Monod (1961), "Genetic Regulatory Mechanisms in the Synthesis of Proteins," *Journal of Molecular Biology* 3: 318–356.

Johannsen, Wilhelm (1911), "The Genotype Conception of Heredity," *American Naturalist* 45: 129–159.

Judson, Horace Freeland (1979), *The Eighth Day of Creation: Makers of Revolution in Biology.* New York: Simon and Schuster.

Kimura, Motoo (1968), "Genetic Variability Maintained in a Finite Population Due to Mutational Production of Neutral and Nearly Neutral Isoalleles," *Genetical Research* 11: 247–269.

Krimsky, S. (1982), *Genetic Alchemy: The Social History of the Recombinant DNA Controversy.* Cambridge, MA: MIT Press.

Luria, Salvator E., and Max Delbrück (1943), "Mutations of Bacteria from Virus Sensitivity to Virus Resistance," *Genetics* 28: 491–511.

Mendel, Gregor (1866), "Versuche über Pflanzenhybriden," *Verhandlungen Naturforscher Verein, Brunn* 4: 3–47.

Meselson, Matthew, and Franklin W. Stahl (1958), "The Replication of DNA in *Escherichia coli,*" *Proceedings of the National Academy of Science, Washington* 44: 671–682.

Morgan, Thomas. H. (1911), "Random Segregation versus Coupling in Mendelian Inheritance," *Science* 34: 384.

——— (1917), "The Theory of the Gene," *American Naturalist* 51: 513–544.

Moss, Lenny (2003), *What Genes Can't Do.* Cambridge, MA: MIT Press.

Muller, Hermann J. (1922), "Variation Due to Change in the Individual Gene," *American Naturalist* 56: 32–50.

Paul, D. B. (1995), *Controlling Human Heredity: 1865 to the Present*. Atlantic Highlands, NJ: Humanities Press.

Sarkar, Sahotra (1998), *Genetics and Reductionism.* New York: Cambridge University Press.

Stadler, Lewis J. (1954), "The Gene," *Science* 120: 811–819.

Watson, James D., and Francis H. C. Crick (1953a), "Molecular Structure of Nucleic Acids," *Nature* 171: 737–738.

——— (1953b), "Genetical Implications of the Structure of Deoxyribose Nucleic Acid," *Nature* 171: 964–967.

See also **Biological Information; Evolution; Molecular Biology; Population Genetics; Reductionism**

H

HANS HAHN

(27 September 1879–24 July 1934)

Hahn was a mathematician whose contributions to analysis and topology were outstanding. In addition, he had a remarkable influence on twentieth-century philosophy—less through his writings (a mere handful of essays) than by bringing together and stimulating other thinkers. Hahn was instrumental in founding and running the Vienna Circle (see Vienna Circle). He was the thesis adviser of Kurt Gödel and a mentor of Karl Popper (see Popper, Karl Raimund). In addition, he had a hand in the chain of events that brought Ludwig Wittgenstein back to philosophy. He was both a front-seat witness and a catalyst of the great foundational debate on mathematics and logic that took place in the first third of the twentieth century.

Hahn was born on September 27, 1879, in Vienna. His father, a former music critic, eventually became one of the highest-ranking officials in the Austro-Hungarian empire. Hans Hahn grew up in the center of the fervid Viennese *fin de-siècle* atmosphere that produced Freud, Mahler, and Kokoschka. The philosophical giants of his youth were Mach and Boltzmann, who, while strongly conflicting in most of their views, were both eminent physicists with a positivistic worldview.

Both Hahn's philosophical inclinations and his networking style became apparent already before the First World War. As a young mathematician, he belonged to a group of intellectuals that included the social scientist Otto Neurath, the applied mathematician Richard von Mises, and the theoretical physicist Philipp Frank (see Neurath, Otto). The group met in Viennese coffeehouses to discuss philosophical topics, influenced by the work of Bertrand Russell, Henri Poincaré, Émile Duhem, and Heinrich Hertz. In retrospect this can be seen as the forerunner of the Vienna Circle. Indeed, when in 1909 Hahn got his first appointment as a professor in far-flung Czernowicz (an outpost of the multiethnic empire of Emperor Franz Josef), he announced to his friends that on his eventual return to a chair in the capital, they would resume their discussions with the participation of a university philosopher.

In Czernowicz, Hahn intensified his philosophical studies, writing to a friend that "last year I almost became faithless to mathematics, seduced by the charms of—philosophy" (Hahn 1906). But it was only in 1921—after years of war service and a professorship in Bonn—that Hahn got his coveted appointment in Vienna and could take steps to resume the philosophical discussions. There was no university philosopher on hand: All three chairs in philosophy happened to be vacant at the time. In particular, Stöhr, successor to Boltzmann and Mach, had died in 1920. Hahn managed to persuade the appointment committee to fill the vacancy with the German, Moritz Schlick, professor at the University of Kiel. Fittingly enough, Schlick was a former physicist, a student of Planck and friend of Einstein (see Schlick, Moritz).

The "Schlick circle," which later became known as the Vienna Circle, met on every second Thursday, during term time, in a small lecture room of the mathematical seminar. The members of the group, who were personally invited by Schlick to attend, were a congenial mixture of philosophers and mathematicians, including Hahn; his sister Olga and her husband, Otto Neurath; Hahn's young colleague Kurt Reidemeister, a professor of geometry; and (later) Hahn's two brightest students, Karl Menger and Kurt Gödel. As Popper later wrote: "It was Hahn who was the founder of the Vienna Circle, and his brother-in-law Neurath who was the organiser. … Schlick was at first, I think, a kind of honorary president … but he became very active" (Popper 1995, 16). Popper went on to state: "What made the Vienna Circle so special, so different from any other philosophical circle was that it was founded not by philosophers but by an important and creative mathematician, who was keenly interested in fundamental problems (also in those belonging to the philosophy of mathematics) and in applications" (ibid). Philipp Frank also designated Hahn as the true founder of the Vienna Circle.

An important part of the discussions in the Circle centered on the theory of knowledge, a topic familiar to Schlick and to Carnap (who joined in 1926), and well in line with the works of Mach and Boltzmann (see Mach, Ernest). Hahn's main contribution to the agenda of the Vienna Circle was his emphasis on the foundations of mathematics. Hahn was not looking for a proof that there exists no contradiction in mathematics, or an explanation for the astonishing efficiency of mathematical tools, or a reduction of mathematical insight to some primordial intuition. What was for him the fundamental problem was the compatibility of mathematics with an empiricist position.

Hahn had encountered foundational problems in mathematics already during his early stay in Göttingen, right after completing his doctoral thesis. He had studied with Hilbert, the foremost advocate of an axiomatization of mathematics, and worked with Zermelo, a highly influential set theorist whose "axiom of choice" aroused fierce debates among mathematicians. Later, Hahn embraced Russell's logicism, the program to reduce mathematics to logic, and engaged in an in-depth study of the *Principia Mathematica* of Russell and Whitehead (see Russell, Bertrand). His sister Olga (who lost her eyesight at the age of twenty-two) had written seminal papers on formal logic. But Hahn himself did not write on mathematical logic. His interest in the foundations of mathematics was of a more philosophical nature, and was an attempt to reconcile Russell with Hume. As Hahn stressed on several occasions, "the only possible way of facing the world seems to me the empiricist position" (Hahn 1980, 31). But since it is unthinkable that an assertion like "two times two is four" is not valid tomorrow, it cannot be based on experience. How, then, "is the empiricist position compatible with the applicability of logic and mathematics to the real world?" (ibid., 32).

Hahn found his answer in a booklet by another Viennese. Following a suggestion by Reidemeister, the Vienna Circle had started reading Ludwig Wittgenstein's *Tractatus Logico-Philosophicus* and spent several semesters discussing it sentence by sentence. Not all members of the Circle were convinced that Wittgenstein had, as he claimed, essentially solved all philosophical questions. But for both Schlick and Hahn, working through the booklet became a key experience. "It was Wittgenstein," he wrote later, "who recognised the tautological character of logic, and who stressed that there exists nothing in the world that corresponds to the so-called logical constants (like 'and,' 'or,' etc.)" (Hahn 1980, 24). Members of the Circle later criticized this view on the grounds that the concept of tautology is not precisely defined except in the realm of first-order logic. But Hahn did not aim at a precise delimitation, and rather used the term to denote any sentence that is true by its logical structure, such as the analytical statement "No object is both red and non-red" (Hahn 1995, 494).

Since mathematics, for a logicist, can be reduced to logic, it also consists of tautologies. Mathematicians often object to the view that their "hard-earned theorems can be dissolved into tautologies" (Hahn 1995, 500). But this, according to Hahn, "overlooks a minor detail, namely the circumstance that we are not omniscient." And indeed, "an omniscient being needs no logic and no mathematics," and "the reason for introducing a symbolic notation which allows to say the same in different ways is that we are not omniscient" (Hahn 1980, 23). Logic, in Hahn's view, "is a set of rules for stating the same in different ways" (ibid., 33). Logic does not deal with the most general properties of objects (this would indeed present insurmountable obstacles to empiricists): "Logic does not deal with any objects at all: it only deals with the way we talk about objects; logic first comes into being by language" (Hahn 1995, 492).

Wittgenstein, at this time, had withdrawn completely from philosophy. After much wooing, he finally condescended to meet with some members of the Circle (not Hahn), on condition that philosophical topics were avoided. But when Hahn invited the celebrated Dutch topologist L. E. J. Brouwer, the founder of the intuitionist movement in the debate on the foundation of mathematics, to give a lecture at the university, Wittgenstein showed up, and started in the after-session to discuss philosophy again. Apparently, there was still something left to say, and Wittgenstein embarked on his second phase, soon leaving Vienna to become, eventually, a professor at the University of Cambridge.

Another member of Brouwer's audience had been Kurt Gödel, a student of Hahn's who first shone in the latter's seminar on the *Principia Mathematica.* That same year, Gödel solved a problem posed by Hilbert as a first step in the latter's program of basing the foundations of mathematics on a formalistic approach. Gödel's proof that first-order logic was complete was published in his doctoral dissertation. In the following year, however, Gödel effectively destroyed Hilbert's program by showing the incompleteness of any consistent mathematical theory rich enough to allow for the natural numbers. Some true statements could not be derived from the axioms and the rules. Hahn praised Gödel and the mathematical importance of his result. The fact that a proof of the consistency of mathematics is impossible—a consequence of Gödel's breakthrough—was taken by Hahn in his stride. In his few philosophical papers, Hahn does not mention Gödel. He seems to have expected the result that "on the basis of present knowledge, an absolute proof of freedom from contradiction is probably unattainable. … For here, as in every sphere of thought, the demand for absolute certainty of knowledge is an exaggerated demand; in no field is such certainty attainable" (Hahn 1980, 121). In the same vein, Hahn appears to have anticipated the independence of the axiom of choice from the axioms of set theory, a result proved only after his death. Hahn wrote: "The question has nothing to do with the nature of reality, as the realists think, or with pure intuition, as the intuitionists think. The question is rather in which sense we decide to use the word 'set'; it is a matter of determining the syntax of that word" (ibid., 118). This approach via analyzing how language is used seems closer to Wittgenstein than to Gödel. Hahn was fifty years old before he wrote his first philosophical essay, a pamphlet named *Occam's Razor.* This and the following philosophical papers—*On Intuition,* for instance, and *What Is Infinity?*—are models of clarity, the outcome of a lifelong concern with popularizing knowledge (Hahn 1980).

Unlike the majority of his colleagues at the university, Hahn was a stalwart member of "Red Vienna," firmly supporting school reforms, free thought, and enlightenment. Some of his fellow members of the Vienna Circle distanced themselves from what they saw as an unseemly involvement in the political quarrels of the day. But after Hahn's death from cancer on July 24, 1934—a year when Austria's political prospects took a distinct turn for the worse—Menger would mourn the demise of this "tireless and effective speaker for progressive causes" (Menger 1994, 215).

Just before the onset of his illness, Hahn had been reading the proofs of Karl Popper's *Logik der Forschung.* "His opinions were as positive as I could only wish [them] to be," wrote Popper (1995, 19) in his last paper. Popper would continue his interactions with the Viennese "Mathematical Colloquium" for the few years until his emigration, but wrote later that "of all the mathematicians at the institute, Hahn was the one who seemed to me the embodiment of mathematical discipline" (Popper 1995, 13).

KARL SIGMUND

References

Hahn, H. (1906), Unpublished letter to Ehrenfest, 26 December 26. Ehrenfest archive, Boerhave Museum, Leiden, Netherlands.

——— (1980), *Empiricism, Logic and Mathematics.* Dordrecht, Holland: Kluwer.

——— (1995), *Introduction to the Collected Works of Hans Hahn.* Edited by L. Schmetterer and K. Sigmund. New York: Springer Verlag.

Menger, K. (1994), *Reminiscences of the Vienna Circle and the Mathematical Colloquium.* Boston: Kluwer Academic Publishers.

Popper, K. (1995), 'Hans Hahn—Reminiscences of a Grateful Student," in L. Schmetterer and K. Sigmund (eds.), *Introduction to the Collected Works of Hans Hahn.* Vienna: Springer Verlag, 11–19.

See also **Logical Empiricism; Neurath, Otto; Schlick, Moritz; Vienna Circle**

NORWOOD RUSSELL HANSON

(17 August 1924–18 April 1967)

After distinguishing himself as a fighter pilot in World War II, Hanson attended the University of Chicago, receiving a B.A. in philosophy in 1946. He then went on to Columbia University, where he received degrees in physics, a B.S. in 1948, and an M.S. in 1949. With the aid of a Fulbright scholarship, Hanson studied philosophy at Oxford and Cambridge, where he also lectured in the philosophy of science. He completed his graduate studies in 1956, earning a D.Phil. from Oxford and a Ph.D. from Cambridge. In 1957, Hanson joined the philosophy department of Indiana University and was the founding chair of Indiana's Graduate Program in the History and Philosophy of Science, the first department of its kind (Hanson 1960; Grau 1999). However, injuries sustained in a plane crash caused Hanson to step down as chair in 1962, and in 1963 he left Indiana for Yale. Hanson died in 1967 in a plane crash on his way to present a paper at Cornell University.

Hanson was a prolific philosopher during his brief career, writing on scientific observation, the role of concepts in accounts of scientific facts and causation, the logic of discovery, the history of discoveries in quantum mechanics and seventeenth-century physics, and the relation between history and the philosophy of science. Hanson represented a fusion of the late Wittgenstein and logical empiricism (see Logical Empiricism). He agreed with Wittgenstein that the meaning of terms, even in science, depends on their use, and he expanded Wittgenstein's account of the conceptual loading of perception to include scientific observation. Hanson, however, shared the logical empiricist view that the function of the philosophy of science is to examine and clarify the conceptual foundations of science.

In a sense, Hanson can be seen as extending the field of conceptual analysis to areas considered off-limits, such as the context of discovery and the conceptualization of perception. Philosophers inspired by logical empiricism often spoke of matters such as observation, factuality, and perhaps causation as fundamental ideas underlying all scientific thought and practice; for Hanson, however, all of these notions can be understood, at a given time, only in terms of the theoretical and notational networks in which they figure. The great revolutions in the history of science were not generally due to observing the world, collecting facts, and finding the causes. Rather, revolutions are made possible by conceptual innovations; after such a conceptual shift, the sense of what the facts are, what has been observed, and what features of phenomena require explanation change as well.

Observation

Hanson's most significant contribution to the philosophy of science was his discussion of observation (see Observation). Hanson argued that observation is "theory-laden": more precisely, in order for perceptual experience to relate to knowledge, experience must already contain some conceptual content. Drawing on Wittgenstein's *Philosophical Investigations* and the findings of psychology on "perceptual sets," Hanson attacked the logical empiricist conception of observation. The logical empiricists generally believed that the edifice of scientific knowledge had basic statements about first-person experience, "protocol sentences," at its foundations, and that these statements were connected to scientific theory via analytic connection rules (see Protocol Sentences). Theories could be confirmed by deriving predictions about observables and then verifying that the appropriate protocol statements were produced in testing the predictions. Among the many complaints Hanson had about this picture of science was that it cannot give an adequate account of scientific controversy and discovery. The deep disputes in the history of science require a better explanation than simply claiming that the disputants were clinging to different interpretations of essentially similar protocol statements.

While other critics of the logical empiricists' conception of observation, such as Feyerabend, focused on the theory-laden character of the terms in observational reports, Hanson's main concern was to show that the *process* of observation is theoretically loaded (see Feyerabend, Paul). Hanson (1961 and 1969) treated vision as illustrative of the general perceptual case. He acknowledged that there is a sense in which seeing is just the stimulation of one's sensory organs or, alternatively, the reception of sense data; however, neither of these senses of seeing are of much epistemological importance. Simply analyzing someone's retinal imprints, or the drawings produced based on sense data, gives one little idea of what is being seen, or what knowledge has been gained through the seeing. In order for seeing to be epistemologically useful, the elements of the visual field must be ordered and categorized with concepts. The logical empiricists explained this by arguing that seeing has two discrete components: acquisition of sense data and interpretation. Hanson objected to this "formula," for he argued that for something to be an interpretation, (i) one must be introspectively aware that one is interpreting, and (ii) there must be a detectible time lag between raw perception and interpretation. Hanson used ambiguous figures from Gestalt psychology to attempt to show that neither of these conditions are met in appreciating figures in an aspect shift: The "interpretation," or the concepts, are already there in the seeing. In order to see something (in the useful sense) as an *X,* one must first have a concept of an *X.* Thus all useful seeing is *seeing as,* and *seeing as* threads experience into knowledge.

The logical empiricist tradition took the reception of sense data to be the paradigmatic case of seeing, because such seeing is incorrigible and provides the foundation for knowledge. Hanson, in contrast, considered the central function of vision to be to provide knowledge about the world, rather than to provide the indubitable foundation of a system of knowledge. Hanson claimed that *seeing that* it is four o'clock or *that* a voltmeter reads 3.5 volts is the sense of seeing of interest in the study of science. Thus, he takes epistemic seeing, or *seeing that,* to be the paradigmatic case of seeing, and asserts that study of the logic of *seeing that* will illuminate the logic of perception generally. To see something as an *X* is to see that, were certain things done to it, other things, which would be expected of *X*s, would follow; more basically, the concept of *X* incorporates our prior knowledge of *X*, such as what *X*s are composed of and what types of interactions *X*s can participate in. To see something as an *X* is to see it in all the connections that the concept of *X* has to other elements of our knowledge.

Hanson's view is open to criticism on the grounds that (i) the statement following *sees that* is always taken to be true in ordinary usage (i.e., *seeing that* is a success verb) and (ii) obsolete concepts (or those that do not apply properly to anything in the world) can be used only for *seeing as,* not *seeing that* (e.g., one cannot now *see that* a bell jar has been saturated by phlogiston). Hanson, however, was very clear that one could be wrong in what one "sees that." He did not take *sees that* to be equivalent to *knows that*; rather, *sees that* is just an indicator of certain (often unconscious) psychological inferences from perception. A person can *see that* the Earth is the center of the solar system, since for Hanson this is just to say that the person's experiences of the Earth are ordered by the concept of a geocentric universe, and thus infers other things in virtue of the experience and the concept. The reason certain conceptual orderings have fallen out of fashion, such as those associated with the geocentric solar system, is that the patterns according to which they order experience render less things intelligible than their successful rivals.

While Hanson thought that observation is theory laden, he seems not to have held that one's theoretical commitments in any sense determine or alter the phenomenology of one's experience. Thus, how one "sees as," or one's conceptual repertoire, does not place absolute constraints on how one will be able to "see as" in the future. However, the production of new conceptual orderings is no trivial, transparent, or easy process, as the theoretical struggles in the history of science attest. One must build new conceptual patterns out of existing frameworks, but it can be extremely difficult to determine which elements of the older frameworks can be transferred and which cannot, and scientists are often blinded by assumptions they have inherited from previous conceptual and notational frameworks.

Hanson's goal in his discussion of observation was not to argue for the subjectivity of science, but rather to clarify the link between perception and knowledge (Hanson 1971). In clarifying this link, Hanson provides a clue to the logic of discovery, for discoveries are achieved through seeing the world differently, which involves appreciating the world through new conceptual arrangements.

Facts

Hanson was also critical of philosophical accounts of facts that attempted to define facts syntactically as phrases following *that*-clauses, arguing that this does little to help make clear the logical and epistemological status of facts. Moreover, faith in the theoretical neutrality of fact-claims obfuscates the effects that language, notation, and idioms have on the way facts are understood. Language and notation provide a sort of template through which facts can be expressed, a pattern through which the world can be understood. Hanson uses ordinary language analysis to indicate the dependence of a fact's significance on the structure of the language in which it is couched. For example, there is a difference between saying "Grass is green" and "Grass greens," for the first formulation assigns merely a property to grass, whereas the second assigns an action; similarly, there is a difference between saying "Electrons are charged" and "Electrons induce electrostatic fields."

Hanson used a historical case to show that while facts inexpressible in a given notation are not impossible to grasp, the practical obstacle such a process involves is very conceptually important for understanding the growth of science. Hanson showed, through a careful analysis of the works of Galileo, Beeckman, and Descartes, how the correct law of free fall was grasped only after a long period of confusion, even though all the requisite data, or "facts," were known from the beginning. All three persisted in thinking of velocity as a direct proportion to the space traversed, rather than (as is correct) a direct proportion of the times. Hanson attributed this apparent obtuseness among geniuses to the geometric notation with which such problems were then treated, which left no room for the expression of a time axis. Spatial properties were more easily measured and represented than temporal ones, and it took the penetrating mind of Galileo to see through this theory-laden factual representation to the correct solution.

Causes

Hanson saw science as primarily a quest for intelligibility, and only secondarily as a search for facts and causes, since these notions are definable only within the organizing conceptual framework. Thus, scientists do not search for succeeding links in the causal chain of nature. According to Hanson, causality is best thought of not as an independent feature of the world, but as a means whereby elements of a theory are bound together inferentially. What is significant, then, about causal language is not what is asserted about objects, but the inferential relations they warrant. Therefore, the adequacy of a proffered explanation cannot be appraised according to some extratheoretical notion of causation, and different scientific programs will differ in terms of what needs to be explained, that is, what causal inferences need be defined. This does not lead to a free-for-all, however, since those conceptual programs that inferentially organize the most phenomena ultimately prevail. For instance, the standard Newtonian line that gravitation requires no explanation was taken as correct until general relativity provided an explanation (see Space-Time); but Newton prevailed over the Cartesians in spite of this inferential omission, since there were so many other inferences he could supply that they could not.

Logic of Discovery

Hanson was also a critic of the logical empiricists' doctrine that discovery is a matter of mere psychology and that philosophical assessment of science should be confined to a logical analysis of the justification of theories. Hanson repeatedly urges that the great conceptual innovations that have fueled science required the genius of a Galileo, Newton, Kepler, or Einstein for their creation, whereas the business of justification is an ordinarily pedestrian affair, better suited to the talents of assistants than geniuses. The production of hypotheses in unsettled domains of inquiry is itself rationally appraisable; there are right and wrong ways of doing it. Revolutionary discoveries are a triumph of reason, and it would be most inaccurate to consider them as such only in retrospect, after they have been justified. Thus, Hanson looked to the history of science in order to adumbrate a set of informal dicta for rational hypothesis creation.

A number of criteria can be used to assess the reasonableness of suggesting a hypothesis. The hypothesis should be consistent with background knowledge, show some capacity to explain the problem at hand, offer testable consequences and an account of the constraints on testing, and have some plausibility (in light of the rest of knowledge). Hanson attempted to outline a set of informal strategies for rational hypothesis production from problematic situations. Hypothesis creation from enumerative induction, abductive reasoning, authority, or symmetry considerations are all approaches that can be used normatively. In addition, since perception and data interpretation are theory-laden activities, they aid in making inferences and offering hypotheses. Previous knowledge shapes new expectations, thus suggesting explanations and inferences about the future. Given a theoretically loaded observation, certain hypotheses, which also make reference to these same concepts, can be reasonably suggested.

Hanson was able to provide a set of only vague criteria for rational hypothesis suggestion. His account was never complete enough to determine whether a particular hypothesis suggestion was rational without some reference to the historical consequences that followed from the suggestion.

Jordi Cat
Matthew Lund

References

Grau, K. T. (1999), "Force and Nature: The Department of History and Philosophy of Science at Indiana University, 1960–1998," *Isis* 90: S295–S318.

Hanson, N. R. (1960), "New Discipline Joins Science to Arts," in *Arts and Sciences: The Review* (published by the Indiana University Alumni Association) 58: 3–11.

——— (1961), *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*, Cambridge: Cambridge University Press.

Hanson, N. R. (1963), *The Concept of the Positron: A Philosophical Analysis*. Cambridge: Cambridge University Press.

Hanson, N. R. (1969), *Perception and Discovery: An Introduction to Scientific Inquiry*. Edited by W. C. Humphreys. San Francisco: Freeman, Cooper and Co.

Hanson, N. R. (1971), *What I Do Not Believe and Other Essays*. Edited by S. Toulmin and H. Woolf. Dordrecht, Holland: Reidel.

CARL GUSTAV HEMPEL

(8 January 1905–9 November 1997)

Carl Gustav Hempel was born in and educated primarily in Germany. He studied mathematics, physics, and philosophy at the universities of Berlin, Gottingen, Heidelberg, and Vienna. He completed most of his doctoral thesis on analyses of probability under Reichenbach but was compelled to find an alternative advisor to complete the project because Reichenbach was dismissed from his position in 1933 when Hitler and the National Socialist Party came to power (see Reichenbach, Hans).

Hempel was opposed to the National Socialist Party and moved to Brussels in 1934 and then to Chicago in 1937. He taught at City College and Queens College in New York from 1939 until 1948, and Yale from 1948 to 1955. Most of his subsequent career was spent at Princeton University from 1955 until mandatory retirement in 1973, and after a two-year sojourn at the University of Pittsburgh, he returned and resided in Princeton, New Jersey, until his death in 1997.

Hempel was one of the youngest members of the Berlin Circle and was in close contact with members of the Vienna Circle (see Vienna Circle; Logical Empiricism). Because of his longevity, his career spanned the rise and decline of the logical empiricist movement. Because of his emphasis on pursuit of truth and clarity, his views underwent considerable change during his career, especially after 1964, when Thomas Kuhn became a colleague in the History and Philosophy of Science program at Princeton (see also Kuhn, Thomas). Although most philosophers who knew Hempel's and Kuhn's views expected that they would be highly antagonistic because their early views were rather divergent, they became close friends and had significant influences on each other.

Hempel contributed to numerous areas of philosophy of science, most notably to explanation, confirmation, analyses of theory and observation, and questions of scientific methodology, each of which is addressed below. However, it is important to note that Hempel also made an important contribution by serving as a personal example of the possibility of combining an unwavering pursuit of truth and clarity with great kindness toward and encouragement of his fellow philosophers. (See Richard Jeffrey's "Introduction" in Hempel 2000 for further details).

Explanation

Hempel's proposals for analyzing scientific explanation were among the most fruitful part of his research (Hempel 1965). Roughly he suggested that explanation involves a relation between a set of sentences (including at least one law) and a statement to be explained. The relations were of three sorts: deductive, inductive, and deductive-statistical (see Prediction). Respectively, the explaining statements provide a deductive argument, an inductive argument, and a deductive argument for a probabilistic conclusion.

These suggestions made depend primarily on syntactic relations. Numerous criticisms and counterexamples were proposed to these suggestions. While Hempel's suggestions are generally discredited, most of the decades after his work were devoted to criticism, defense, and modification of his views, so that it is fair to say that his ideas shaped the field for a significant period of time. The alternative views incorporate psychological, social, and pragmatic factors, as well as syntactic ones, and can be seen as developments that add further factors to his syntactic approach. (For a more detailed discussion, see Explanation; Inductive Logic).

Confirmation and the Raven Paradox

The puzzle known sometimes as *Hempel's paradox* and sometimes as the Raven paradox attracted less attention initially than some of Hempel's later work, but it has proved to be an enduring topic. It was first sketched by Hempel in 1937, but the first full development came later (Hempel 1943), and his *Studies in the Logic of Confirmation* (Hempel 1945) is more accessible both in terms of content and physically as reprinted still later (Hempel 1965). Most writers (Giere 1970; Good 1967) do not see the Raven paradox as a major puzzle because they are each convinced that they have solved it; yet there are numerous conflicting solutions with no consensus (see Confirmation Theory). The puzzle is easily described.

It seems plausible that the universal generalization "All ravens are black" is confirmed by reports of black ravens. (The general statement of this principle, that a universal generalization is confirmed by a report of a positive instance, is often called *Nicod's criterion,* though this is somewhat misleading historically, since Nicod's (1930, 219) original criterion made this a *necessary* as well as a *sufficient* condition. It is also highly plausible that whatever confirms a statement confirms any statement that is logically equivalent to the first. After all, if two statements are logically equivalent, then they are guaranteed to say the same thing about the world. (For further discussion of these principles, their justification, and history, see Hempel 1945.) The standard formalization of "All ravens are black" renders it as the universal generalization of a conditional, that is, $(\forall x)(Rx \rightarrow Bx)$ (where Rx is "x is a raven" and Bx is "x is black"). By the standard formalization, "All non-black things are non-ravens" is to be formalized as $(\forall x)(\neg Bx \rightarrow \neg Rx)$. By Nicod's criterion, a report of a non-black non-raven, for example, a white shoe, confirms that all non-black things are non-ravens. But by the logical equivalence condition, since $(\forall x)(Rx \rightarrow Bx)$ and $(\forall x)(\neg Bx \rightarrow \neg Rx)$ are equivalent, observation of a white shoe confirms that all ravens are black.

Many people find this last conclusion unacceptable, but some solutions to the paradox attempt to make it palatable. Hempel propounded the paradox in the context of attempting to develop principles for a qualitative theory of confirmation, and he accepted the conclusion in spite of its counterintuitiveness. Later writers have attempted to make the conclusion more palatable by embedding the argument in a quantitative context and saying that although the report of a white shoe confirms that all ravens are black, it does so only to a very minute degree, and thus one has the illusion of irrelevance (see Induction, Problem of).

More precisely, if one accepts a Bayesian (see Bayesianism) account of confirmation, then a hypothesis H is confirmed by evidence E just in case the evidence increases the probability of H. There are various exact formulations of confirmation, but one rather natural one is that the degree of confirmation of a hypothesis H by evidence E is the extent to which the probability of H given E, $P(H \mid E)$, exceeds the prior probability of hypothesis $P(H)$. Thus one plausible measure is $P(H \mid E) - P(H)$.

Let H be the hypothesis that all ravens (R) are black (B) and consider the two evidential statements, the statement, E_1, that the observed object is a black raven, that is, $Ra \wedge Ba$, and E_2, the statement that b is a non-black non-raven, that is, $\neg Bb \wedge \neg Rb$. Since $P(H \mid E) = P(H \wedge E)/P(E) = [P(E \mid H) \times P(H)] \mid \mathrm{P(E)}$, an expression for the confirmation measure is given by $\{P(E \mid H) \times P(H)]/\mathrm{P(E)}\} - P(H)$. Substituting the particular positive evidence $Ra \wedge Ba$ for E produces $\{[P(Ra \wedge Ba \mid H) \times P(H)]/P(Ra \wedge Ba)\} - P(H)$.

Using the definition of conditional probability again, $[P(Ra \wedge BaH) \times P(H)] = P(Ba \mid (Ra \wedge H)) \times P(Ra \mid H) \times P(H) = P(Ra) \times P(H)$, on the assumptions that $P(Ba \mid (Ra \wedge H)) = 1$ and $P(Ra \mid H) = P(Ra)$. The first is a theorem of probability, the second an assumption about the irrelevance of H to whether something is a raven (given no other information). Since $P(Ba \wedge Ra) = P(Ba \mid Ra) \times P(Ra)$, the expression for confirmation turns into

$\{P(H)/P(Ba \mid Ra)\} - P(H)$. If it is taken as a background assumption that black objects are relatively uncommon and that ravenhood is antecedently thought irrelevant in this situation, the term $P(Ba \mid Ra)$ will be small and the confirmation large.

Working through the parallel calculation for the negative evidence, $\neg Ba \wedge \neg Ra$, leads to the parallel expression $\{P(H)/P(\neg Ba \mid \neg Ra)\} - P(H)$. But since the relative frequency of non-black objects is presumed high, $P(\neg Ba \mid \neg Ra)$ is close to 1 and the confirmation is minimal, though not presumably not zero.

This analysis vindicates Hempel's qualitative claim that the non-black non-raven does confirm the hypothesis H, but is supposed to allay the sense of paradox by showing that the degree of confirmation is extremely small and is considerably less than the confirmation provided by a black raven. However, the derivation depends on some assumptions about the probabilities, that is, $P(Ra \mid H) = P(Ra)$ and that $P(\neg Ra)$ is very close to 1, and these assumptions can be questioned. (See Vranas 2004 for further discussion and references.)

A second family of proposed solutions to the paradox appeals to Goodman's (1983) conception of projectiblity and his arguments that generalizations are confirmed by positive instances only if the predicates they contain are projectible. Projectibility is relative to a language community and a history of actual predictions. Roughly, a predicate is projectible if it has been used successfully in making predictions by the community in the past. New predicates may be projected if, for example, they are coextensive with prior projectible predicates. Goodman presents a complicated set of rules for projectibility (Goodman 1983, ch. 4), but what is essential for present purposes is that the class of projectible predicates is not closed under negation. Thus the presumed projectibility of "raven" and "black" does not transfer to "non-raven" and "non-black" (see Induction, Problem of).

A third family of solutions changes the subject from simple confirmation of a hypothesis by evidence, and substitutes the question whether a particular piece of evidence selectively confirms H from among a set of competing hypotheses. In this context, Hempel's reasoning shows that being a positive instance of a logically equivalent hypothesis does not generally provide selective confirmation. For example, if the alternative hypotheses to "All ravens are black" are "All ravens are white," "All ravens are green," etc., then a white shoe instantiates all of these equally. The only evidence that would selectively instantiate "All ravens are black" is a black raven.

There are at least two suggestions of how the alternatives for selective confirmation are specified. According to one alternative, specification would be in linguistic terms, so that the contrast would always use expressions in the same semantic field as the term that was the focus. The other alternative is to see the competing hypotheses as those that would be seriously entertained by members of the scientific community that investigates the relevant domain. Thus the first version makes the selective confirmation relation a function whose arguments include the language, and the second makes the function depend on the community as well as the evidential statement and the candidate for confirmation.

Arguments for one as opposed to the other of these alternatives are probably not compelling when the sentences are qualitative statements such as the Raven hypotheses. But for quantitative hypotheses, it would appear that the community approach is more plausible, since it is difficult to envision linguistic grounds for preferring some equations over others.

Yet another distinct approach to solving the paradox questions a basic assumption about the logical form of the hypotheses. Consider for example, the theory of conditionals espoused by McDermott (1996), in which a conditional is true if the antecedent and consequent are both true, false if the antecedent is true and the consequent false, and has no truth value otherwise. If this conditional is symbolized by $\Rightarrow$, so that "All ravens are black" is translated as $(\forall x)(Rx \Rightarrow Bx)$ and "All non-black things are non-ravens" as $(\forall x)(\neg Bx \Rightarrow \neg Rx)$, the two sentences are not equivalent and the paradox is resolved.

In summary, Hempel's Raven paradox continues to command the attention of philosophers of science. There are four major kinds of responses to the puzzle. The first response accepts the puzzling conclusion and attempts to explain it away. The second—projectibility—appeals to a combination of language and community practices to disarm the puzzle. The third appeals to a slightly different relation and to the role of the scientific community to change the subject. And the fourth claims to solve the problem by placing the blame on the choice of the material conditional as a way of representing the hypothesis.

Problems and Changes

Hempel's earliest work dealt with truth in science and mathematics (Hempel 2000, Essays 1–5), but his focus shifted fairly soon to more accessible and less metaphysical issues such as confirmation and explanation. Truth and realism deal with the relation between sentences and the world, whereas confirmation and explanation, on the surface at least, deal with relations among sentences. One of his most valuable later papers was "Problems and Changes in the Empiricist Criterion of Meaning" (reprinted in Hempel 1965), which chronicles the shift over several decades in the attempts to make a sharp demarcation between statements that are cognitively significant and those that are not. This piece lacks the rhetorical flourishes and the metaphoric ending of Quine's (1980) *Two Dogmas of Empiricism* (see Quine, Willard Van Orman) but is perhaps a more telling argument against the attempt to make a demarcation of the boundary of the cognitively significant. In this, as in numerous other areas, Hempel's views shifted away from attempts at purely syntactic or even semantic characterizations and toward conceptions that included social, psychological, and historical elements. (For a more detailed discussion, see Cognitive Significance).

Theories and Observation

Hempel's views on theory and observation evolved as he continued to ponder basic questions of confirmation and explanation. In evaluating the extent of this evolution, it is important to read closely the formulations that Hempel provides. For example, in *The Theoretician's Dilemma,* he states: "Formally, a scientific theory may be considered as a set of sentences expressed in terms of a specific vocabulary" (Hempel [1958] 1965, 182–183). Logical empiricists are often criticized for *identifying* theories with sets of sentences in first-order logic, but note that here Hempel is not arguing for identification, but proposing that from a particular perspective, a theory may be *considered* as a set of sentences. The distinction between formal questions about artificial languages and the related questions about natural languages was clear in Hempel's mind from very early. For example, in his one venture into the topic of vagueness and logic (Hempel 1939), he argues that since vagueness is a phenomenon of natural language, it does not provide any leverage for an argument for relinquishing two-valued formal logic.

The theoretician's dilemma is the following: Divide the vocabulary of a theory into two portions, that which is observation and that which is theoretical. If the sole function of the theory is to provide derivations of observational statements from observational statements by means of the intermediary use of theoretical statements and vocabulary, then it can be shown that the theoretical statements are dispensable. It will be useful to expand on the terms of the dilemma.

Hempel's construal of "observational" at this stage is to be distinguished from one of the earlier logical empiricist notions of "observational," which was to provide an absolute epistemological foundation for science (see Logical Empiricism; Observation). Rather, the relevant criteria for observationality is that intersubjective agreement is obtainable for the statement in question: "The observational data. . .are. . .couched in terms whose applicability in a given situation different individuals can ascertain with high agreement, by means of direct observation" (Hempel [1958] 1965, 179). Notice that this characterization of the observational terms is not syntactic or semantic, but pragmatic, in the sense of the scientific community agreeing on applicability of the terms (see Observation). A closely related conception of an observation sentence is developed by Quine (1960); it too is relativized to a time and a linguistic community, as well as other parameters (see Quine, Willard Van).

Shortly thereafter, in *Philosophy of Natural Science,* Hempel (1966) took two further steps in modifying his view. First, he abandoned the characterization of the term opposed to "theoretical" as "observational," and instead proposes to distinguish between what is an understood and agreed-upon antecedent to a particular theory and that which is not. This is, of course, a distinction that shifts over time, as what were new theories become accepted and incorporated as part of the "antecedently understood." The second shift was that he recognized that being antecedently understood is not a characteristic of terms *simpliciter,* but of statements including those terms. A term may be antecedently understood within a particular range of application, but not outside that range:

> For example, a characterization of the concept of temperature by reference to the readings of a mercury thermometer affords no general definition of temperature; it assigns no temperature below the freezing point or above the boiling point of mercury. (Hempel 1966, 79)

This illustrates that a term such as "temperature" may be antecedently understood in the context of assigning temperatures in a certain range to liquids, but be highly theoretical when applied to other objects or at much higher or lower temperatures.

Even with the change from "observation" to "antecedently understood," the theoretician's dilemma can still be stated. Consider the theory and the antecedently understood expressions as part of a formal language. Divide the sentences into those that are antecedently understood and those that are not, and consider what the function might be for the sentences that are not antecedently understood. If their role is to provide appropriate deductive connections between antecedently understood sentences, that is, predictions in the broad sense that include past statements, then it can be shown on formal grounds (Hempel [1958] 1965) that an alternative formal structure is available that does not utilize the other sentences at all.

Hempel's conclusion in 1958 was not that the theoretical is dispensable on these grounds because there are nondeductive characteristics of theoretical systematizations that are of great importance: "If it is recognized that a satisfactory theory should provide possibilities also for inductive explanatory and predictive use and that it should achieve systematic economy and heuristic fertility, then it is clear that theoretical formulations cannot be replaced by expressions in terms of observables only" (Hempel [1958] 1965, 222).

Although he came to recognize the importance of nonformal characteristics of scientific theories, Hempel never entirely lost sight of the usefulness of axiomatization and formalization for some purposes. As the commentator on an exchange between Suppes and Kuhn on whether axiomatization is valuable, Hempel clearly favored a compromise position. On the one hand, "axiomatization of theories can be of value for certain philosophical or scientific purposes," but on the other hand:

> Professor Kuhn's paper is highly relevant, for it explores ways in which the requisite agreement in the understanding and the use of scientific terms may be attained by the members of the scientific community without reliance on, or even availability of, explicitly formulated criteria of application. (Hempel 1977, 257)

Later Work

Hempel's later thought continued to be innovative and influential, but while his later conclusions were also clear, they were more tentative and less definite. He had recognized from both his own work and the criticisms of others, especially Kuhn, that the logical empiricist tradition that emphasized syntactic, and to a lesser extent semantic, relations among sentences omitted a great deal of importance for the purpose of understanding how science develops and theories are evaluated, and for assessing rationality. However, his drive for clarity and precision left him dissatisfied with the formulations provided by Kuhn and others in the more pragmatic and historical traditions that emerged from the 1950s on.

Hempel (1983a, and 1983b) provides examples of his working through the issues concerning theory choice and rationality, but the most thorough is *Scientific Rationality* (Hempel [1979] 2001). He begins by noting the goal held by analytic empiricism of formulating explicit, logically precise criteria of rationality for the formulation, testing, and evaluation of scientific claims (358). However, the historical quest for such criteria has not been successful, and the arguments of Kuhn and others seem persuasive that the quest cannot be fulfilled. Scientific inquiry, at its best, does involve shared preferences that shape theory evaluation, such as a preference for quantitative theories "whose predictions show a close fit with experimental findings; for theories covering a wide variety of phenomena; for theories that correctly predict novel phenomena; for fruitful theories; for simple theories rather than complex ones" (359).

However, these preferences do not generally suffice for the unambiguous selection among competing theories. Unless such crucial terms as "close fit," "wide variety," "fruitful," and "simple" can be explicated clearly and rigorously, there remains room for disagreement in theory choice. Hempel remained optimistic that some progress could be made on these questions in specific contexts but was persuaded by the arguments from history of science and the logical problems that no general solution would be forthcoming.

Although Hempel's views evolved considerably and in many regards moved toward convergence with those of Kuhn, the two continued to differ on some important issues, including the locus of rationality. For Kuhn the rationality of theory choice resides in the relevant scientific community and is a holistic property of that group. Hempel criticizes Kuhn's analogy between evolutionary change and scientific change because evolutionary change selects from among more or less randomly produced variations. In contrast, Hempel believed that the rationality of science is indicated by the fact that later theories are superior to earlier ones, because they have been consciously designed to be better, and this is explained by the goal-directed character of scientific change at the level of individual scientists (Hempel [1979] 2001).

Richard E. Grandy

References

Giere, R. (1970), "An Orthodox Statistical Resolution of the Paradox of Confirmation," *Philosophy of Science* 37: 354–362.

Good, I. J. (1967), "The White Shoe Is a Red Herring," *British Journal for the Philosophy of Science* 17: 322.

Goodman, Nelson (1983), *Fact, Fiction and Forecast* (4th ed.). Cambridge, MA: Harvard University Press.

Hempel, C. G. (1939), "Vagueness and Logic," *Philosophy of Science* 6: 163–180.

——— (1943), "A Purely Syntactical Definition of Confirmation," *The Journal of Symbolic Logic* 8: 122–143.

——— (1945), "Studies in the Logic of Confirmation," *Mind* 54: 1–26 and 97–121.

——— (1952), *Fundamentals of Concept Formation in Empirical Science: International Encyclopedia of Unified Science* (Vol. II, no. 7). Chicago: University of Chicago Press.

——— ([1958] 1965), "The Theoretician's Dilemma: A Study in the Logic of Theory Construction," in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* New York: Free Press, 173–226.

——— (1965), *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

——— (1966), *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall.

——— (1977), "Formulation and Formalization of Scientific Theories," in Suppe, F. (ed.), *The Structure of Scientific Theories*. Urbana: University of Illinois Press, 245–254.

——— ([1979] 2001), "Scientific Rationality: Normative Versus Descriptive Construals," in James Fetzer (ed.), *The Philosophy of Carl G. Hempel*. Oxford: Oxford University Press, 357–371.

——— (1983a), "Valuation and Objectivity in Science," in R. S. Cohen and L. Laudan (eds.), *Physics, Philosophy and Psychoanalysis in Honor of Adolf Grunbaum*. Dordrecht, Holland: D. Reidel Publishing, 73–100.

——— (1983b), "Kuhn and Salmon on Rationality and Theory Choice," *Journal of Philosophy* 80: 570–572.

——— (1988a), "Limits of a Deductive Construal of the Function of Scientific Theories," in Edna Ullman-Margalit (ed.), *Science in Reflection*. The Israel Colloquium, vol. 3. Dordrecht, Holland: Kluwer Academic Publishers, 1–15.

——— (2000), *Selected Philosophical Essays*. Edited by Richard Jeffrey. Cambridge: Cambridge University Press.

Maher, P. (1999), "Inductive Logic and the Ravens Paradox," *Philosophy of Science* 66: 50–70.

McDermott, M. (1996), "On the Truth Conditions of Certain 'If ' Sentences," *PhilosophicalReview* 105: 1–38.

Nicod, J. (1930), *Foundations of Geometry and Induction*. Translated by N. Weiner. London: Routledge and Kegan Paul.

Quine, Willard Van (1960), *Word and Object*. Cambridge, MA: MIT Press.

——— (1980) "Two Dogmas of Empiricism," in *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 20–46.

Vranas, P. (2004), "Hempel's Raven Paradox: A Lacuna in the Standard Bayesian Solution" *British Journal for the Philosophy of Science* 55: 545–560.

See also **Carnap, Rudolf; Cognitive Significance; Confirmation; Explanation; Induction, Problem of; Inductive Logic; Kuhn, Thomas; Logical Empiricism; Observation; Prediction; Quine, Willard Van; Reichenbach, Hans; Verifiability; Vienna Circle**

HERITABILITY

Offspring resemble their parents. This simple observation of family resemblances is one of the oldest conceptions of inheritance. It is also the foundation of the scientific study of heredity. For centuries ideas about heredity were based on this simple qualitative measurement. As might be expected, explanations of this observation changed in the course of history, and these changing explanations reflected the prevailing (scientific) attitudes within historical periods. For Aristotle the resemblance between parents and offspring was based on the fusion of male and female "fluids" and the subsequent action of the four causes (*materialis, efficiens, formalis, finalis*), with the female providing the material cause and the male providing the semen, "that which generates," the active stimulus for the developmental dynamics. For Aristotle, development was epigenetic: The new organism was not preformed in any of the parental contributions; rather it realized its own potential in the dynamic process of development. Heredity, in Aristotle's conception, was a consequence of generation (Aristotle 1979).

During the seventeenth and eighteenth centuries, two competing positions were put forward to account for heredity and generation: preformationism and epigenesis. Each of these positions progressed through several versions that reflected both new empirical observations and changing theoretical assumptions. Preformationists, such as Malebranche, Malpighi, and Bonnet, based their argument both on new microscopic observations that showed that semen and eggs had internal structures and on a rejection of ideas of spontaneous generation. Another criticism was the lack of a satisfactory mechanism that would explain generation epigenetically. Proponents of epigenesis, such as Harvey, Gassendi, and Wolff, also claimed empirical support (especially based upon observations of chick embryos). They emphasized several phenomena that could not easily be explained within a preformationist framework, such as hybridization, regeneration, or the existence of so-called monsters (teratology). Even though they did not resolve the problem of inheritance, these seventeenth- and eighteenth-century debates brought the problem of generation and heredity within the scientific focus. It soon became clear that heredity and generation were linked, even though it was not yet obvious what mechanism could account for each (see also Roe 1981; Pinto-Correia 1997; Maienschein 2004).

Ideas about the variability and transformation of species, which emerged in the late eighteenth and early ninenteenth centuries, complicated things even further. While early approaches were mostly concerned with transformations between forms (Goethe, Lamarck, Oken) or the correlations between embryological transformations and morphological complexity (Meckel, Serres, von Baer), attention soon shifted to variation at the subspecies level. This shift first happened in the context of animal and plant breeding programs, the study of geographic variation (fostered by the consolidation of colonial power and the increase in global trade), and the emerging science of anthropology, with its focus on the concept of race. In the nineteenth century, an experimental approach to the problem of heredity emerged, as well as conceptual transformations of the older notion of generation. The latter can be seen most prominently in the work of Darwin (see Evolution).

Darwin's theory of the transmutation of species was predicated on the existence of variation within populations, the action of natural selection on these variants, and a principle of heredity that would guarantee that the offspring of these selected variants would also bear the same traits that enabled the survival and reproductive success of their parents (see Natural Selection). The often repeated formal requirement of Darwin's theory of natural selection consists of phenotypic variation within a population correlated with a corresponding variation in fitness that is also heritable; in other words, offspring share the same traits that helped their parents succeed in the "struggle for existence" (Lewontin 1974a). The essence of Darwin's theory of natural selection is thus based on strictly phenotypic observations. It was, at least initially, also largely a qualitative theory, based on common sense and supported by an overwhelming body of empirical observations.

In the years after the publication of *On the Origin of Species,* the nature of biological variation was the subject of intense debates. Was variation primarily continuous, as in many quantitative characteristics, such as height, or was it primarily discontinuous and discrete, as in qualitative characteristics, such as many color variants? This was an important question, as the views about the nature of biological variation corresponded to different ideas about the mechanisms of evolutionary change (see Population Genetics). Evolution, and the origin of new variants and species, was considered to have either followed the gradual path that Darwin proposed or to have happened by means of larger changes. Some even thought that different mechanisms accounted for the gradual adaptation to changing environmental conditions and for the discontinuous origin of new species.

Two problems that were left unanswered in Darwin's theory are of special interest for the discussion of the problem of heritability: (1) How can the qualitative observations of heredity by Darwin and others be made quantitative and therefore predictive? and (2) What is the material basis of heredity? The latter problem initiated a century-long research program that led from Mendel's experiments (and the factors that he postulated would represent each of the discrete variants in them), to Johannsen's distinction between genotype and phenotype and the associated concept of a pure line, to Boveri and Sutton's chromosomal theory of inheritance, to Morgan's gene maps, and finally to the discovery of the double-helical structure of DNA (see Genetics). Though this line of research (eventually) elucidated the molecular basis of heredity, it contributed very little, with the exception of the genotype/phenotype distinction, to the second problem of establishing a quantitative theory of inheritance and natural selection. For this a different approach was needed, one that was rooted in the simultaneous development of statistics.

It was Darwin's cousin, Galton, who first compared the properties of the phenotypes of parents with those of their children as well as with the rest of a population. He found that the mean value of all the offspring tended to be closer to the overall population mean than to the mean value of their parents. The analysis of this phenomenon, which he called regression to the mean, initiated the statistical analysis of the problem of inheritance. This was a separate approach to the study of inheritance, one that was focused on statistical correlations rather than on material entities. As Provine (1971) has shown, after the turn of the twentieth century, these questions became the foundation not only of theoretical population genetics, but also of quantitative genetics, which was more closely allied with traditional efforts of animal and plant breeding. It was in this context that the concept of heritability was first formulated.

Current Definitions and Problems

Today the concept of heritability is at the heart of the discipline of quantitative genetics and is increasingly also employed in medical contexts, where it is used to assess the probability for the occurrence of certain genetic conditions. Probably the best known, and also the most controversial, applications of the concept of heritability have been attempts to quantify the genetic component of the observed variance in individual values of IQ (e.g., Herrnstein and Murray 1994; Fraser 1995; Jacoby and Glauberman 1995). These debates, especially the more popular discussions, often confused several crucial components of heritability (such as broad- and narrow-sense heritability) and also unjustly equated heritability with a specific form of genetic causation. It is therefore crucial to distinguish several important dimensions and assumptions of the statistical concept of heritability as it is currently used in genetics.

In quantitative genetics, there are two definitions of heritability, broad-sense heritability and narrow-sense heritability. The former is defined simply as the ratio of the total genetic variance, V_G, and the phenotypic variance, V_P. In the case of broad-sense heritability, one is interested in quantifying the total genetic contribution, including all dominance and interaction effects, to the phenotypic variance and to distinguish those from the environmental contributions. This measurement then provides a broad estimate of the degree of genetic determination. In the simple case, one assumes that the total phenotypic variance is simply the sum of the genetic and environmental variance ($V_P = V_G + V_E$). However, this is an idealization, as it assumes that there is no variation in the interaction between a genotype and the environment, that is, that any difference in the environment has the same effect on all the different phenotypes. As this is extremely unrealistic, one should include another variance term that accounts for the variance in the genotype/environment interaction, V_{GE}. Thus we have $V_P = V_G + V_E + V_{GE}$. (Even this is an approximate relation. For a full quantitative treatment, see Sarkar 1998, Ch. 4.)

In sexually reproducing diploid species, there is the additional problem that the genotype is not passed on directly to the next generation, only gametes are. Thus, from an evolutionary point of view, the total genetic variance is not what helps an understanding of the dynamics of natural selection (see Natural Selection). The fraction of the total genetic variance V_G that is relevant for the evolutionary consequences of natural selection is called the additive genetic variance V_A. Fisher (1918) first introduced the concept of the additive effect of an allele in order to account for the additive genetic variance. He first analyzed a one-locus two-allele case, with *A* and *a* as the two alleles. If the substitution from the *aa* homozygote to the *aA* heterozygote genotype produces the same phenotypic effect as the second substitution of the *a* allele (from the *aA* to the *AA* genotype), then all of the genetic variance at that locus is considered additive. These ideal cases are, of course, exceptionally rare. In actuality, one has to also account for a within-locus deviation from additivity (the so-called dominance deviation), as well as for nonadditive effects between different loci (the so-called epistasis effect) that contribute to a given phenotype. The total genetic variance is thus a combination of three different genetic variance components, the *additive, dominance,* and *epistatic* (interaction):

$$V_G = V_A + V_D + V_E.$$

Based on this decomposition of the total genetic variance V_G, the narrow-sense heritability is then defined as the fraction of the additive genetic variance and the phenotypic variance:

$$h^2 = \frac{V_A}{V_P}.$$

The open question thus is, How can one estimate the additive genetic variance in order to calculate the heritability of a trait, as it is not usually possible to measure the additive effects of alleles directly? In the context of quantitative genetics, the additive genetic variance is also given by the variance of the breeding values of the genotypes within a population (Falconer and Mackay 1996). The breeding value of a genotype is defined by the mean genotypic value of its offspring. It therefore can be measured. The idea behind the concept of a breeding value is similar to Galton's regression analysis between midparent and midoffspring values. It provides a measure of how much of the phenotypic variance of the parents is passed on to their offspring. The slope of the regression line between midparent and midoffspring values is therefore another measure of heritability.

Heritability is clearly a central concept in evolutionary and quantitative genetics. It is therefore important to be aware of its assumptions and limitations. There are two main problems of special relevance for philosophers of science: the first is related to causal inferences of genetic determination is based on heritability estimates and the second related to the consequences of epistatic interactions. As has already been seen, heritability is a statistical concept, defined as a ratio between variances or as a regression coefficient. It does not imply a specific model of genetic or environmental causation, and therefore no such model or any specific interpretation of genetic causality can be inferred from a particular value of heritability. In order to establish support for a specific interpretation of genetic causality from the type of variance analysis that is part of heritability assessments, one would have to develop a rather rigorous experimental design (Lewontin 1974b; Feldman and Lewontin 1975). While this is possible in certain breeding experiments with Drosophila (see Falconer and Mackay 1996), these conditions are almost never realized in studies involving humans. The fact that there is a widespread tendency to use heritability values in support of a genetic etiology of human conditions and diseases thus points more to the existence of an underlying genetic ideology than to a well-supported understanding of the role of genes in human disease (Laubichler and Sarkar 2002).

Besides the problems related to causal inference from statistical and population-dependent values, nonlinear or epistatic interactions between genes also complicate the interpretation of heritability. The consequences of epistatic interactions for heritability are also related to the *unit of selection problem* (Wimsatt 1981; Lloyd 1988). In short, recent studies have shown that (1) in all cases of multilevel selection, there will be a certain amount of the total additive genetic variance that will be a consequence of additive effects of alleles at individual loci (Sarkar 1994) and (2) based on a physiological definition of epistasis as the effect of a gene substitution at one locus on a subsequent gene substitution at another locus (Cheverud and Routman 1995; Wagner, Laubichler, and Bagheri-Chaichian 1998), it can be shown that in the case of epistasis between loci, there will (a) always be an epistatic component to the additive genetic variance and (b) under certain conditions, such as a population that is far away from its equilibrium point, there will be irreducible higher-order additive effects that can be attributed to sets of interacting genes (or gene complexes) rather than individual genes. The latter is important for considerations of heritability, as these higher-order additive effects will also contribute to the total additive genetic variance. Consequently, there will also be covariance terms between the additive effects of individual alleles and the irreducible effects of interacting gene complexes. As these covariance terms can be negative under certain circumstances, any value of heritability that is derived from an estimate of the additive genetic variance based solely on the additive effects of individual alleles can thus potentially overestimate the actual value of heritability. As there are only a small number of diseases that are caused by a single gene, these recent studies offer some corrective to the often unrealistically high estimates of heritability for genetic disorders reported in the medical literature.

Today, the concept of heritability continues to be central to quantitative, evolutionary, and medical genetics. It represents a culmination of a century-long quest to quantify and understand the consequences and phenomena of inheritance. However, there continues to be a discordance between the technical interpretation of heritability and the many roles the concept plays in medical, popular, and philosophical discourses.

Manfred D. Laubichler

References

Aristotle (1979), *Generation of Animals*. Cambridge, MA: Harvard University Press.

Cheverud, J. M., and E. J. Routman (1995), "Epistasis and Its Contribution to Genetic Variance Components," *Genetics* 130: 1455–1461.

Falconer, D. S., and T. F. C. Mackay (1996), *Introduction to Quantitative Genetics*. London: Longmans.

Feldman, M. W., and R. Lewontin (1975), "The Heritability Hang-Up," *Science* 190: 1163–1168.

Fisher, R. A. (1918), "The Correlation Between Relatives and the Supposition of Mendelian Inheritance," *Transactions of the Royal Society, Edinburgh* 52: 399–433.

Fraser, S. (1995), *The Bell Curve Culture Wars: Race, Intelligence, and the Future of America.* New York: BasicBooks.

Herrnstein, R. J., and C. A. Murray (1994), *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Jacoby, R., N. Glauberman (eds.) (1995), *The Bell Curve Debate: History, Documents, Opinions*. New York: Times Books.

Laubichler, M. D., and S. Sarkar (2002), "Flies, Genes, and Brains: Oskar Vogt, Nicolai Timofeeff-Ressovsky, and the Origin of the Concepts of Penetrance and Expressivity in Classical Genetics," in L. Parker and R. Ankeny (eds.), *Medical Genetics. Conceptual Foundations and Classical Questions*. Dordrecht, Netherlands: Kluwer, 63–85.

Lewontin, R. (1974a), *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

——— (1974b), "The Analysis of Variance and the Analysis of Causes," *American Journal of Human Genetics* 26: 400–411.

Lloyd, E. A. (1988), *Structure and Confirmation of Evolutionary Theory*. Westport, CT: Greenwood.

Maienschein, J. (2004), *Whose View of Life? Embryos, Cloning, and Stem Cells*. Cambridge, MA: Harvard University Press.

Pinto-Correia, C. (1997), *The Ovary of Eve: Egg and Sperm and Preformation*. Chicago: University of Chicago Press.

Provine, W. B. (1971), *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press.

Roe, S. A. (1981), *Matter, Life, and Generation: Eighteenth-Century Embryology and the Haller-Wolff Debate*. Cambridge: Cambridge University Press.

Sarkar, S. (1994), "The Selection of Alleles and the Additivity of Variance," in D. Hull, M. Forbes, and R. M. Burian (eds.), *PSA 1994: Proceedings of the 1994 Meeting of the Philosophy of Science Association*. East Lansing, MI: Philosophy of Science Association, 3–12.

——— (1998), *Genetics and Reductionism*. New York: Cambridge University Press.

Wagner, G. P., M. Laubichler, and H. Bagheri-Chaichian (1998), "Genetic Measurement Theory of Epistatic Effects," *Genetica* 102/103: 569–580.

Wimsatt, W. C. (1981), "Units of Selection and the Structure of the Multi-Level Genome," *PSA 1980* 2: 122–183.

See also **Adaptation and Adaptationism; Evolution; Fitness; Genetics; Natural Selection; Population Genetics**

DAVID HILBERT

(23 January 1862–14 February 1943)

Hilbert was, and still is, known as one of the greatest mathematicians of the first half of the twentieth century. Although this view is doubtlessly correct, it is one-sided and incomplete because it neglects two important aspects of his work:

1. Throughout his career, Hilbert was interested in the foundations of all exact sciences—not only mathematics, but also the natural sciences. As a truly universal scientist, he contributed to fields other than pure mathematics, in particular, to theoretical physics and its dramatic development during the first quarter of the twentieth century.
2. Hilbert consciously and deliberately transcended the borders between mathematics and the exact sciences, on the one side, and epistemology and philosophy, on the other. This situates him with other twentieth-century figures such as Einstein, Bohr, Born, Schrödinger, and Weyl, who, like Hilbert, aimed to tear down the wall between traditional philosophy and the exact sciences.

In spite of numerous accounts of Hilbert's achievements in mathematics, his work in other areas—in particular, his contributions to modern physics and its philosophical implications—remains relatively neglected. Because the present volume is an encyclopedia of the philosophy of science, this article focuses on these other aspects of Hilbert's work. However, his mathematical achievements will not be entirely ignored: It will be pointed out that there is an intimate relationship between his mathematical work and his contributions to modern physics, especially its conceptual clarification.

David Hilbert was born in Königsberg, then the capital of East Prussia. He attended the local Gymnasium and spent most of his student life in Königsberg. During these years, he spent much time with Hurwitz and Minkowski studying mathematics and physics. In 1886, he became *Privatdozent* in Königsberg with a highly regarded work on the theory of invariants. Six years later, he was appointed as *Extraordinarius* and became full professor for mathematics in Königsberg a year later. In 1895, Felix Klein brought him to Göttingen, where he remained, despite many offers from other distinguished universities, until his death. In 1925, Hilbert suffered from "pernicious anemia" but recovered soon thanks to a new medication. During his academic career in Göttingen, he established (together with Klein and Courant) the best-known and most esteemed center for mathematics in the world. The fruits of his research in the foundations of mathematics and the sciences are still significant.

The principal means by which Hilbert achieved most of his fundamental results in the foundations of mathematics and science was the so-called *axiomatic method.* Since there are many obscure and confused opinions about this method, especially about its essential role (as well as its limits) in the logical analysis of the exact sciences, the subsequent sections discuss what the "essence" of this method is (see also Hilbert 2004). A similar clarification is necessary with respect to Hilbert's so-called *formalistic* approach as an alternative to Brouwer's *intuitionistic* and Frege's *logicistic* views about the foundations of mathematics (Sieg 1990).

Foundations of Geometry and the Axiomatic Method

Hilbert's inquiries into the foundations of mathematics fall into two periods, which are separated (judged by his publications) by about fifteen years. The first period began in 1893 with a lecture on "projective geometry," reached its zenith with *Grundlagen der Geometrie* (Hilbert [1899] 2004), and ended with *Über die Grundlagen der Geometrie* (Hilbert 1902). The second period, which started in 1917–1918 with the programmatic essay "Axiomatisches Denken" (Hilbert [1918] 1932), includes most of Hilbert's investigations of the foundations of arithmetic and the establishment of a radical new program, called *proof theory,* to prove the consistency of arithmetic by finite means. This important program came to a halt (at least temporarily) in 1931, when Gödel published his famous paper showing the existence of undecidable sentences within Peano arithmetic (Gödel 1986). This happened just as Hilbert was going to retire from his position at Göttingen.

This division of Hilbert's career into two periods gives the impression that the two topics that characterized them were for Hilbert unrelated, which is not the case. In fact, the latter is the continuation of the former by other much more radical means. This becomes obvious if one considers Hilbert's early development more closely, taking into account both his published papers and his unpublished lectures. The first point that should be noted is the trivial fact that Hilbert started his research in geometry quite conventionally. He did not have the axiomatic method at his disposal. Instead this method first emerged in connection with his "meta-theoretical inquiries of the logical structure of geometry." What this phrase means will become clear in the next section.

Around 1890, when Hilbert began his studies in geometry, the intellectual situation in that discipline was rather complicated. Geometry had become torn asunder into a confusing number of different branches and competing programs, such as projective versus differential geometry, Euclidean versus non-Euclidean geometry, and synthetic versus analytic geometry. Interesting but unconnected results were being discovered, and important books appeared, such as those by von Staudt (1847) and Pasch (1882). Hilbert was not acquainted with all of them initially. But once he had read Pasch's book (in about 1893), he knew, at least in principle, what his main goal was: He intended to resuscitate Euclid's axiomatic point of view, not exactly in the same way as used by Euclid, but in a very similar form. The main difference was to do it more transparently and perfectly. This meant that the desired axiom system should have a "perspicuous" logical-deductive order, that is, it should be complete in the sense that no essential assumption is missing, and simple in the sense that it contains no superfluous assumptions.

To achieve this goal, Hilbert had to develop a device by which he could prove whether a given sentence is logically dependent on or independent of a certain set of sentences. The principal idea is the following: A given sentence S is logically independent of a set of sentences $S_1, \ldots, S_n$, if there is a structure in which all sentences $S_1, \ldots, S_n$ are fulfilled (or true, as is now said), but S is not, and, instead, $\neg S$ is. If there is no such structure, that is, if in all structures in which $S_1, \ldots, S_n$ hold, S holds, S is a *logical consequence* of $S_1, \ldots, S_n$. This idea of a structure is the core of model theory (if "fulfilled" is replaced by "true"); it is also an essential ingredient of what Hilbert called the *axiomatic method.* The latter is the deliberate change or *variation* (Hilbert's term) of an axiom system in order to study the logical dependence of a specific sentence (e.g., the axiom of parallels) from a given set of axioms by model-theoretic means. Hilbert took primarily algebraic number fields as models for his proofs. In this way he was able to prove many interesting results—for example, the independence of Archimedes' axiom of continuity from the remaining axioms of his axiomatization of Euclidean geometry, the nonprovability of the Desargues sentence in Euclidean geometry without the axioms of congruence, the nonprovability of Pascal's sentence without the axiom of Archimedes, and others (see Hilbert [1899] 2004, Chs. 2, 3, and 5, for more details).

From an epistemological point of view, more important than these particular results is another, more general, aspect of the axiomatic method. With this method Hilbert could not only answer logical questions of independence and dependence, but also analyze meta-theoretical problems like the consistency and completeness of an axiom system, which roughly means (syntactically) that it is impossible to deduce a sentence *A* and its negation, and furthermore that every sentence which is "intuitively" true can be deduced from the axiom system.

These questions had become particularly pressing since the consistency of "non-Euclidean geometry" (taken quite generally) was still unproven (respectively the existing "proofs" were doubted). Hence, the consistency of the different non-Euclidean systems could not simply be taken for granted, but had to be shown definitively. Hilbert's idea to prove the consistency of non-Euclidean axiom systems was strikingly simple: If arithmetic is consistent (as everyone believed), then the consistency of a geometrical axiom system can be proved by relating it to the consistency of a suitably chosen number field. (This idea was quite original; it cannot be found in Pasch's work.) Today a geometrical axiom system is considered consistent if it has a numerical model. Hence, consistency means only "relative consistency": If it is not possible to deduce a contradiction within the numerical model, then the coordinated geometrical axiom system is consistent.

Execution of this program is very tricky, but Hilbert could show that his full axiom system of Euclidean geometry (including the axioms of continuity) is "complete" in the sense that it has a "numerical" model (the real numbers) whose domain of individual elements cannot be expanded without introducing a contradiction. Hence, Hilbert's axiom system of Euclidean geometry is consistent in virtue of its completeness if the real-number field is consistent, which no one doubts seriously. Hilbert's concept of completeness is today called *categoricity.* A theory is categorical iff it has, up to isomorphism, exactly one model (Majer 1998).

New Foundations of Mathematics and the Genesis of Proof Theory

With the proof of the relative consistency of a large number of non-Euclidean geometries a new problem emerged: How could the consistency of arithmetic itself be proved? Although it was immediately clear to Hilbert that this could not be done in the same style as in geometry (otherwise, one would be trapped in an infinite regress), around 1900 he did not know how to achieve this goal. It wasn't until 1920 that his first proposals occurred, in a pair of unpublished lectures given in Göttingen, which was the beginning of modern proof theory. Its basic assumptions and procedures include:

1. Arithmetic cannot be reduced to logic, because it has a (nonlogical) content, given in intuition.
2. Contradictions can be avoided totally if all operations used in calculating and reasoning remain finite.
3. Because mathematics was and shall remain a "free" science, it cannot be restricted to the finite.
4. Consequently, to save mathematics from the danger of inconsistency (by entering the transfinite), it has to be proven by *finite means* that no contradiction can be derived from an *appropriate* system of axioms for arithmetic.
5. In order to achieve this goal, the axiom system itself has to be "formalized" so that a meta-theoretical investigation about its deductive structure is possible.

"Formalization" here means two things: (a) all axioms have to be expressed as formulas of a definite language (*Zeichensprache*), with clear syntactical rules for the formation and transformation of formulas, and (b) all logical rules of deduction such as *modus ponens* and substitution *salva veritate* have to be made explicit. Once this is done, the formalized system must be investigated to determine whether it is possible to derive a pair of formulas A and $\neg A$ from the axioms. If this is impossible, the system is consistent. The real difficulty lies in the proof that this is indeed impossible, because the proof of the impossibility has to be finite, whereas the number of possible proofs within a formal system can (and usually is) infinite.

Today (70 years after Gödel) most scientists believe that such a *direct* (nonrelative) proof of the consistency of an axiom system for arithmetic is impossible. But this is incorrect for several reasons. First, Hilbert and his school presented such proofs for certain "weak" axiom systems of arithmetic. Second, if one gives up the finitist restrictions on the meta-theoretical proofs and permits, as Gentzen did, transfinite induction, the consistency of arithmetic can be proved. Third, Gödel's proof of the existence of undecidable sentences within Peano arithmetic, and hence the infeasibility of a formal proof of consistency of Peano arithmetic by finite means, is no absolute verdict. Its correctness depends on the meaning of the phrase "formal proof by finite means." There are a number of proposals about how to "sidestep" Gödel's verdict without betraying Hilbert's finite point of view (modern proof theory, inverse arithmetic, weak arithmetic, etc.) (Simpson 1999).

Hilbert's Contributions to Physics and Its Axiomatic Foundations

Hilbert delivered his first lecture in physics in 1898, a year before the *Foundations of Geometry* appeared. During the next decade, he lectured six times on classical and continuum mechanics, including hydrodynamics, electrodynamics, and thermodynamics. During this period, mechanics was for Hilbert *the* fundamental theory of all physics, as it was for Heinrich Hertz (1895), whose book *Die Prinzipien der Mechanik* Hilbert admired as an exemplar for his own axiomatic point of view in physics. The period of classical physics (based exclusively on Galilean space-time theory) ended when Hilbert (in cooperation with Minkowski) began studying Einstein's theory of special relativity. This led to a complete revision of the mechanics lecture of 1911 in which Hilbert presented Einstein's theory of special relativity and discussed its consequences for electrodynamics and thermodynamics.

Hilbert's first publications in physics date from 1912 and are closely related to his monumental work on the theory of integral equations (Hilbert 1912). In fact, "Begründung der kinetischen Gastheorie" first appeared as Chapter 22 of that book. This gives the impression that Hilbert's interest in physics was only that of a mathematician, who is "simply looking for another possible application of his mathematical theories" (Brush 1976, 448) without any real interest in physics. A cursory examination of the unpublished lectures shows that this claim is unjustified. Although Hilbert's main concern in the paper is the search for solutions to the Maxwell-Boltzmann equation, his primary concern in the lectures on the kinetic theory of gases is different. The main question he pursued is whether a logical derivation of the Maxwell-Boltzmann equation from the time-reversible equations of mechanics is possible (see Irreversibility; Time). This is a very interesting question, and the different answers proposed so far are still controversial. Hilbert himself favored a negative answer in the sense that no strict logical deduction is possible. This leads to a new problem: What does "irreversibility" mean in an objective sense, if the irreversible Maxwell-Boltzmann equation cannot be deduced from the fundamental equations of motions? (see Majer 2002 and Scheibe 1997).

Similar points can be made regarding the reception of Hilbert's work in physics (such as his papers on radiation theory) as merely applied mathematics. As in the former case, this work was closely related to his new theory of integral equations and was thought by Pringsheim (a leading figure in radiation theory) as entailing nothing physically new about Kirchhoff's law of radiation. This is a misapprehension. Hilbert pointed out a serious logical gap in the foundation of radiation theory and the deduction of Kirchhoff's law from more fundamental theories, but his work was dismissed as of merely mathematical interest.

Better received were his two papers *The Foundations of Physics* in 1915 and 1916 (republished in Hilbert 1924), in which Hilbert presented several generalized field equations, which turned out to be equivalent to Einstein's field equations of 1915. (Therefore, they are also sometimes called the Hilbert-Einstein equations.)

Nineteen fifteen was, in retrospect, the year of the most intensive research on the theory of general relativity, in both Hilbert's and Einstein's careers. Both struggled in searching for *universal* field equations, in which two fundamental forces would be united: electromagnetism and gravitation. They approached the problem from different points of view. To Hilbert as "mathematician" it was clear from the very beginning that the field equations had to be independent of the choice of the coordinates, and, more important, it was clear what this precisely meant in physical terms. Einstein, on the other hand, had to struggle as a self-taught mathematician with the problem of invariance for several years before finding an acceptable solution.

More important than their different technical approaches are their differences in physical perspectives. Hilbert, having once abandoned the *mechanical* worldview, followed Gustav Mie, who tried to develop a universal field theory from which the existence of the electron could be explained. Einstein, however, looked for a generalized field theory in which Newton's theory of gravitation could be embedded, at least approximately. The solutions Einstein and Hilbert found toward the end of 1915 are, seen from this perspective, only contingently the same. In "essence" they are rather different (Weyl 1988). Recently, a priority debate among historians of science has emerged about whether Hilbert "acquired" his field equations from Einstein (Sauer 1999).

In his last period of active research in physics (1926–1927), Hilbert lectured on "Mathematical

Methods of Quantum Theory" (unpublished typescript). This was just after the *new* quantum mechanics had been formulated by Heisenberg, Born, and Jordan, and independently by Schrödinger and Dirac ("new" in distinction to the old "quantum theory" of Bohr) (see Quantum Mechanics). Although Hilbert did not belong to the group of physicists who created the new quantum mechanics, he is doubtless one of its intellectual progenitors, since the mathematical foundations of the new theory were precisely his theory of integral equations with infinitely many variables, which he had developed fifteen years earlier. It is therefore not accidental that in modern textbooks of quantum theory, the basic concept is the so-called *Hilbert space.* Hilbert's lectures significantly influenced the further development of quantum theory, in particular its axiomatic presentation by von Neumann. The lectures predate the publication of the famous paper by Hilbert, von Neumann, and Northeim (1928), which became the starting point of von Neumann's (1932) *Mathematische Grundlagen.*

Hilbert's "Finite Point of View" and Recursive Epistemology

Hilbert was not a "professional" philosopher, but he studied Kant's *Critique of Pure Reason* and acquired an intimate knowledge of the writings of contemporary philosophers such as Husserl, Frege, and Russell. Traces of these authors (and some minor figures) can be found in Hilbert's work. Hilbert was, however, too independent a thinker to simply adopt their philosophical views. Instead he selected only those aspects that he found acceptable in light of the extraordinary progress of mathematics and science in the late nineteenth and early twentieth centuries. He tried to "unify" these aspects into a coherent view. This was not easy, because there were conflicts in these views, which had to be resolved. The best example of such a resolution is his "finite point of view" regarding the foundations of mathematics, by which the actual infinite of Cantor's set theory should be tamed. This led to the idea of proof theory, in which the "infinite" of arithmetic should be controlled by "finite means," that is, by proof of its consistency within a "finite" fraction of itself. Although there are some doubts as to what precisely this finite fraction is, there are strong indications in Hilbert's work that he thought it was primitive recursive arithmetic.

Perhaps more important than this example from pure mathematics are the conflicting moments or centrifugal forces in Hilbert's conception of geometry. They are best understood by considering the following three statements:

1. Geometry is a natural science.
2. The task of geometry is the logical analysis of human spatial intuition.
3. Geometrical conventionalism (a la Poincarè) is an untenable position in spite of the fact that experimental results (e.g., the light deflection in the gravitational field of the sun) are easier explained by assuming a "variable metric" than by clinging to Euclidean geometry and introducing new material forces; because the introduction of such forces is quite ad hoc (see Conventionalism; Poincarè, Henri).

At first glance, the three statements seem incompatible. There are, however, several ways to make them coherent.

The easiest way is to drop statement 2 and consider geometry as a purely empirical science. This is, roughly, what the logical empiricists did. But this is not Hilbert's point of view. He insists on statement 2 (Hilbert [1899] 2004). The second option is to take Einstein's view and distinguish geometry as a formal mathematical discipline (which can be known a priori) from a physical discipline (which can be known only a posteriori). This again is not Hilbert's position, because in his view such a separation is totally arbitrary and problematic. The third option is to take Poincarè's position seriously and regard the choice of geometry as a matter of convention (like the choice of meter or yard as unit of length) and then stick to Euclidean geometry as the simplest one. Hilbert rejects this opinion as confused, because it confounds two concepts of simplicity, which have to be sharply distinguished: an old intuitive notion and Hertz's new methodical notion of simplicity. Poincarè introduces a metrical structure into geometry as simple, which is unnecessary, and this violates Hertz's principle of simplicity: Do not introduce superfluous elements into a theory.

But what is Hilbert's solution? How can he make the three statements coherent? The answer can be stated in a very much abbreviated form as this: Human beings have a spatial intuition of external objects, which they use in normal life. For the first level of conscious reflection, the most significant facts of human intuition were conceptually identified and put into an axiomatic order of deduction. This is roughly what Euclid achieved in his *Elements.* In the second stage, mathematicians began making spatial intuition the object of logical investigation. This led to a multiplicity of geometries, whose logical relations could be studied by the axiomatic method (including model theory). For a correct understanding of Hilbert's epistemology, it is important to note that this process took place without any input from the natural sciences. The application of non-Euclidean geometry to physics came after the logical analysis had been achieved. There was a second, equally important part of Hilbert's view: Euclidean geometry was not simply abandoned. It still played a decisive role, intellectually as well as practically, but not, as Poincarè supposes, because it is simpler to cling to Euclidean geometry. Hilbert rejected this view because he thought it was like an ''idle wheel'' in general relativity. The true reason that Euclidean geometry is used not only in daily life but also in science is that the deviations from it are so unimaginably small that it would be ridiculous in most cases to replace it by a non-Euclidean geometry. This remains true also when measuring devices are constructed, by which Einstein's or any other theory of general relativity is tested.

Ulrich Majer

References

Brush, S. G. (1976), *The Kind of Motion We Call Heat: A History of the Kinetic Theory of Gases* (Vol. 2). Amsterdam: Elsevier.

Gödel, K. (1986), *Collected Works* (Vol 1). Edited by S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, and R. Solovay. New York: Oxford University Press.

Hertz, H. (1895), ''Die Prinzipien der Mechanik,'' in *Gesammelte Werke* (Vol. 3). Leipzig: J. A. Barth.

Hilbert, D. ([1899] 2004), ''Grundlagen der Geometrie,'' in M. Hallett and U. Majer (eds.), *David Hilbert's Lectures on the Foundations of Geometry: 1891–1902*. Berlin: Springer, 72–123.

——— (1902), ''Über die Grundlagen der Geometrie,'' *Mathematische Annalen* 56: 381–422.

——— (1912), *Grundzügeeiner allgemeinen Theorie der linearen Gleichungen*. Leipzig: Teubner.

——— ([1918] 1932), ''Axiomatisches Denken,'' in *Gesammelte Abhandlungen* (Vol. 3). Berlin: Springer, 146–156.

——— (1924), ''Die Grundlagen der Physik,'' *Mathematische Annalen* 92: 1–32.

——— (2004), *David Hilbert's Lectures on the Foundations of Geometry: 1891–1902*. Edited by M. Hallett and U. Majer. Berlin: Springer.

Hilbert, David, J. von Neumann, and L. Northeim (1928), ''Über die Grundlagen der Quantenmechanik,'' *Mathematische Annalen* 98: 1–30.

Majer, U. (1998), ''Husserl and Hilbert on Completeness,'' *Synthese* 110: 37–56.

——— (2002), ''Lassen sich Phänomenologische Gesetze *im Prinzip* auf mikro-physikalische Theorien reduzieren?'' in M. Pauen and A. Stephan (eds.), *Phänomenales Bewusstsein: Rückkehr zur Identitätstheorie?* Paderborn, Germany: Mentis Verlag.

Pasch, M. (1882), *Vorlesungen über neuere Geometrie*. Leipzig: Teubner.

Sauer, T. (1999), ''The Relativity of Discovery: Hilbert's First Note on the Foundations of Physics,'' *Archive for the History of the Exact Sciences* 53: 529–575.

Scheibe, E. (1997), *Die Reduktion physikalischer Theorien. Ein Beitrag zur Einheit der Physik* (Vol. 2). Berlin: Springer.

Sieg, W. (1990), ''Relative Consistency and Admissible Domains,'' *Synthese* 84: 259–297.

Simpson, S. G. (1999), *Subsystems of Second-Order Arithmetic*. New York: Springer.

von Neumann, J. (1932), *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer.

von Staudt, K. G. C. (1847), *Geometrie der Lage*. Nürnberg: Bauer and Raspe.

Weyl, H. 1988. *Riemanns geometrische Ideen, ihre Auswirkungen und ihre Verknuepfungen mit der Gruppentheorie*. Berlin: Springer.

See also **Carnap, Rudolf; Conventionalism; Logical Empiricism; Quantum Mechanics; Physical Sciences, Philosophy of; Poincarè, Henri; Space-Time; Vienna Circle**

I

IDEALIZATION

See **Approximation**

IMMUNOLOGY

Because of its eclectic contributions to pathology, clinical medicine, and basic biology, immunology cannot be defined by a single, unifying experimental framework. Rather, it is (and has been) characterized by multiple, even competing, thought styles (Crist and Tauber 1997), each requiring a different methodological apparatus to order its experimental program—from receptor biology to molecular biology, from allergy to xeno-transplantation, from infectious diseases to rheumatoid arthritis and diabetes. The discipline is experimentally divided by the examination of two broad arenas of immune function:

1. Innate immunity, employing more ancient phylogenetic mechanisms, which deploys various identifying proteins (lectins and complement) to target pathogens for destruction by phagocytes, and
2. Acquired immunity (found only in vertebrates), which consists of antibodies (immunoglobulins) and lymphocytes (T cells and B cells); it is more specific in its identification capabilities and its memory of prior immune encounters.

The lymphocyte, because of its central role in contemporary clinical immunology (ranging from vaccination to transplantation to neuroendocrinology), has become the intense focus of current investigations. But underlying each branch of immunology, the concept of an identified and protected "self," a theoretical construction and fecund metaphor, has served as the central theme that integrates this diverse discipline.

During the last three decades of the twentieth century, immunology has commonly been described as the science that distinguishes 'self' from 'nonself,' and upon this distinction the means to preserve organismal integrity has defined the scope of immunity. In this formulation, the host organism, perceiving an invasion by microbial pathogens, mounts a defensive response. Contemporary immunology has broadened this agenda to include surveillance of the body for malignant, effete, damaged, or dead host constituents (altered "normal" cells), as well as autoimmune processes directed against undamaged elements—some of which may be part of ordinary physiological economy, while others are pathological. The challenges to define a basis for immune identity, within the coupled ambiguities of autoimmunity and tolerance (the reciprocal nonreactivity to host constituents), has generated debate about selfhood as an organizing concept for the discipline. The immune self, an implicit entity in the late nineteenth century (Tauber and Chernyak 1991; Mazumdar 1995) and a hotly contested one today (Langman 2000), is a rich philosophical topic, in terms of both its epistemological standing as well as its metaphysical foundations (Tauber 1994 and 1999). Note that while the immune self is rooted historically in the problematics of biological individuality (Loeb 1945; Buss 1987), its philosophical attention is distinct from those concerns (Wilson 1999) and is subsumed in the broader questions of reductionism (see Reductionism).

This article will outline in a historical context the two principal theories governing immunology's research program: the theory of immune identity and the more recent one that challenges the very notion of selfhood. In those constructions are reflected the prevailing attempts to define the concept of organism.

Historical Antecedents

The first medical use of the term "immunity" (originally a legal designation conferring exemption and distinction) appears in 1775, when van Sweiten, a Dutch physician, used *immunitas* to describe the effects induced by an early attempt at variolization (Moulin 1991, 24). But the concept did not develop until the mid-nineteenth century, when Claude Bernard set the theoretical stage for the autonomous organism (Cohen 2001). In contradistinction to an animal in humoral balance (i.e., the body conceived as composed of various "humors" that were in balance during health and unbalanced in disease with a pervasive environment), Bernard postulated the primacy of the organism's essential independence. In this view, animals provided discrete sites for methodological medical experimentation, as well as a focus for a theoretical reductive strategy based on positivist principles. Together, these two views provide medicine with its modern experimental basis.

Bernard furnished biology with a new concept of the organism, which would have wider ramifications than the establishment of physiology and biochemistry. Obviously, interchange with the environment was a necessary requirement for life, but Bernard emphasized how boundaries provided the crucial metabolic limits required for normal physiological function. With his concept of the *milieu interieur*, the body was envisioned as a demarcated, interdependent, yet autonomous entity ("corporeal atomism" [Cohen 2001, 190]), thereby establishing the theoretical grounding that became the *sine qua non* for the development of the models for infectious diseases, genetics, neurosciences, and immunology in all of its various guises. But as important as Bernard's concept proved to be for certain sciences, his construction also obfuscated certain aspects of biology's complexity. Most importantly, the ecological consciousness that emerged in the twentieth century found itself enmeshed in a conceptual struggle to promote a contextualist approach to complex biological environments populated by multiple species, against a biology dominated by the centrality of the autonomous organism. Even within the confines of biomedicine, Bernard's focus on the individual proved inadequate for the hygienic movement of his own period and later developments in public health. But Bernard introduced a revolutionary formulation, notwithstanding its limitations, and immunology became one of its defining sciences—indeed, immunity was alien to the older humoral view. By radically changing the inside/outside topology so that the organism's interior became the determining context of function, Bernard effectively isolated the organism from its environment and joined a complex cultural movement of redefining the body more generally.

Bernard's notion of the body as independent of the environment complemented Malthusian economics, liberal political philosophy, and Comtian sociology. From these and other disparate sources, the autonomous, atomistic body as a political, social, economic, and medical entity was redefined in the nineteenth century (Foucault 1973; Agamben

1998), and Bernard played a central role in providing a theoretical biological foundation for its critical nexus in various discourses. Notwithstanding that "independence" is a political term and fairly represents neither the dialectical relationships of the organism and its environment (Levins and Lewontin 1985) nor the evolutionary peculiarities of individuality itself (Buss 1987), the formulation has served as the touchstone for various cultural constructions of identity. Indeed, culture critics have seized on immunology as paradigmatic for the modern notions of identity, where boundaries are contested and the body becomes the localized site of battle between self and other (Haraway 1989; Martin 1994). The warfare metaphors—"attack," "defense," "invaders"—so prevalent in immunology's lexicon, dramatically illustrate this construction, in terms of both the self/other dichotomy and the privileged regard of individuality over community.

Origins of the Immune Self

Immunology's history is generally regarded as intimately tied to those discoveries leading to the elucidation of the bacterial etiology of infectious diseases, which draws together twin disciplines—microbiology (the study of the offenders) and immunology (the examination of host defenses). Thus, in this pathological context, immunology began as the study of how a host animal reacts to pathogenic injury and defends itself against the deleterious effects from such microbial insult. This is the typical historical account of immunology as a clinical science, a tool of medicine; and as such, it focused almost exclusively on the role of immunity as a defender of the infected. The paradigmatic host is the patient, an infected "self," which is the critical element for the power of this view. The clinical orientation, which *assumes* a given entity—the self—is obviously a dominant organizing perspective, but another perspective turns this assumption into a question or a problem: Rather than the science that seeks to discern the basis of self/nonself discrimination, immunology may also be regarded as more fundamentally concerned with the *establishment* of organismal identity.

This latter point of view was offered by Elie Metchnikoff, who came to the nascent field of immunology from an unexpected theoretical and methodological perspective—that of an embryologist—and sought to discover genealogical relationships in the context of Darwinism (Tauber and Chernyak 1991). Intrigued with the problem of how divergent cell lineages were integrated into a coherent, functioning organism, Metchnikoff was thus preoccupied with the problems of development as process, which he regarded as analogous to Darwinian interspecies struggle: Cell lineages were inherently in conflict to establish their own hegemony, but he hypothesized that unlike nature writ large, a regulatory system was required to impose order, or what he called "harmony," on the disharmonious elements of the animal. He found such an agent in the phagocyte, which retained its ancient phylogenetic eating function, to devour effete, dead, or injured cells that violated the phagocyte's sense of organismal identity. When pathogenic microbes were discovered in the 1870s, Metchnikoff soon assigned his phagocyte the new role of defending the organism against invaders. Indeed, in this context, the phagocyte became an exemplary combatant of Darwinian struggle, now occurring within the organism.

In Metchnikoff's theory, immunity was a particular case of physiological inflammation, a normal process of animal economy. But there was a more subtle message: (1) Immunity was an active process, with the phagocyte's response seemingly mounted with a sense of independent arbitration, and (2) organismal identity was a problem bequeathed from a Darwinian perspective that placed all life in an evolutionary context. In short, he combined a Darwinian sensibility to a Bernardian conceptualization of autonomy.

Metchnikoff's overall representation constituted the phagocyte as an *agent* (Crist and Tauber 2001), an actor that was the cause of its own action, as a matter of endogenously generated and directed behaviors. The portrayal of the phagocyte as autonomous is largely derived from the linked features of its capacity to sense its environment and move freely within it, and the various degrees of unpredictability and meaningfulness that characterize this behavior. Indeed, the phagocyte, as an agent, becomes a metaphorical 'self,' a primordial microcosmic expression of what later immunologists would extend into an epistemology of biological identity. But while placing the identity function at the nexus of immunology's concern, Metchnikoff failed to provide the necessary preconditions for those who would seek to demonstrate those reactions that conferred protection of such an *entity.* Much of the subsequent history of immunology may be traced to the attempts to establish a definition and experimental basis that fulfills such an identity function, an effort that may be fairly regarded as remaining unresolved, as an ambiguity at the very heart of the discipline.

Twentieth-Century Constructions of the Immune Self

In the first half of the twentieth century, immunology was devoted to establishing the chemical basis of specificity, which unreflectively assumed the parameters of selfhood (Silverstein 1989; Mazumdar 1995). Paul Ehrlich, whose early scientific research concerned the chemical specificity of dyes, applied his general notions of biological affinity to immunology and thereby provided the first theory of immune specificity. Analogous to Fisher's model of "lock and key" binding of organic compounds, Ehrlich proposed that antibodies and their targets bind according to corresponding structural fittings. His postulated "side chains" were *cellular receptors* (a term he coined) for bacteria and their products. When confronted with infection, proliferation of side chains (which in solution were liberated as "anti-bodies") bound and thus neutralized pathogens and their toxic products. This mechanism, coupled with phagocytes, provided the host organism with a defense against microbes, and Ehrlich shared the Nobel Prize with Metchnikoff in 1908 in recognition of the synthesis of their respective (cellular versus immunochemical) points of view.

By World War I, Karl Landsteiner had demonstrated the extraordinary finesse of chemical recognition, and the biological mechanism that accounted for antibody generation to a seemingly infinite array of antigens (targets of antibody recognition) accommodated itself to the colloid theory of protein structure. In this view, antibodies were thought to form upon an antigenic template, which then would serve as a model for the multiplication of identical antibody molecules (Silverstein 1989). With the understanding of protein synthesis inspired by Watson and Crick in the 1950s, such template models violated DNA-directed protein synthesis, and a new theory soon followed. In 1955, Niels Jerne postulated that antibodies were "selected" from a pool of "natural antibodies" on the basis of their respective affinities for antigen. From this subpopulation of the antibody pool, an array of appropriate neutralizers would be conscripted. This suggestion was soon followed by a biological model to explain how such "natural selection" operated (Tauber 1994). David Talmage and Frank Burnet's better developed "clonal selection theory" (CST) (Burnet 1959), predicted that antibody selection occurred at the level of the antibody-producing lymphocyte, whose singular antibody receptor had high affinity for antigen. Consistent with the peptide model, they proposed that antigen binding stimulated cellular proliferation and differentiation of those cells (clones) that shared an appropriate affinity profile for those pathogens, toxins, allergens, and any other "foreign substance" that was so recognized.

Within a decade, compelling evidence confirmed these theoretical musings, and the central question then became one of the intracellular mechanism of antibody generation. By the late 1970s, this beguiling puzzle was solved when it was shown that immunoglobulin (antibody) was made up of segments that were put together like so many "cards" from a genotypic "deck," and that these synthesized proteins also underwent somatic mutation. (The elucidation of antibody generation was generally important, for it demonstrated the plasticity of the genome and the genotypic variability of individual cells.) Thus the bewildering specificity of the immune reaction could be accounted for by the shuffling of a finite number of genes (coding for the cards in the deck) and a mechanism of fine-tuning somatic mutation that gave rise to "custom" antibodies with highly specific binding characteristics (Podolsky and Tauber 1997). This breakthrough was only the most celebrated of immunology's molecularization. Indeed, the entire field of immunology was now committed to defining the molecular pathways of immune effector functions, the structure/function relationships of various mediators, and the molecular control mechanisms of what became an increasingly complex system of interactive components (Tauber 1996).

But this shift to a highly sophisticated molecular approach should not obscure the underlying theoretical questions being addressed of a new biologically oriented program. While Ehrlich's chemical perspective had dominated immunology until World War II, these new hypotheses concerning antibody generation, coupled to clinical demands in the fields of organ transplantation and autoimmunity, drove immunology toward a new biological theory of immunity that was both more comprehensive than earlier chemical models and far-reaching in its theoretical implications. Burnet not only provided a mechanism for antibody selection and biological generation, he also presented a theory of immunological "tolerance" that was to henceforth dominate the field (Burnet and Fenner 1949; Tauber 1994). From Burnet's perspective, foreign bodies are destroyed by immune cells and their products, whereas the normal constituents of the animal are ignored, that is, "tolerated." In other words, the identity of the host organism was a given within the Bernardian construct, with implicit boundaries as defined by immune reactivity. What was "attacked" was "other"; that which was regarded by immune silence became the self. What was, perhaps, implicit in pre–World War II immunology now declared its theoretical basis: Without a theory of self/nonself differentiation, immune reactivity had no biological basis for control. Indeed, the 'self' was introduced by Burnet into the immunological lexicon specifically to address immunity as an organismal phenomenon.

Unlike Metchnikoff, Burnet sought a firm definition of the immune self. Burnet's theory proposed that the animal, during prenatal development, exercises a purging function of self-reactive lymphocytes (the cells responsible for synthesizing reactive antibodies and mediating so-called cellular reactions) so that all antigens (substances that initiate immune responses) encountered during this period would attain a neutrality status. Thus, lymphocytes with reactivity against host constituents are putatively destroyed during development, and only those tolerant lymphocytes that are nonreactive are left to engage the antigens of the foreign universe. The hypothesis (first presented in 1949 and later developed into the CST [Burnet 1959]) contained two key challenges that dominated immunology: (1) How was tolerance induced and autoimmunity controlled? and (2) What was the mechanism that accounted for antibody and lymphocyte diversity? As already noted, the latter issue was solved by molecular biologists by the late-1970s (Podolsky and Tauber 1997); the former question, involving systems analysis, apparently requires a comprehensive model of the immune system as a whole and remains enigmatic.

Aside from incomplete accounts of tolerance, there were early discrepancies arising from a continuum of autoimmune reactions, ranging from normal physiological and inflammatory processes to uncontrolled disease initiated by an immune reaction gone awry. Bountiful evidence in recent years has shown that autoimmunity is also a normal finding, and in these newer views, such functions are regarded as integrated within a more complex normal physiology. Thus, immune reactivity, rather than functioning only in an "other-directed" mode, is in fact bidirectional. This position contrasts with the "one-way" definition of selfhood, where there is a genetic self whose constitutive agents see the foreign bodies, and immune reactivity arises from this polarization with attack directed only against nonself (Tauber 1998). Not unexpectedly, in this turn inward, the immune self becomes increasingly difficult to define, and the concept becomes unable to easily accommodate these new appraisals. There are at least half a dozen different conceptions of what constitutes the immune self (Matzinger 1994, 993):

1. Everything encoded by the genome
2. Everything under the skin including/excluding immune "privileged" sites
3. The set of peptides complexed with T-lymphocyte antigen-presenting complexes, of which various subsets vie for inclusion
4. Cell surface and soluble molecules of B-lymphocytes
5. A set of bodily proteins that exist above a certain concentration
6. The immune network itself, variously conceived (detailed below)

While these versions may be situated along a continuum between a severe genetic reductionism and complex organismal view (Tauber 1998 and 1999), each shares an unsettled relationship to Burnet's original dichotomous model of self and other.

Assaults on the Immune Self

Well before the current debate about the immune self (Langman 2000), Niels Jerne attempted to dispel the many ad hoc caveats and paradoxes encumbering it by eliminating the self concept altogether. He went beyond the current notion of the immune network composed of lymphocyte subsets, secreting immunostimulatory and inhibitory substances (essentially a simple mechanical model with interlaced, first-order feedback loops) to propose a novel conception of immune regulation (Tauber 1994; Podolsky and Tauber 1997). His network theory was, from its very inception, a complex amalgam of pieces of the regulation puzzle fitted into place, with the overriding goal of understanding the immune system as a cognitive enterprise that spawned different formulations (e.g., Varela et al. 1988; Atlan and Cohen 1989; Stewart 1994a). In introducing this metaphoric construction of the immune system as analogous to the nervous system as early as 1960, Jerne set the stage for understanding newer immune metaphors (recognition, memory, learning) that built on the parallel with human cognition.

Jerne's idiotypic network theory hypothesis proposed that antibodies form a highly complex interwoven system, where the various specificities "referred" to each other (Jerne 1974). Under the general rubric of "cognition," he conceived of the immune system as self-regulating, where antibodies recognize not only foreign antigens, but self constituents as antigens (the so-called *idiotopes*). There was no essential difference between the recognized and the recognizer, since any given antibody might serve either, or both, functions. In other words, immune regulation was based on the reactivity of antibodies (and later lymphocytes) with their own repertoire forming a set of self-reactive, self-reflective, self-defining immune activities. There is no self and other for the immune system, for according to Jerne's theory, the system is complete unto itself, consisting of interlocking recognizing units: Each component reacts with certain other constituents to form a complex network or lattice structure. When the system is perturbated by the introduction of a substance that is recognized (i.e., it reacts with members of the system), this disturbance initiates immune responsiveness. Thus, foreignness per se does not exist in this formulation.

Jerne's theory presents a radically altered view of immune selfhood. In Burnet's simplified world of self/nonself discrimination, the immune system learned host/foreign distinctions, generated an army of reactive antibodies and lymphocytes, and acted accordingly when "antigen" was encountered. But Jerne coupled the simple antibody/antigen interactions to the far more complex and nondiscriminatory functions of the immune system, which built upon self-recognition. In his view, autoimmunity, instead of an aberration, became the organizational rule to explain immune function. Strikingly, there is no explicit mechanism for self/nonself discrimination, and this apparent lacuna served as the nexus of critiques (reviewed in Podolsky and Tauber 1997; Tauber 1999; Tauber 2000). But for Jerne, the need to define the self as distinct from other receded from his primary theoretical concerns, and this posture was to have important repercussions.

When the immune system is regarded as essentially self-reactive and interconnected, the "meaning" of immunogenicity, that is, reactivity, must be sought in some larger framework. Antigenicity then is only a question of degree, where self evokes one kind of response, and the foreign evokes another, not because of its intrinsic foreignness, but because the immune system sees the foreign antigen in the context of invasion or degeneracy. From the immune system's perspective, it only "knows" itself (Varela et al. 1988). Indeed, for Jerne, if a self was at all needed, it would be simply the immune system. Most importantly, the singular defensive purpose of immunity was widened to include an array of physiological functions, each of them now regarded as fully integrated within the immune system (Matzinger 1994). If eventually successful, this heralds a decisive shift in immunology's theoretical foundations, one more attuned to the diversity of immune functions that contribute to evolutionary fitness (Cohen 1992 and 1994; Stewart 1994a). While host defense is a critical function, it is hardly the only one of interest. Indeed, the immune system might be regarded as primarily fulfilling an altogether different role if its phylogeny is carefully examined. On this basis, John Stewart has provocatively suggested that the immune system became defensive only after its primordial neuroendocrine communicative capabilities (Ader, Felton, and Cohen 2001) were usurped for immunity (Stewart 1994b).

Biologists have increasingly come to appreciate that such systems are highly integrated within larger wholes and require analysis of how adjustments are made in relation to these other systems. This means, simply, that immune reactivity is determined by context (Cohen 1994; Podolsky and Tauber 1997), where agent and object play upon each other. Specific recognition of an antigen by a lymphocyte receptor is not sufficient for activation, for additional signals determine whether a cellular response or cell inactivation follows. In short, an antigen is neither self nor nonself except as it attains its meaning, so to speak, within a broader construct. Orthodox immune theory encompasses this idea in the so-called two-signal model, which does not require any of Jerne's hypotheses to fulfill its agenda. But there are more radical readings of the contextualist setting by which antigens are sensed, and debate concerning what constitutes the milieu of meaning of antigenicity and ensuing reaction have spawned certain provocative, and potentially important, models of immune regulation (reviewed in Podolsky and Tauber 1997; Tauber 2000).

In summary, immunology may be seen as structured on two major theoretical developments. The first was made by Metchnikoff in framing immunology with dual functions:

- establishment of organismal identity and
- protection of this integrity.

His immunochemical contemporaries and their direct heirs followed the second agenda to the exclusion of the first. The primacy of the identity issue was reintroduced by Burnet, and his program defined lymphocyte biology for the latter half of the twentieth century.

The second theoretical advance was made by Jerne, who moved past the identity issue altogether. No longer in service to a self, the immune system functioned within a greater whole as a cognitive faculty, perceiving only what it might know—*itself*. Patterns, context, and interlocution become organizing principles, so that the self metaphor, assuming a Jernian perspective, is eclipsed by another catchall metaphor, *cognition*. Even within such new formulations, the self still resides, reflecting a deep struggle over the character of biology, one that has its roots in Bernard's original understanding of autonomy, and now linked to our own more complex ecological views of agency and determinism.

Alfred I. Tauber

References

Ader, Robert, David L. Felten, and Nicholas Cohen (2001), *Psychoneuroimmunology*, 3rd ed. San Diego: Academic Press.

Agamben, Giorgio (1998), *Homo Sacer: Sovereign Power and Bare Life*. Stanford, CA: Stanford University Press.

Atlan, Henri, and Irun R. Cohen (eds.) (1989), *Theories of Immune Networks*. Berlin: Springer-Verlag.

Burnet, Frank Macfarlane (1959), *The Clonal Selection Theory of Acquired Immunity*. Nashville, TN: Vanderbilt University Press.

Burnet, Frank Macfarlane, and Frank Fenner (1949), *The Production of Antibodies*, 2nd ed. Melbourne, Australia: Macmillan and Co.

Buss, Leo (1987), *The Evolution of Individuality*. Princeton, NJ: Princeton University Press.

Cohen, Edward (2001), "Figuring Immunity: Towards the Genealogy of a Metaphor," in A. M. Moulin and A. Cambrosio (eds.), *Singular Selves: Historical Issues and Contemporary Debates in Immunology*. Amsterdam: Elsevier, 179–201.

Cohen, Irun R. (1992), "The Cognitive Paradigm and the Immunological Hommunculus," *Immunology Today* 13: 490–494.

——— (1994), "Kadishman's Tree, Escher's Angels, and Immunological Homunculus," in Antonio Coutinho and Michel D. Kazatchkine (eds.), *Autoimmunity: Physiology and Disease*. New York: Wiley-Liss, 7–18.

Crist, Eileen, and Alfred I. Tauber (1997), "Debating Humoral Immunity and Epistemology: The Rivalry of the Immunochemists Jules Bordet and Paul Ehrlich," *Journal of the History of Biology* 30: 321–356.

——— (2001), "The Phagocyte, the Antibody, and Agency: Contending Turn-of-the-Century Approaches to Immunity," in Anne Marie Moulin and Alberto Cambrosio (eds.), *Singular Selves: Historical Issues and Contemporary Debates in Immunology*. Amsterdam: Elsevier, 115–139.

Foucault, Michel (1973), *The Birth of the Clinic: An Archaeology of Medical Perception*. New York: Vintage.

Haraway, Donna (1989), "The Biopolitics of Postmodern Bodies: Determinations of Self in Immune System Discourse," *Differences* 1: 3–43.

Jerne, Niels K. (1974), "Towards a Network Theory of the Immune System," *Annals of Institute Pasteur/Immunology (Paris)* 125C: 373–389.

Langman, Rodney (ed.) (2000), *Self-Nonself Discrimination Revisited*. Special issue of *Seminars in Immunology* 12(3).

Levins, Richard, and Richard Lewontin (1985), *The Dialectical Biologist*. Cambridge, MA: Harvard University Press.

Loeb, Leo (1945), *The Biological Basis of Individuality*. Springfield, IL: C. C. Thomas.

Martin, Emily (1994), *Flexible Bodies: The Role of Immunity in American Culture from the Days of Polio to the Age of AIDS*. Boston: Beacon Press.

Matzinger, Polly (1994), "Tolerance, Danger, and the Extended Family," *Annual Review of Immunology* 12: 991–1045.

Mazumdar, Pauline M. H. (1995), *Species and Specificity. An Interpretation of the History of Immunology*. Cambridge: Cambridge University Press.

Moulin, Anne Marie (1991), *Le Dernier Langage de la Medicine: Histoire de l'Immunologie de Pasteur au Sida*. Paris: Presses Universitaires de France.

Podolsky, Scott H., and Alfred I. Tauber (1997), *The Generation of Diversity: Clonal Selection Theory and the Rise of Molecular Immunology*. Cambridge, MA: Harvard University Press.

Silverstein, Arthur (1989), *A History of Immunology*. San Diego: Academic Press.

Stewart, John (1994a), "Cognition Without Neurons: Adaptation, Learning and Memory in the Immune System," *Communication and Cognition—Artificial Intelligence* 11: 7–30.

——— (1994b), *The Primordial VRM System and the Evolution of Vertebrate Immunity*. Austin, TX: R. G. Landes.

Tauber, Alfred I. (1994), *The Immune Self: Theory or Metaphor?* New York and Cambridge: Cambridge University Press.

——— (1996), "The Molecularization of Immunology," in S. Sarkar (ed.), *The Philosophy and History of Molecular Biology: New Perspectives*. Dordrecht, Netherlands: Kluwer Academic Publishers, 125–169.

——— (1998), "Conceptual Shifts in Immunology: Comments on the 'Two-Way Paradigm,'" *Theoretical Medicine and Bioethics* 19: 457–473.

——— (1999), "The Elusive Self: A Case of Category Errors," *Perspectives in Biology and Medicine* 42: 459–474.

——— (2000), "Moving Beyond the Immune Self?" *Seminars in Immunology* 12: 241–248.

Tauber, Alfred I., and Leon Chernyak (1991), *Metchnikoff and the Origins of Immunology: From Metaphor to Theory*. New York and Oxford: Oxford University Press.

Varela, Francisco J., Antonio Coutinho, B. Dupire, and N. N. Vaz (1988), "Cognitive Networks: Immune, Neural, and Otherwise," in Alan S. Perelson (ed.), *Theoretical Immunology*, part 2. Redwood City, CA: Addison-Wesley, 359–375.

Wilson, Jack (1999), *Biological Individuality: The Identity and Persistence of Living Entities*. Cambridge: Cambridge University Press.

INCOMMENSURABILITY

Incommensurability is a relation of incomparability, or limited comparability, purported to obtain between some pairs of successive or competing scientific theories. The thesis that scientific theories may be incommensurable was proposed by Paul Feyerabend and Thomas Kuhn in separate publications in 1962 (see Feyerabend, Paul; Kuhn, Thomas). Due to perceived negative consequences of incommensurability, the thesis has been the focus of considerable controversy. Before considering the objections to it, the thesis of incommensurability will first be examined.

Feyerabend on Incommensurability

Feyerabend's claim that some theories are incommensurable derives from his critique of the empiricist idea of a theory-neutral observation language. Neither experience nor pragmatic conditions of use determine the meaning of observational terms. Instead, "the interpretation of an observation language is determined by the theories which we use to explain what we observe, and it changes as soon as those theories change" (Feyerabend [1958] 1981, 31). In contrast to the empiricist view that the meaning of observational terms is independent of theory, Feyerabend holds that the meaning of such terms varies with theory.

Feyerabend introduced the concept of incommensurability in the context of a discussion of the empiricist account of inter-theory reduction by means of deductive subsumption. Against reduction, Feyerabend argues:

> What happens ... when a transition is made from a theory T′ to a wider theory T (which ... is capable of covering all the phenomena that have been covered by T′) is something much more radical than incorporation of the *unchanged* theory T′ (unchanged, that is, with respect to the meanings of its main descriptive terms as well as to the meanings of the terms of its observation language) into the context of T. What does happen is, rather, a *replacement* of the ontology (and perhaps even of the formalism) of T′ by the ontology (and formalism) of T, and a corresponding change of the meanings of the descriptive elements of the formalism of T′ (provided these elements and this formalism are still used). This replacement affects not only the theoretical terms of T′ but also at least some of the observational terms which occurred in its test statements. (Feyerabend [1962] 1981, 44–45)

For Feyerabend, change in theoretical ontology leads to variation in the meaning of the vocabulary employed by theories. One theory cannot be deductively subsumed by the other, given differences in the meaning (due to untranslatability) of the terminology employed by the theories.

According to Feyerabend, reduction fails because of incommensurability. Theories are incommensurable due to lack of semantic equivalence between terms employed by the theories. On the one hand, the concepts of one theory cannot be defined on the basis of concepts of the other. On the other hand, no empirical statement may be formulated that correlates terms of one theory with terms of the other theory. Because no neutral observation language exists in which to express the empirical consequences of such theories, Feyerabend concludes that "incommensurable theories may not possess any comparable consequences, observational or otherwise" (Feyerabend [1962] 1981, 93). The contents of incommensurable theories are unable to be compared because no consequence of one theory may either assert or deny the same thing as any consequence of a theory with which it is incommensurable (see Feyerabend, Paul).

Incommensurability in Kuhn's *Structure of Scientific Revolutions*

In *The Structure of Scientific Revolutions*, Kuhn (1962) proposed a model of the development of science divided into periods of normal science grounded in consensus on a shared scientific paradigm. Normal science is broken at intervals by periods of extraordinary science, brought on by anomaly and crisis, which may ultimately result in revolutionary displacement of a paradigm (see Scientific Revolutions).

Once a new candidate for paradigm emerges in the midst of a crisis, debate ensues between defenders of the reigning paradigm and advocates of the candidate paradigm. This debate is characterized by failure of communication that arises because of the incommensurability of the old paradigm and the new candidate paradigm. As a result of incommensurability, debate about which paradigm to adopt is unable to be brought to closure by purely rational means.

According to Kuhn, the incommensurability of competing paradigms is due to differences that arise at three levels between paradigms. The first difference involves variation at the methodological level. Paradigms address different problem-solving agendas and employ different standards of theory appraisal:

> [P]roponents of competing paradigms will often disagree about the list of problems that any candidate for paradigm must resolve. Their standards or their definitions of science are not the same. (Kuhn 1962, 148)

The second difference is at the semantic level. There is variation in the concepts employed by paradigms, which leads to change in the meanings of the terms that express key scientific concepts:

> Within the new paradigm, old terms, concepts, and experiments fall into new relationships one with the other. . . . To make the transition to Einstein's universe, the whole conceptual web whose strands are space, time, matter, force, and so on, had to be shifted and laid down again on nature whole. (149)

The third difference relates to the theory-dependence of observation. Not only may scientists observe different things, but the content of their perceptual experience when they observe the same thing depends upon the paradigm in which they work:

> [P]roponents of competing paradigms practice their trades in different worlds. . . . [P]racticing in different worlds, the two groups of scientists see different things when they look from the same point in the same direction. (150)

Kuhn's claim that scientists work in different worlds may be taken to suggest a stronger thesis than that which states that scientists' perceptual experience depends on paradigm. In his book *Reconstructing Scientific Revolutions*, Paul Hoyningen-Huene (1993) has argued that Kuhn's position is best understood as a neo-Kantian position in which the phenomenal world of scientists varies with paradigms, while the unknowable noumenal world remains constant (see Kuhn, Thomas).

Taxonomic Incommensurability

In later work, Kuhn continued to refine his concept of incommensurability. In contrast to Feyerabend, Kuhn's original concept of incommensurability included nonsemantic elements. Kuhn came to view incommensurability as a semantic issue distinct from methodological variation and dependence of observation on theory. Semantic issues relating to translation failure are the focus of Kuhn's later work on incommensurability, of which he proposes a taxonomic version that involves localized translation failure between subsets of the special terminology employed by theories.

In *The Road Since Structure*, Kuhn (2000) claims that scientific revolutions are characterized by changes in the taxonomic schemes by means of which theories classify entities in their domain (30). In the transition between theories, both criteria of classification and membership of taxonomic categories undergo change. At the semantic level, taxonomic change gives rise to variation in meaning of some of the preserved vocabulary, as well as to introduction of vocabulary with new meaning. Because taxonomic change involves change of interconnected categories, the meanings of the terms affected by such change are related in a holistic manner. Each theory possesses a central set of interdefined terms, which cannot be translated in piecemeal fashion into the vocabulary of a theory with a different taxonomic structure (Kuhn 2000, 43–44). Translation failure between theories is a localized phenomenon that is restricted to such central sets of interdefined terms.

Objections to Incommensurability

As indicated above, the incommensurability thesis is controversial because of negative outcomes to which it gives rise. If, as Kuhn initially suggested, there are no neutral standards of theory appraisal, and communication is obstructed, it is unclear how choice between theories may proceed on a rational basis. If, as Kuhn and Feyerabend both suggest, the content of theories may not be compared due to semantic variance, it is unclear how to conduct crucial tests between rival theories or to determine whether one theory marks an advance over another. Indeed, Dudley Shapere raises the question of whether incommensurable theories may constitute rivals at all (Shapere 1984, 73). But the two objections that have proven the most telling have been Donald Davidson's critique of untranslatability and Israel Scheffler's referential objection to incomparability.

The Incoherence of Untranslatability

Davidson raises serious doubts regarding the coherence of the idea of an untranslatable language. He notes that there is an air of paradox about incommensurability: "Kuhn is brilliant at saying what things were like before the revolution using—what else?—our post-revolutionary idiom" (Davidson 1984, 184). If one provides an example of an untranslatable concept in the language into which translation fails, the example belies the untranslatability. It is also puzzling how one might understand an untranslatable concept in the first place if it cannot be translated into a language that one understands. It is not, moreover, clear what would count as evidence of untranslatability. Failure to translate a language is indeterminate between being evidence that the language is untranslatable and evidence that it is not a language at all. Davidson suggests that the idea of an untranslatable language depends on a distinction between conceptual scheme and content, which gives substance to the idea of a language independent of translation. But, he argues, no intelligible sense can be made of the distinction between scheme and content.

Davidson's objections may be defused by noting two ways in which the incommensurability thesis is less extreme than he supposes. First, failure of translation between incommensurable theories is restricted to the vocabulary employed by theories, or to a subset of such vocabulary, rather than extending to the entirety of a natural language. Thus, the thesis of incommensurability does not require sense to be made of radically alternative conceptual schemes, but only of localized translation failure within a language. Nor need untranslatable concepts be formulated in the language into which translation fails. Untranslatability is restricted to semantically variant fragments of an embracing natural language. The latter may therefore serve as metalanguage within which semantic relations between the vocabulary of meaning-variant theories may be analyzed. Second, incommensurability need entail only failure to translate between the vocabulary of theories, rather than failure to understand the content of a meaning-variant theory. One may understand what is said in another language even if it cannot be translated into one's own language. Equally, one may understand concepts of a theory that are untranslatable into one's own theory due to incommensurability.

The Referential Objection

In his book *Science and Subjectivity*, Scheffler (1967) notes that discussion of meaning variance in relation to incommensurability runs foul of the distinction between sense and reference. Variation in theoretical context may lead to variation in the sense of a scientific term. But it does not follow that the term's reference is thereby similarly affected. Terms that differ in sense may refer to the same thing. Coreference is all that is needed for claims about the world to enter into conflict. Hence, the content of meaning-variant theories may be compared if the terms employed by the theories share common reference, regardless of variation in sense.

However, Scheffler's referential objection is not entirely successful. Meaning variance in science need not be restricted to variation in the sense of scientific terms. If reference is determined by sense, then significant variation in sense may result in variation of reference. Moreover, the reference of scientific terms may be subject to variation independent of sense as a result of changes in classification or revision of linguistic use.

The Causal Theory of Reference

Since the 1970s, the referential objection has been based on the causal theory of reference. Causal theorists argue that the reference of a term is not determined by an associated description that specifies the term's sense. Rather, reference is determined in a direct manner by ostensive introduction of a term in the presence of the referent. What determines reference is the causal relation (e.g., perception) between term-introducer and referent. Subsequent use of the term by later speakers is connected by means of a causal-historical chain to the original term-introduction.

If reference is determined independently of description, then terms employed by one theory may be employed in the context of a later theory to refer to the same things as they referred to in the earlier theory. Terms employed by successive theories may continue to refer to the same things despite variation in descriptive content associated with the terms in different theories. The claims made by such theories may be compared directly on the basis of the shared referents of the terms that the theories employ.

But application of the causal theory of reference in the context of scientific theory change is not without difficulties. First, to secure reference for a kind-term, it does not suffice to identify sample members of the kind in an ostensive manner. The sample may belong to multiple kinds. Ostension must be supplemented by a description that specifies the relevant kind by means of a sortal expression. Second, theoretical terms may fail to refer if the unobservable entities to which they purport to refer do not in fact exist. But if reference is determined by the causal relation to the real cause of an actual phenomenon, then it may be impossible for reference to fail. To allow for such failure, descriptive characterization of putative theoretical entities must enter into the reference determination of theoretical terms. Third, to allow reference change, reference must be sensitive to the use of terms on occasions subsequent to their initial introduction, rather than being permanently fixed at the outset.

In light of the need for a descriptive element in the determination of reference, recent authors propose causal descriptive accounts of reference in which either causal relation combines with description to fix reference or else reference-fixing description is cast in causal terms. Causal descriptive accounts allow that the descriptive content of scientific theories affects reference. However, such accounts provide little scope for incommensurability due to radical divergence of reference. For while description may play a role in reference determination, neither is reference fully determined by description, nor is the entirety of the descriptive content associated with a term relevant to reference. Thus, so far as variation of reference is concerned, the prospects for incommensurability are greatly diminished.

HOWARD SANKEY

References

Bird, Alexander (2000), *Thomas Kuhn*. Chesham, UK: Acumen Publishing.

Davidson, Donald (1984), *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Feyerabend, Paul K. ([1958] 1981), "An Attempt at a Realistic Interpretation of Experience," in *Realism, Rationalism and Scientific Method: Philosophical Papers*, vol. 1. Cambridge: Cambridge University Press, 17–36. Originally published in *Proceedings of the Aristotelian Society, New Series* 58, 143–170.

——— ([1962] 1981), "Explanation, Reduction and Empiricism," in *Realism, Rationalism and Scientific Method: Philosophical Papers*, vol. 1. Cambridge: Cambridge University Press, 44–96. Originally published in Herbert Feigl and Grover Maxwell (eds.), *Scientific Explanation, Space and Time: Minnesota Studies in the Philosophy of Science*, vol. 3. Minneapolis: University of Minnesota Press, 28–97.

Hoyningen-Huene, Paul (1993), *Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science*. Chicago: University of Chicago Press.

Hoyningen-Huene, Paul, and Sankey, Howard (eds.) (2001), *Incommensurability and Related Matters: Boston Studies in Philosophy of Science*, vol. 216. Dordrecht, Netherlands: Kluwer Academic Publishers.

Kuhn, Thomas S. (2000), *The Road Since Structure*. Edited by James Conant and John Haugeland. Chicago: University of Chicago Press.

——— (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Preston, John (1997), *Feyerabend: Philosophy, Science and Society*. Cambridge: Polity Press.

Sankey, Howard (1993), "Kuhn's Changing Concept of Incommensurability," *British Journal for the Philosophy of Science* 44: 775–791.

——— (1997), *Rationality, Relativism and Incommensurability*. Aldershot, UK: Ashgate.

——— (1998), "Taxonomic Incommensurability," *International Studies in the Philosophy of Science* 2: 7–16.

——— (1994), *The Incommensurability Thesis*. Aldershot, UK: Avebury.

Scheffler, Israel (1967), *Science and Subjectivity*. Indianapolis: Bobbs-Merrill.

Shapere, Dudley (1984), *Reason and the Search for Knowledge*. Dordrecht, Netherlands: Reidel.

See also **Feyerabend, Paul; Kuhn, Thomas; Logical Empiricism; Scientific Change; Unity and Disunity of Science**

INDETERMINISM

See **Determinism; Quantum Mechanics**

INDIVIDUALISM

See **Methodological Individualism**

INDIVIDUALITY

One of the most fundamental distinctions in philosophy is between individuals (also called particulars) and such things as classes, sets, kinds, universals, or whatever. As the number of synonyms might indicate, philosophers have spent much more effort clarifying classes than they have explicating the polar notion of individuals. In the briefest form of this relationship, classes range over individuals, and individuals belong to one or more classes. For example, in the claim that Socrates is mortal, Socrates is an individual human being, while ‘mortal’ denotes a class of beings all of whom are born and die. Socrates is mortal, but so are all other living creatures.

Philosophers have dealt with the distinction between individuals and classes in two ways. First, some have relied on ordinary folk notions of individual and class. What ordinary people take to be individuals are individuals. What ordinary people take to be classes are classes. Socrates is an individual, and ‘mortal’ is a class. For most philosophers, however, ordinary notions of individual and class are not good enough. They have devised technical notions of individuals and classes to conform strictly to the needs of their technical philosophies: for instance, a bare particular and the class of all bare particulars. A bare particular is an individual that has no properties of its own, an entity shorn of all characteristics. Needless to say, bare particulars are not common sense entities.

Biological Individuals

Although most of the entities that biologists treat as individuals are common sense individuals, many are not. Ordinary people consider a Portuguese man-of-war a single organism—a single individual. Biologists do not. Like philosophers, biologists have had to develop their own technical notions of individuality to fulfill their own special needs. The analyses of the ‘individual’ produced by biologists have several advantages over comparable philosophical accounts. All such analyses must take place in some context or other. In philosophy these contexts are supplied by very general philosophical systems, while in biology they take place within specific scientific theories. Unfortunately for philosophers, none of their systems have gained much in the way of consensus. Scientists have the advantage that many scientific theories are widely accepted. Right now, evolutionary biology is going through some fundamental revisions (see Evolution). Even so, there is more agreement among evolutionary biologists about the evolutionary process than among philosophers with respect to their systems. As a result, biologists’ notions of the individual are likely to have more content and lasting influence than such notions devised by philosophers.

A second advantage that scientists, particularly biologists, have over philosophers is the wealth of examples open to them. Philosophers have a habit of making up science fiction examples to illustrate and test their analyses. Unfortunately, such examples are highly malleable; philosophers can make them serve just about any purpose they desire. Real examples, on the contrary, can force a reexamination of the decisions that are made. They can force one to see that some of one’s deepest intuitions are mistaken. Nature provides much more bizarre examples than any philosopher has ever been able to dream up, and—more importantly—it provides good reason to accept one analysis over another.

Nihilists can be found claiming that how one divides up the world simply does not matter. This claim is usually made in the context of classes. Group females of one species with males of another. It makes no difference. Ignore the distinction between sexual and asexual organisms. Comparable claims are also made with respect to individuals. Consider a clone of a single organism or thousands. It simply does not matter. For scientists, it does matter. For scientists, certain ways of dividing up the world are preferable to others. Perhaps scientists cannot always settle on one way and only one way of dividing up the world, but from this state of affairs it does not follow that anything goes. In actual fact, developing alternative classifications is quite difficult. Only a very few serious alternatives can be found.

In biology at least, biologists and philosophers have pooled their conceptual resources to deal with topics such as individuality. For example, biologists commonly use "individual" and "organism" interchangeably, but such a linguistic convention leads to all sorts of confusion. All organisms are individuals, but not all individuals are organisms. "Individual" is a much broader term, and one needs such a broad term. A running argument in the biological literature concerns the levels of organization at which selection can occur (Keller 1999; Michod 1999). Those who think that organisms are the primary units of selection are called "individual selectionists," because organisms are individuals; but another group of biologists think that genes are the primary units of selection, and they too are called "individual selectionists," because genes are just as much individuals as organisms are. Of course, genes are not organisms. To avoid such confusion, "individual" is used here in a generic sense to refer not only to organisms but also to other individuals.

The literature on individuality as this notion is used in biology is replete with all sorts of bizarre problem cases. A few years ago, several mycologists discovered a clonal population of the fungus *Armillaria bulbosa* that occupied fifteen hectares in the Upper Peninsula of Michigan. As far as these mycologists could tell, all parts of this clonal entity were attached and had the same genome. Why not think of this huge fungus as a single organism or, if not a single organism, at least a single individual (Gould 1992; Wilson 1999)? To answer this question (and others), philosophers and biologists have set out a list of criteria for an entity counting as an individual in the generic sense and the implications that these criteria have for a variety of problem cases. But before these criteria are examined, a word needs to be said about the importance of developing a clear notion of individuality. As real as the fungus is, one might wonder why deciding whether or not it is one organism or a million different organisms is of any significance (Hull 1992; Wilson 1999).

The Cost of Meiosis

Such questions are important because counting is a central activity of scientists, and they have to know what it is they are counting. Like must be counted with like. For example, one commonly hears that sexual reproduction is extremely prevalent and that this prevalence poses a problem for evolutionary biologists. In sexual reproduction, homologous chromosomes line up at meiosis and separate in the formation of germ cells. At every locus where different alleles reside, sexual reproduction has a 50 percent cost. More generally, asexual organisms can pass on all their genetic material to the next generation, whereas each sexual organism can pass on only half. A 50 percent cost is extremely high. The usual conclusion is that sexual reproduction must be doing a lot of good, or else it would not be so prevalent.

If meiosis is so costly, why is it so prevalent? Several answers to this question have been suggested, but one issue is passed over too quickly. Is sexual reproduction actually all that prevalent, compared with asexual reproduction? This difference is usually presented in terms of species. Sexual species are vastly more prevalent than asexual species. But there are some difficulties here. The most popular definition of species in the past century or so includes reference to interbreeding. Those organisms that produce a genealogical nexus by mating with each other form species. Since asexual organisms do not form such a network, they do not form species. Hence, contrasting the number of sexual species with the number of asexual "species" is illicit.

Only if one resorts to defining species in terms of character distributions do sexual and asexual organisms form the same sort of species. However, the concept of morphospecies has numerous problems. At one time systematists thought that something called "overall similarity" existed out there in nature and that one degree of overall similarity could be found that was the same across all organisms. Such is not the case. Various degrees of similarity exist, and no reason can be found for choosing one degree of similarity over any of the others as the level of species. Sexual organisms produce biologically significant units, but these units are not comparable to anything found in asexual organisms. Comparable units of overall similarity can be found in both, but these units are not biologically significant. In short, comparable units at the traditional level of species cannot be found for both sexual and asexual organisms.

One way out of this difficulty is to move down to the level of organisms. Sexual and asexual organisms may not form species of the same kind, but they are 'organisms' in the same sense. Hence, the prevalence of sexual reproduction can best be determined at the level of organisms, not species. When one makes such a conceptual shift, the relative percentages change. Sexual reproduction ceases to be so overwhelmingly prevalent. For the first half of life on earth, no sexual reproduction occurred. Since then it has become increasingly prevalent, but not as prevalent as is commonly claimed. In comparing organisms with organisms, problems with the cost of meiosis remain, but their scope is greatly reduced.

Units of Selection

Are there units of selection? and if so, what are they? The different answers to these questions result from yet another failure to make necessary distinctions. Selection is not one process but two intricately connected processes—replication and environmental interaction. In replication, information is passed on from one generation to the next. Certain entities also interact with their environment in such a way that replication is differential. If selection is to occur, both replication and environmental interaction are necessary. What are the units of replication? By and large, replication is limited to the genetic material, though in special circumstances replication can occur at higher levels of organization. What are the units of environmental interaction? Environmental interaction can occur at a variety of levels, from single genes and cells through organisms and hives to demes and possibly entire species.

The discussion above mentions no entities that are uniquely units of selection. There are units of replication and units of environmental interaction, but no units of selection. When biologists such as Dawkins (1976) claim that genes are units of selection, they are actually claiming that genes are the units of replication. When Dawkins's opponents claim that selection occurs at a variety of hierarchically organized levels, they are referring to environmental interaction. The issue here is individuality—what counts as individuals and what roles these individuals play in the evolutionary process. There are entities that function as replicators, and there are entities that function in environmental interaction; but there are no entities that function as "selectors."

Species as Individuals

Species present another example of the importance of individuality in biology. According to Ghiselin (1974) and Hull (1976), species are not classes but historical entities. As such, they exhibit all the characteristics of individuals, not of classes. For example, species can go extinct. Classes cannot. A class can temporarily have no individuals exemplifying it; that is all. For example, during the first few moments of the big bang, no heavy elements existed. Through time, elements with higher atomic numbers came into existence here and there in the universe. It is possible that at any one moment, gold could cease to exist. Then, later, additional atoms of gold could reemerge.

Such occurrences pose no problems for physicists, but what about the emergence, extinction, and reemergence of biological species? Could not dinosaurs reevolve? For species as historical entities, continuity through time is required. Once a taxon as a monophyletic unit goes extinct, it cannot reevolve; once extinct, no taxon can come into existence again. This claim depends not on empirical contingencies but on the individuation of species in terms of descent. For an atom to count as a gold atom, no historical connections are required. For evolving lineages, such connections are necessary. Descent is required because natural selection requires descent. Natural selection is not the only mechanism involved in the evolution of species, but it is certainly the main mechanism.

Reinterpreting species as historical entities has numerous important implications for an understanding of the evolutionary process. Critics of present-day biology have made much of the contention that biology, unlike physics, has no laws. As an example, they cite the millions upon millions of claims made by biologists about particular species. It is often heard that all swans are white, all ravens are black, and human beings are rational animals. All these claims inevitably have exceptions because evolution proceeds by means of variations, and these variations cannot be explained away as monsters. Variability is essential to the evolutionary process. However, if species are individuals, this is exactly as they should be. 'All swans are white' is no more a candidate for a law of nature than 'the earth is the third planet from the sun.' If biology includes any genuine laws of nature, they must be found elsewhere.

Criteria for Individuality

Outside biology, the criteria for individuality are frequently quite strong—so strong that no biological entities can meet these standards. For example:

(a) *Same substance.* Individuals must retain the same substance throughout their existence. Nothing in the living world fulfills this requirement. For example, organisms such as people exchange their substance many times over during the course of their existence. Even nerve cells are lost, gained, and reconfigured, albeit quite slowly.

(b) *Same form.* Individuals must retain the same form throughout their existence. Individuating forms is far from easy, but no matter how one deals with this knotty issue, many organisms undergo dramatic metamorphosis during the course of their development. A caterpillar and the butterfly into which it develops do not share the same form. The choice is either to consider these two as stages in the life cycle of a single organism or to consider the caterpillar a separate organism from the butterfly it produces. The problem is only magnified when a single organism produces numerous individuals at a later stage in its development.

Perhaps some individuals in physics are composed of the same substance and exhibit the same form throughout their existence, but nothing fulfills these requirements in biology. Two additional criteria do apply equally inside and outside of biology:

(c) *Spatiotemporal localization* (*boundedness*). Individuals have locations as well as beginnings and endings in space and time. These beginnings and endings can be sharp or fuzzy. Individuals can cease to exist terminally, but they also can cease to exist by becoming another individual. That is what happens when one cell splits into two cells or one species splits into two species.

(d) *Spatiotemporal continuity.* In criteria (a) and (b) above, no change whatsoever can take place in substance or essential form. However, all sorts of things change in the natural world. If the entity involved is to count as the same individual throughout, the change must be continuous, even if it is not gradual. An organism, when it undergoes metamorphosis, may proceed from stage to stage quite abruptly, but it remains the same organism because the change is continuous. As a result, numerically, the same individual cannot come into existence more than once.

(e) Spatiotemporal localization and continuity define what is commonly called a "historical entity" (Wiggins 1967). Such entities can be found both in and outside biology. For example, organisms are clear cases of historical entities, but so are planets and stars. As long as the nine solar planets keep revolving around the sun, each will remain the same historical entity. However, if a very large comet were to smash into Pluto, that event might well be the end of both of them.

Biologists are not content to limit themselves to historical entities of the more generic sort. Instead, they add more criteria, which characterize few if any individuals outside biology. For example:

(f) *Structural heterogeneity.* Biological individuals are structurally heterogeneous. Outside biology, certain entities approach structural homogeneity (e.g., the proverbial billiard ball), but even the simplest biological entity, such as a single codon, exhibits structural heterogeneity. However, the most important sort of structural heterogeneity is that which plays a role in functional organization.

(g) *Functional organization.* This facility is limited to human artifacts and evolved biological organisms. Although there is considerable latitude for fluctuations in functionally organized systems, eventually this organization can be destroyed. For highly organized systems, just about any modification results in loss of function. In others, especially those that exhibit modular organization, considerable disruption can be tolerated (Buss 1987). For example, if a lobster were torn into half a dozen pieces and the pieces thrown back into the ocean, all of these pieces would die. A lobster can regenerate a leg or two, but little more. However, if the same thing happened to a starfish, one would be likely to get six new starfish.

(h) *Genetic homogeneity.* Just as biological individuals are quite heterogeneous internally, they are quite homogeneous genetically. In order for multicellular organisms to develop, some way had to be found to allow cells

> to cooperate. The roadblocks that hinder entities with different genetic constitutions from cooperating are well known, especially as a result of Richard Dawkins's famous book *The Selfish Gene* (1976). One way around this roadblock is to have all the genes in a multicellular organism genetically identical.

The criterion of genetic homogeneity prevents entities more inclusive than single organisms from functioning as units in the evolutionary process. A beehive can fulfill all the criteria listed above save for (a) and (b), sameness of substance and form, but not all the individual bees living in the same hive contain the same genes. In particular, the drones that succeed in mating with the queen differ from her and from the workers. The entire hive might succeed in functioning as a unit of selection, but not one that is as efficient as it should be.

Kinds of Biological Individuals

Given these considerations, Wilson (1999, 60) distinguishes four sorts of biological historical entities:

1. *Functional individuals* are entities whose heterogeneous parts are currently so causally integrated that they tend to return the individual to the same state in the face of appreciable though not unlimited alterations.
2. *Developmental individuals* are entities that are programmed to develop through time. Whereas functional individuals return to a preferred state, developmental individuals proceed from state to state.
3. *Genetic individuals* are entities that possess the same genetic makeup derived from descent from a common ancestor. For example, all the descendants of a single zygote count as genetic individuals.
4. *Units of evolution* are entities that play important roles in evolutionary change, for example, units of replication and environmental interaction (Brandon and Burian 1984).

As a result of this analysis of biological individuality, certain "individuals" do not accord with common sense notions of individuality. For example, a clone counts as a single genetic individual even though it may be made up of hundreds of distinct developmental individuals. In such situations, growth is indistinguishable from reproduction. If the clone is considered a single individual, then the production of additional individuals counts as growth. If, however, the clone is thought of as being made up of independent organisms, then the very same changes count as asexual reproduction (see, e.g., Jackson, Buss, and Cook 1986).

This entry has concentrated on only one half of the polar concepts of 'individual' and 'class.' Several fascinating issues concerning the relation between individuals and their kinds have been ignored (Lowe 1989; Wiggins 1967). For example, can an individual remain the same individual when it changes from one kind to another? Some philosophers argue that an entity's ability to change from one kind to another proves that these were not genuine kinds in the first place. For example, in certain species of fish, a female can change into a male with only minor alterations. It would seem rather strange to consider this fish two distinct individuals in the process. Hence, it would seem that male and female are not genuine kinds, but they play this role in several areas of biology. Here is but one more instance in which real examples can make a difference.

David L. Hull

References

Brandon, R., and R. Burian (eds.) (1984), *Genes, Organisms, Populations: Controversies over the Units of Selection*. Cambridge, MA: MIT Press.

Buss, L. (1987), *The Evolution of Individuality*. Princeton, NJ: Princeton University Press.

Dawkins, R. (1976), *The Selfish Gene*. Oxford: Oxford University Press.

Ghiselin, M. T. (1974), "A Radical Solution to the Species Problem," *Systematic Zoology* 23, 536–544.

Gould, S. J. (1992), "A Humungous Fungus Among Us," *Natural History*, July, 10–16.

Hull, D. L. (1976), "Are Species Really Individuals?" *Systematic Zoology* 25, 174–191.

———. (1992), "Individual," in E. Fox Keller and E. A. Lloyd (eds.), *Keywords in Evolutionary Biology*. Cambridge, MA: Harvard University Press, 180–187.

Jackson, Jeremy B. C., Leo W. Buss, and Robert E. Cook (eds.) (1986), *Population Biology and Evolution of Clonal Organisms*. New Haven, CT: Yale University Press.

Keller, L. (ed.) (1999), *Levels of Selection in Evolution*. Princeton, NJ: Princeton University Press.

Lowe, E. L. (1989), *Kinds of Being: A Study of Individuation, Identity, and the Logic of Sortal Terms*. Oxford: Basil Blackwell.

Michod, R. E. (1999), *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*. Princeton, NJ: Princeton University Press.

Wiggins, D. (1967), *Identity and Spatio-Temporal Continuity*. Oxford: Oxford University Press.

Wilson, J. (1999), *Biological Individuality: The Identity and Persistence of Living Entities*. Cambridge: Cambridge University Press.

See also **Biological Information; Evolution; Natural Selection; Species**

PROBLEM OF INDUCTION

The Scottish philosopher David Hume (1711–1776) first focused attention on the question of what the grounds are for believing that the future will resemble the past or, more generally, that what has not been observed will resemble what has (Hume 1965 and 1999). Inference in which one takes the past as grounds for beliefs about the future, or the observed as grounds for beliefs about the unobserved, or in general an inference that is ampliative—having more content in the conclusion than in the premises—has come to be called *induction.* ("Induction" is also sometimes used in a more specific sense to refer to induction by enumeration, the inference in which one simply generalizes from instances to all cases or to the next case.) In addition, the powerful arguments that led Hume to suppose that there was no rational ground whatsoever for inference from the observed to the unobserved are difficult to answer. Hume's question has thus come to be called the *problem* of *induction.*

Hume divided all reasoning into two mutually exclusive types, that concerning relations between ideas and that concerning matters of fact and existence. All mathematical and logical reasoning fell into the former category—also called "demonstration," what would now be called "deduction"—and was regarded as unproblematic, though also nonampliative. Reasoning about relations between ideas could not suffice for natural science, which frequently makes inferences from the observed to the unobserved in making predictions, retrodictions, and generalizations, which count as Hume's second sort of reasoning. Hume asks what these ampliative inferences are based on, and he concludes that all of them are founded on beliefs about relations of cause and effect. The question of how to justify ampliative inferences thereby becomes the question of how to justify judgments about cause and effect.

One thread of Hume's further argument depends to some extent on his empiricism and his psychological views. He is committed to the views that all ampliative knowledge comes from experience and that experience is composed entirely of impressions. Thus when he asks what one can know of cause and effect, he asks what one experiences of it and answers with what one can have an impression of. He considers one billiard ball hitting another and points out that while people seem to think that there are three things here—cause, effect, and necessary connection between the two—one has impressions only of the first two, the first ball hitting the second, and the second moving. Thus, the only possible way of knowing that this effect must follow the cause, as opposed to merely that it happens to have done so in the past, gives no grounds for believing in this necessity or connection.

Hume's main argumentation does not depend on special assumptions about causation or the psychology of experience, and this is the version of his argument that has received the most attention from philosophers of science subsequently. He claims first that the inference from the collision of the first ball with the second to the belief that this will be followed by the second ball moving cannot be made by demonstration. This is because, in a demonstration, supposing the premises true and the conclusion false is always a contradiction, but there is no contradiction in supposing that after the first billiard ball hits the second, the second ball rises one foot and levitates. There is no contradiction, that is, in supposing that the second ball will not do what it usually does. By assumption, the only other form of reasoning by which the usual inference that the second billiard ball will move could be justified is the very reasoning about matters of fact that Hume was seeking a justification of. To appeal to such reasoning at this juncture would be circular. Hume concludes that the expectation that the future will resemble the past has no rational justification and is based only on instinct and custom.

To be sure, the inference to the conclusion that the second billiard ball will move in the usual way on being hit by the first could be understood as a deduction with a suppressed premise, also known as an *enthymeme.* The suppressed premise on which one's confidence depends in this view would be a claim that nature is uniform—a "uniformity of nature" assumption—and that regularities that have been observed to hold up to now will continue to hold. However, this strategy is ineffectual unless one can justify the premise that nature is uniform. One cannot show by demonstration from the claim that what has been seen of nature is uniform that nature's as yet unseen parts are uniform, because denying the latter does not contradict the former. However, one would have to know the latter to know that nature is uniform. One cannot in this context appeal to reasoning about matters of fact to defend the claim that all of nature is uniform, and not only the parts already seen. That would be circular because the assumption about the uniformity of nature was supposed to justify all of our reasoning about matters of fact. This strategy for answering Hume's question is thus best understood as an alternative way of presenting Hume's problem.

The uniformity-of-nature assumption presented has further problems. One is that not all regularities that have been observed to hold up to now will continue to hold. One does not expect regularities that are regarded as coincidences to continue, and one recognizes at once the folly of a person who jumps out of a window on the fortieth floor and when passing the tenth floor says, "So far so good." The uniformity-of-nature assumption can be made weaker, to say that everything that happens is an instance of some exceptionless general law, a claim that resembles Kant's response to Hume about causation. It may be that such a claim could be argued for on a priori grounds, but since this principle would provide no basis for identifying which events are invariably followed by which others, it also would provide no basis for distinguishing sound from unsound inductions. If the uniformity-of-nature assumption were, on the other hand, so specific as to identify the regularities that will continue, it would gain content at the expense of being difficult to defend without circularity.

Twentieth-Century Responses

There are three popular ways of responding to Hume's question by rejecting the problem. The first, the explicit formulation of which is due to Strawson, though it was also suggested by the Ayer (1952), says that attention to the meaning of 'reasonableness' shows that the supposed problem is merely a misunderstanding (Strawson 1952) (see Ayer, Alfred Jules). Induction is part of the standards for reasonableness, so the question why *it* is reasonable does not, strictly speaking, make sense. To ask why it is reasonable to make inductive inferences is to ask why it is reasonable to be reasonable, a question to which one should not expect an answer. However, the problem is not so easily dissolved. A reasonable reply to this line of argument is that the problem of induction is not generated by asking whether induction is acceptable according to itself, but rather by asking whether the inductive standards of reasonableness that are in fact employed are likely to serve the end of making true predictions as often as possible, and what grounds one has for thinking so. Thus, one is not asking why it is reasonable to be reasonable, but why it is reasonable to think that these standards serve (one of) their purpose(s) of making true predictions.

A second way of rejecting the problem is to protest that the claim that justifying induction is a problem rests on a mistaken demand for a guarantee where, as Hume has shown, a guarantee is logically impossible. This response is commonly invoked to say that it is only by holding the bar too high that one can think there is a problem of justification about induction; it is only by expecting induction to be deduction, to yield certainty, that one sees a problem when it is not and does not. However, this line involves an erroneous understanding of what Hume's arguments purport to show—for, it is not argued merely that there is no deductive (demonstrative) guarantee that future events will behave like past events. It is argued rather persuasively that there can be no rational ground whatsoever. That is, one would be happy if one could show even that a great number of occasions on which *A* is associated with *B* make it more *likely* than it would have been without those instances, that *A* will be associated with *B* in the future (Russell 1959). Hume argues that there can be none but a question-begging justification of this probability claim.

A third complaint about Hume's problem says that he did not concern himself with the kinds of complex inferences scientists actually engage in, and skeptical conclusions about the simple induction by enumeration that he had in mind are thus irrelevant to the justification of scientific claims, whatever other significance they may have. However, whether or not Hume had enumerative induction explicitly in mind, the applicability of his main argumentation is by no means restricted to that. Hume's main argument applies to any ampliative inference, which the complex scientific inferences referred to certainly involve. For any ampliative inference, it will be the case that no demonstration can secure it, for a conclusion that goes beyond the premises in content could well be different in the content that goes beyond the premises without contradicting the premises. And ampliative inference cannot be used to justify ampliative inferences because that would be relying on the very sort of inference that is in need of justification. Appeal to the complexity of real science does not erase Hume's problem.

In the 1950s, a popular response to Hume's problem was to question easy acceptance of the claim that induction could not legitimately justify itself. Black (1954) pointed out that if one appeals to the past success of induction to justify belief that induction will be successful in a new instance, the conclusion that induction will be successful in the new instance does not appear as a premise. Such an argument thus does not commit what is called "premise circularity." However, as Salmon (1966) pointed out, the argument does commit rule circularity, since appeal is made to the rule of induction in making the transition from the premises to the conclusion that recommends the rule of induction. Rule circularity has consequences as untoward as does premise circularity in that there are rules that are patently invalid or crazy that can justify themselves in the same way. Consider the following argument:

> If affirming the consequent is valid, then grass is green.
> Grass is green.
> Therefore, affirming the consequent is valid.

The argument itself proceeds by affirming the consequent, and in so doing leads to the conclusion that affirming the consequent is a legitimate procedure. It is known independently, however, that affirming the consequent is an invalid procedure that can lead from true premises to false conclusion.

In an example closer to the present topic, if one believes that induction is reasonable, one will not believe in the legitimacy of counterinduction, in which one takes past negative instances to be a good indication of future positive instances. However, counterinduction can recommend itself in the same way that affirming the consequent did:

> Counterinduction has usually been unsuccessful in the past.
> Therefore, counterinduction is likely to be successful in the next instance.

The premise is presumably true, even according to an inductivist. The conclusion recommends the rule of counterinduction. The conclusion is inferred from the premise by that same rule of counterinduction, so counterinduction is capable of the same kind of rule-circular defense of itself that was lately recommended for induction. Showing that induction can defend itself in a rule-circular though not premise-circular fashion cannot assuage worries if a rule regarded as illegitimate can do the same (Salmon 1966; Skyrms 2000).

Feigl's distinction between two types of justification, validation and vindication, helps to clarify the so-called 'pragmatic' attempt to justify inductive behavior (see Feigl, Herbert). A principle is validated when it is derived from other, more basic, principles that one accepts, as when a theorem is derived from geometric axioms. A principle is vindicated, on the other hand, when it is shown that following its rule serves the purpose for which that rule was designed (Feigl 1950). Incidentally, this distinction shows that even if inductive rules are basic, not derivable from other principles, as Strawson suggested, that alone does not excuse one from the task of justifying them. They may still require vindication.

The pragmatic justification of inductive behavior was developed by Reichenbach (1949, 469–482) (see Reichenbach, Hans). In this strategy one does not try to show that induction is likely to succeed, but only that it is likely to succeed *if any method will.* In Reichenbach's precise treatment, the rule of induction is to infer from the fact that the frequency of *A*s among *B*s in a large sample is *m*/*n* to the claim that the limit of the relative frequency of *A*s among *B*s in an infinite number of trials is *m*/*n*. This limit either exists or it does not. It must exist if the inductive procedure is to be successful, but according to Reichenbach it need not be known whether it exists when it is asked whether the inductive procedure is justified. The inductive procedure is justified as an attempt to find the limit. There is no known sufficient condition for finding the limit of the relative frequency. However, Reichenbach argued that the attainability of the limit by means of the rule of induction is a necessary condition for the existence of the desired limit. If that limit exists, then it follows analytically that the rule of induction will yield the limit at some point, for a set degree of approximation. If the limit does not exist, then there is no probability for any method to ascertain. More vividly, if the clairvoyant's predictions will identify the limit of the relative frequency, then that limit exists. If the limit exists, the rule of induction will also identify it. Hence, one cannot do better at identifying the limit than by employing the rule of induction.

The main defect in this strategy is that it does not so far defend the standard rule of induction as uniquely qualified for the specified role. The familiar "straight rule" of induction, described above, is one of an infinite number of inductive rules that satisfy the demand of giving the limit of the relative frequency if that limit exists, a class that

Reichenbach called the "asymptotic" rules. The existence of an infinite number of such rules would not be a problem if the limits they yielded in a given case were similar, but in fact they vary arbitrarily widely from each other. For a finite sample, the class of asymptotic rules tolerates any identification of the limit of the relative frequency (Salmon 1966). However, adding to asymptoticity further natural criteria, such as speed-optimality, does narrow the class of acceptable rules (Juhl 1994).

Popper accepted Hume's skeptical conclusion but thought he solved the problem Hume created by declaring that people never use induction anyway (Popper 1972). He developed a deductivist view of science according to which all scientific inference takes the form of falsification; and positive, ampliative inferences asserting the truth or probability of theories, or claims about their future performance, are never made (see Popper, Karl Raimund). When scientists appear to take confirming instances as offering positive evidence for their theories, what they are really doing is finding *corroboration*, a term Popper used for the summation of the theory's past performance under test (see Corroboration). A theory is more highly corroborated the more, and more severe, the tests it has passed that it might have failed, but corroboration never gives grounds for positive conclusions about the theory. Corroboration is thus similar to but crucially distinct from confirmation. One problem with this line of argument is the difficulty even Popper sometimes had eschewing the disallowed positive claims for theories while maintaining a plausible view of scientists' behavior and motivation. Another problem is that, despite its association with deduction, falsification can be seen to be no more decisive than verification, since falsification must assume background claims themselves in need of defense (see Duhem Thesis). Finally, defining corroboration in such a way that theories can be compared with each other when neither is a logical consequence of the other has proved troublesome. One problem is that Popper never adequately defined the notion of severity for tests, a concept on which much depended, since the more severe the test a theory passed, the better its corroboration. Mayo (1996) has recently shed light on this topic by developing the insights inherent in error statistics.

Confirmation Theory

At some point in the early to mid-twentieth century, it was recognized that however the problem of justifying induction turned out, there was much descriptive work to do in order to characterize what exactly the rules of inference in question were. The most basic rules of deduction have been known since the time of Aristotle, but the same degree of clarity had not been achieved for the other major branch of reasoning, *viz.,* induction. (Arguably, it still has not, perhaps testifying to the difficulty of the task.) Perhaps the reason there were few complaints about the fact that the rules of deduction cannot be given a noncircular justification was that those rules were so clear. Work on the description of induction, known as "confirmation theory," might shed light on, if not entirely transform, the justificatory question (see Confirmation Theory). Inductive logic is a similarly descriptive enterprise but is based on the calculus of probability (see Inductive Logic).

However, confirmation theory encountered several paradoxes on its way, including the Raven paradox discovered by Hempel (Hempel 1937, 1943, and 1945; Earman and Salmon 1992) (see Hempel, Carl Gustav). It seems reasonable to assume that any instance lends some confirmation to its generalization, and also that, if a given piece of evidence lends confirmation to a hypothesis, it also lends confirmation to every statement logically equivalent to that hypothesis. However, the statement that all non-black things are non-ravens is logically equivalent to the statement that all ravens are black. Yet a white tennis shoe, which is an instance of the former statement, supports the hypothesis that all ravens are black only on pain of elevating indoor ornithology to the status of a science. More generally, three conditions are sufficient for this paradox. The first is the instantiation condition (also known as "Nicod's condition," after Jean Nicod), by which any instance that is both P and Q provides confirmation for the hypothesis $(x)(Px \rightarrow Qx)$, and any instance that is P and not Q provides disconfirmation of that hypothesis. Another is the equivalence condition, by which an instance that provides confirmation for a hypothesis, H, provides confirmation for any statement logically equivalent to H. The third, the irrelevance condition, says that for any hypothesis, H, there are some instances that provide neither confirming nor disconfirming evidence for H. Orthodox statistics, and some Bayesian approaches, resolve the paradox by denying the instantiation condition (Giere 1970). The most well known Bayesian approach involves denying the irrelevance condition but arguing that the confirmation that instances like the white tennis shoe provide is of negligible size (Howson and Urbach 1996, 126–130) (see Bayesianism).

The paradox of confirmation discovered by Goodman, which introduced what is called the "new riddle of induction," may present a more serious problem, since it seems to show that a purely syntactical confirmation theory is impossible (Goodman 1983; Stalker 1994). Let the predicate "grue" be true of an emerald just in case it is green and observed before January 2100 or not observed before that new year and blue. Why should this not be the predicate one projects on the basis of all of the green emeralds one has seen? Syntactically speaking, the generalization that all emeralds are grue is instantiated by, and so confirmed if anything is by, all the same instances that confirm the generalization that all emeralds are green, since those are also instances in which the emeralds are grue. Indeed, the generalization that all emeralds are grue is apparently confirmed in just the same way, to just the same degree, as the generalization that all emeralds are green. Syntactically, Goodman argues, the two predicates are symmetrical. Common sense says it surely cannot be that the two are equally confirmed by emeralds that have been observed, yet what reason can be given to project the predicate "green" rather than the predicate grue? Nothing in syntax alone, Goodman submitted, can show why the predicate grue should be shunned in favor of the predicate green.

Goodman's answer to this problem was in the tradition of Hume when he developed the notion that projectible predicates can be distinguished from nonprojectible predicates on the basis of the former's superior entrenchment, that is, on the fact that the former predicates have actually been projected more in the past. "Regularities are where you find them, and you can find them anywhere," wrote Goodman, who focused attention on the fact that Hume did not emphasize—which was touched on above with uniformity-of-nature assumptions—that there are many repeated instances that are *not* expected to continue. Goodman's view that what distinguishes projectible from nonprojectible predicates is the former's entrenchment is a descriptive account, as he intended it to be. Goodman was convinced that he had dissolved the traditional problem about the justification of induction in favor of descriptive questions and answers. However, this may not be persuasive, since one can ask about entrenched predicates by what right one's ancestors chose them over other possibilities, and of oneself what reason one has for continuing that tradition.

Sherrilyn Roush

References

Ayer, Alfred Jules (1952), *Language, Truth and Logic*. New York: Dover, 49–50.

Black, Max (1954), "The Inductive Support of Inductive Rules," in *Problems of Analysis*. Ithaca, NY: Cornell University Press, 191–208.

Earman, John, and Wesley C. Salmon (1992), "The Confirmation of Scientific Theories," in Merrilee Salmon et al. (eds.), *Introduction to the Philosophy of Science*. Indianapolis: Hackett Publishing Company.

Feigl, Herbert (1950), "De principiis non disputandum," in Max Black (ed.), *Philosophical Analysis*. Ithaca, NY: Cornell University Press.

Giere, Ronald N. (1970), "An Orthodox Statistical Resolution to the Paradox of Confirmation," *Philosophy of Science* 37: 354–362.

Goodman, Nelson (1983), *Fact, Fiction, and Forecast* (4th ed.). Cambridge, MA: Harvard University Press, 59–124.

Hempel, Carl G. (1937), "Le Problème de la Vérité," *Theoria* (Goteberg), vol. 3.

——— (1943), "A Purely Syntactical Definition of Confirmation," *Journal of Symbolic Logic*, vol. 8.

——— (1945), "Studies in the Logic of Confirmation," *Mind* 54: 1–26, 97–121.

——— (1981), "Turns in the Evolution of the Problem of Induction," *Synthese* 46: 389–404.

Howson, Colin (2000), *Hume's Problem: Induction and the Justification of Belief*. Oxford: Oxford University Press.

Howson, Colin, and Peter Urbach (1996), *Scientific Reasoning: The Bayesian Approach* (2nd ed.). Chicago: Open Court.

Hume, David (1965), *Treatise of Human Nature*. Edited by L. A. Selby-Bigge. Oxford: Oxford University Press.

——— (1999), *An Enquiry Concerning Human Understanding*. Edited by T. L. Beauchamp. Oxford: Oxford University Press, 108–130.

Juhl, Cory (1994), "The Speed-Optimality of Reichenbach's Straight Rule of Induction," *British Journal for the Philosophy of Science* 45: 857–863.

Mayo, Deborah G. (1996), *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Popper, Karl R. (1972), "Conjectural Knowledge: My Solution of the Problem of Induction," in *Objective Knowledge*. Oxford: Clarendon Press, 1–31.

Reichenbach, Hans (1949), "The Justification of Induction," in *The Theory of Probability*, Section 91. Berkeley and Los Angeles: University of California Press.

Russell, Bertrand (1959), "On Induction," in *The Problems of Philosophy*. New York: Oxford University Press, 60–69.

Salmon, Wesley C. (1966), *The Foundations of Scientific Inference*. Pittsburgh, PA: University of Pittsburgh Press.

Skyrms, Brian (2000), *Choice and Chance: An Introduction to Inductive Logic*. Stamford, CT: Wadsworth, 2000.

Stalker, Douglas (ed.) (1994), *Grue! The new riddle of induction*. Chicago: Open Court.

Strawson, P. F. (1952), "Inductive Reasoning and Support," in *Introduction to Logical Theory*. London: Methuen, 233–263.

See also **Causation; Confirmation Theory; Inductive Logic; Laws of Nature**

INDUCTIVE LOGIC

The idea of inductive logic as providing a general, quantitative way of evaluating arguments is a relatively modern one. Aristotle's conception of 'induction' (επαγωγή)—which he contrasted with 'reasoning' (συλλογισμός)—involved moving only from particulars to universals (Kneale and Kneale 1962, 36). This rather narrow way of thinking about inductive reasoning seems to have held sway through the Middle Ages and into the seventeenth century, when Francis Bacon (1620) developed an elaborate account of such reasoning. During the eighteenth and nineteenth centuries, the scope of thinking about induction began to broaden considerably with the description of more sophisticated inductive techniques (e.g., those of Mill [1843]), and with precise mathematical accounts of the notion of probability. Intuitive and quasi-mathematical notions of probability had long been used to codify various aspects of uncertain reasoning in the contexts of games of chance and statistical inference (see Stigler 1986 and Dale 1999), but a more abstract and formal approach to probability theory would be necessary to formulate the general modern inductive-logical theories of nondemonstrative inference. In particular, the pioneering work in probability theory by Bayes (1764), Laplace (1812), Boole (1854), and many others in the eighteenth and nineteenth centuries laid the groundwork for a much more general framework for inductive reasoning. (Philosophical thinking about the possibility of inductive knowledge was most famously articulated by David Hume 1739–1740 and 1758) (See Problem of Induction).

The contemporary idea of inductive logic (as a general, logical theory of argument evaluation) did not begin to appear in a mature form until the late nineteenth and early twentieth centuries. Some of the most eloquent articulations of the basic ideas behind inductive logic in this modern sense appear in John Maynard Keynes's *Treatise on Probability.* Keynes (1921, 8) describes a "logical relation between two sets of propositions in cases where it is not possible to argue demonstratively from one to another." Nearly thirty years later, Rudolf Carnap (1950) published his encyclopedic work *Logical Foundations of Probability*, in which he very clearly explicates the idea of an inductive-logical relation called "confirmation," which is a quantitative generalization of deductive entailment (See Carnap, Rudolf; Confirmation Theory).

Carnap (1950) gives some insight into the modern project of inductive logic and its relation to classical deductive logic:

> Deductive logic may be regarded as the theory of the relation of logical consequence, and inductive logic as the theory of another concept ["c"] which is likewise objective and logical, *viz.,* . . . degree of confirmation. (43)

More precisely, the following three fundamental tenets have been accepted by the vast majority of proponents as desiderata of modern inductive logic:

1. Inductive logic should provide a quantitative generalization of (classical) deductive logic. That is, the relations of deductive entailment and deductive refutation should be captured as limiting (extreme) cases with cases of "partial entailment" and "partial refutation" lying somewhere on a continuum (or range) between these extremes.
2. Inductive logic should use probability (in its modern sense) as its central conceptual building block.
3. Inductive logic (i.e., the nondeductive relations between propositions that are characterized by inductive logic) should be objective and logical.

(Skyrms 2000, chap. 2, provides a contemporary overview.) In other words, the aim of inductive logic is to characterize a quantitative relation (of inductive strength or confirmation), c, which satisfies desiderata 1–3 above. The first two of these desiderata are relatively clear (or will quickly become clear below). The third is less clear. What does it mean for the quantitative relation c to be objective and logical? Carnap (1950) explains his understanding as follows:

> That c is an objective concept means this: if a certain c value holds for a certain hypothesis with respect to a certain evidence, then this value is entirely independent of what any person may happen to think about these sentences, just as the relation of logical consequence is independent in this respect. [43] . . . The principal common characteristic of the statements in both fields

> [deductive and inductive logic] is their independence of the contingency of facts [of nature]. This characteristic justifies the application of the common term 'logic' to both fields. [200]

This entry will examine a few of the prevailing modern theories of inductive logic and discuss how they fare with respect to these three central desiderata. The meaning and significance of these desiderata will be clarified and the received view about inductive logic critically evaluated.

Some Basic Terminology and Machinery for Inductive Logic

It is often said (e.g., in many contemporary introductory logic texts) that there are two kinds of argument: deductive and inductive, where the premises of deductive arguments are intended to guarantee the truth of their conclusions, while inductive arguments involve some risk of their conclusions being false even if all of their premises are true (see, e.g., Hurley 2003). It seems better to say that there is just one kind of argument: An argument is a set of propositions, one of which is the conclusion, the rest are premises. There are many ways of evaluating arguments. Deductive logic offers strict, qualitative standards of evaluation: the conclusion either follows from the premises or it does not, whereas inductive logic provides a finer-grained (and thereby more liberal) quantitative range of evaluation standards for arguments. One can also define comparative and/or qualitative notions of inductive support or confirmation. Carnap (1950, §8) and Hempel (1945) both provide penetrating discussions of the contrast between quantitative and comparative/qualitative notions. For simplicity, the focus here will be on quantitative approaches to inductive logic, but most of the main issues and arguments discussed below can be recast in comparative or qualitative terms.

Let $\{P_1, \ldots, P_n\}$ be a finite set of propositions constituting the premises of an (arbitrary) argument, and let C be its conclusion. Deductive logic aims to explicate the concept of *validity* (i.e., deductive-logical goodness) of arguments. Inductive logic aims to explicate a quantitative generalization of this deductive concept. This generalization is often called the "inductive strength" of an argument (Carnap 1950 uses the word "confirmation" here). Following Carnap, the notation $c(C, \{P_1, \ldots, P_n\})$ will denote the degree to which $\{P_1, \ldots, P_n\}$ jointly inductively support (or "confirm") C.

As desideratum 2 indicates, the concept of probability is central to the modern project of inductive logic. The notation $P(\bullet)$ and $P(\bullet|\bullet)$ will denote unconditional and conditional probability functions, respectively. Informally (and roughly), "$P(p)$" can be read "the probability that proposition p is true," and "$P(p|q)$" can be read "the probability that proposition p is true, given that proposition q is true." The nature of probability functions and their relation to the project of inductive logic will be a central theme in what follows.

A Naive Version of Basic Inductive Logic and the Received View

According to classical deductive propositional logic, the argument from $\{P_1, \ldots, P_n\}$ to C is *valid* iff ("if and only if") the material conditional $(P_1 \wedge, \ldots, \wedge P_n) \rightarrow C$ is (logically) necessarily true. Naively, one might try to define "inductively strong" as follows: The argument from $\{P_1, \ldots, P_n\}$ to C is *inductively strong* iff the material conditional $(P_1 \wedge \ldots \wedge P_n) \rightarrow C$ is (logically?) probably true. More formally, one can express this naive inductive logic (NIL) proposal as follows:

$$c(C, \{P_1, \ldots, P_n\}) \text{ is high iff } P((P_1 \wedge, \ldots, \wedge P_n) \rightarrow C) \text{ is high.}$$

There are problems with this first, naive attempt to use probability to generalize deductive validity quantitatively. As Skyrms (2000, 19–22) points out, there are (intuitively) cases in which the material conditional $(P_1 \wedge, \ldots, \wedge P_n) \rightarrow C$ is probable but the argument from $\{P_1, \ldots, P_n\}$ to C is not a strong one. Skyrms (21) gives the following example:

> (**P**) There is a man in Cleveland who is 1,999 years and 11 months old and in good health.
> (**C**) No man will live to be 2,000 years old.

Skyrms argues that $P(\mathbf{P} \rightarrow \mathbf{C})$ is high, simply because $P(\mathbf{C})$ is high and not because there is any evidential relation between **P** and **C**. Indeed, intuitively, the argument from (**P**) to (**C**) is not strong, since (**P**) seems to disconfirm or countersupport (**C**). Thus, $P((P_1 \wedge \ldots \wedge P_n) \rightarrow C)$ being high is not sufficient for $c(C, \{P_1, \ldots, P_n\})$ being high. Note also that $P((P_1 \wedge \ldots \wedge P_n) \rightarrow C)$ cannot serve as $\mathrm{c}(C, \{P_1, \ldots, P_n\})$, since it violates desideratum 1. If $\{P_1, \ldots, P_n\}$ refutes C, then $Pr((P_1 \wedge \ldots \wedge P_n) \rightarrow C) = Pr(\neg(P_1 \wedge \ldots \wedge P_n))$, which is not minimal, since the conjunction of the premises of an argument need not have probability one.

Skyrms suggests that the mistake that NIL makes is one of conflating the probability of the material conditional $Pr((P_1 \wedge \ldots \wedge P_n) \rightarrow C)$ with the conditional probability of C, given $P_1 \wedge \ldots \wedge P_n$, that is, $P(C|P_1 \wedge \ldots \wedge P_n)$.. According to Skyrms, it is the latter that should be used as a definition of $\mathsf{c}(C, \{P_1, \ldots, P_n\})$. The reason for this preference is that $P((P_1 \wedge \ldots \wedge P_n) \rightarrow C)$ fails to capture the evidential relation between the premises and conclusion, since $P((P_1 \wedge \ldots \wedge P_n) \rightarrow C)$ can be high solely in virtue of the unconditional probability of (C) being high or solely in virtue of the unconditional probability of $P_1 \wedge \ldots \wedge P_n$ being low. As Skyrms (20) stresses, $\mathsf{c}(C, \{P_1, \ldots, P_n\})$ should measure the "evidential relation between the premises and the conclusion." This leads Skyrms (and many others) to defend the following account, which might be called the received view (RV) about inductive logic:

$$\mathsf{c}(C, \{P_1, \ldots, P_n\}) = Pr(C|P_1 \wedge \ldots \wedge P_n).$$

The idea that $\mathsf{c}(C, \{P_1, \ldots, P_n\})$ should be identified with the conditional probability of C, given $P_1 \wedge \ldots \wedge P_n$, has been nearly universally accepted by inductive logicians since the inception of the contemporary discipline. Recent pedagogical advocates of the RV include Copi and Cohen (2001), Hurley (2003), and Layman (2002); and historical champions of various versions of the RV include Keynes (1921), Carnap (1950), Kyburg (1970), and Skyrms (2000), among many others. There are nevertheless some compelling reasons to doubt the correctness of the RV. These reasons, which are analogous to Skyrms's reasons for rejecting the NIL, will be discussed below. But before one can adequately assess the merits of the NIL, RV, and other proposals concerning inductive logic, one needs to say more about probability models and their relation to inductive logic (see Probability).

Probability: Its Interpretation and Role in Traditional Inductive Logic

The Mathematical Theory of Probability

For present purposes, assume that a probability function $P(\bullet)$ is a finitely additive measure function over a Boolean algebra of propositions (or sentences in some formal language). That is, assume that $P(\bullet)$ is a function from a Boolean algebra B of propositions (or sentences) to the unit interval [0,1] satisfying the following three axioms (this is Kolmogorov's (1950) axiomatization), for all propositions X and Y in B:

i. $P(X) \geq 0$.
ii. If X is a (logically) necessary truth, then $P(X) = 1$.
iii. If X and Y are mutually exclusive, then $P(X \vee Y) = Pr(X) + Pr(Y)$.

Following Kolmogorov, define conditional probability $P(\bullet|\bullet)$ in terms of unconditional probability $P(\bullet)$, as follows:

$$Pr(X|Y) = Pr(X \wedge Y)/Pr(Y), \text{ provided that } Pr(Y) \neq 0.$$

A probability model $M = \langle B, P_M \rangle$ consists of a Boolean algebra B of propositions (or sentences in some language), together with a particular probability function $P_M(\bullet)$ over the elements of B.

These axioms (and the definition of conditional probability) say what the mathematical properties of probability models are, but they do not say anything about the interpretation or application of such models. The latter issue is philosophically more central and controversial than the former (but see Popper 1992, appendix *iv, Roeper and Leblanc 1999, and Hájek 2003 for dissenting views on the formal theory of conditional probability). There are various ways in which one can interpret or understand probabilities (see Probability for a thorough discussion). The two interpretations that are most commonly encountered in the context of applications to inductive logic are the so-called "epistemic" and "logical" interpretations of probability.

Epistemic Interpretations of Probability

In epistemic interpretations of probability, $P_M(H)$ is (roughly) the degree of belief that an epistemically rational agent assigns to H, according to a probability model M of the agent's epistemic state. A rational agent's background knowledge K is assumed (in orthodox theories of epistemic probability) to be "included" in any epistemic probability model M, and therefore K is assumed to have an unconditional probability of 1 in M. $P_M(H|E)$ is the degree of belief an epistemically rational agent assigns to H upon learning that E is true (or on the supposition that E is true; see Joyce 1999, chap. 6, for discussion), according to a probability model M of the agent's epistemic state. According to standard theories of epistemic probability, agents learn by conditionalizing on evidence. So, roughly speaking, the probabilistic structure of a rational agent's epistemic state evolves (in time t) through a series of probability models $\{M_t\}$, where evidence learned at time t has probability 1 in all subsequent models $\{M_{t'}\}, t' > t$.

Keynes (1921) seems to be employing an epistemic interpretation of probability in his inductive logic when he says:

> Let our premises consist of any set of propositions *h*, and our conclusion consist of any set of propositions *a*, then,

> if a knowledge of *h* justifies a rational degree of belief in *a* of degree *x*, we say that there is a *probability-relation* of degree *x* between *a* and *h* [$P(a|h) = x$]. (4)

It is not obvious that the RV can satisfy desideratum 3—that c be logical and objective—if the probability function P that is used to explicate c in the RV is given an epistemic interpretation of this kind. After all, whether "a knowledge of *h* justifies a rational degree of belief in *a* of degree *x*" seems to depend on what one's background knowledge *K* is. And while this is arguably an objective fact, it also seems to be a contingent fact and not something that can be determined a priori (on the basis of *a* and *h* alone). As Keynes (1921) explains, his probability function $P(a|h)$ is not subjective, since "once the facts are given which determine our knowledge [background and *h*], what is probable or improbable [*viz*., *a*] in these circumstances has been fixed objectively, and is independent of our opinion" (4). But he later suggests that the function is contingent on what the agent's background knowledge *K* is, in the sense that $P(a|h)$ can vary "depending upon the knowledge to which it is related."

Carnap (1950, §45B) is keenly aware of this problem. He suggests that Keynes should have characterized $P(a|h)$ as the degree of belief in *a* that is justified by knowledge of *h—and nothing else* (the reader may want to ponder what it might mean for an agent to "know *h and nothing else*"). As Keynes's remarks suggest (and as Maher 1996 explains), the problem is even deeper than this, since even a complete specification of an agent's background knowledge *K* may not be sufficient to pick out a unique (rational) epistemic probability model *M* for an agent. (Keynes's reaction to this was to conclude that sometimes quantitative judgments of inductive strength or degree of conditional probability are not possible and that in these cases one must settle for qualitative or comparative judgments.) The problem here is that "$P(X|K)$" ("the probability of *X*, given background knowledge *K*") will not (in general) be determined unless an epistemic probability model *M* is specified, which (*a fortiori*) gives $Pr_M(X)$, for each *X* in *M*. And, without a determination of these fundamental or a priori probabilities $P_M(X)$, a general (quantitative) theory of inductive logic based on epistemic probabilities seems all but hopeless. This raises the problem of specifying an appropriate a priori probability model *M*. Keynes (1921, chap. 4) and Carnap (see below) both look to the principle of indifference at this point, as a guide to choosing a priori probability models. Before discussing the role of the principle of indifference, logical interpretations of probability require a brief discussion.

Logical Interpretations of Probability

Philosophers who accepted the RV and were concerned about the inductive-logical ramifications (mainly, regarding the satisfaction of desideratum 3) of interpreting probabilities epistemically began to formulate logical interpretations of probability. In such interpretations, conditional probabilities $P(X \mid Y)$ are themselves understood as quantitative generalizations of a logical entailment (or deducibility) relation between propositions *Y* and *X*. The motivation for this should be clear—it seems like the most direct way to guarantee that an RV-type theory of inductive logic will satisfy desideratum 3. If $P(\bullet \mid \bullet)$ is itself logical, then c(•,•), which is defined by the RV as $P(\bullet \mid \bullet)$, should also be logical, and the satisfaction of desideratum 3 (as well as the other two) seems automatic. Below it will become clear that RV + logical probability is not the only way (and not necessarily the best way) to satisfy the three desiderata for providing an adequate account of the logical relation of inductive support. In preparation, the notion of logical probability must be examined in some detail.

Typically, logical interpretations of probability attempt to define $Pr(q \mid p)$, where *p* and *q* are sentences in some formal first-order language *L*, in terms of the syntactical features of *p* and *q* (in *L*). The most famous logical interpretations of probability are those of Carnap. It is interesting to note that Carnap's (1950 and 1952) systems are almost identical to those described 20–30 years earlier by W. E. Johnson (1921 and 1932) (Paris 1994; Kyburg 1970, Ch. 5). His later work (Carnap 1971 and 1980) became increasingly complicated, involving two-dimensional continua, and was less tightly coupled with the syntax of *L* (Maher 2000 and 2001; Skyrms 1996 discusses some recent applications of Carnapian techniques to Bayesian statistical models involving continuous random variables; Glaister 2001 and Festa 1993 provide broad surveys of Carnapian theories of logical probability and inductive logic).

Begin with a standard first-order logical language *L* containing a finite number of monadic predicates *F*, *G*, *H*, ... and a finite or denumerable number of individual constants *a*, *b*, *c*, ... Define an unconditional probability function $P(\bullet)$ over the sentences of *L*. Finally, following the standard Kolmogorovian approach, construct a conditional probability function $P(\bullet \mid \bullet)$ over pairs of sentences of *L*, using the ratio definition of conditional probability given above. To fix ideas, consider a very simple toy language L with only two monadic predicates, F and G and only two individual constants a and b. In this language, there are only sixteen possible states of the world that can be described. These sixteen maximally specific descriptions are called the *state descriptions* of L, and they are as follows:

$Fa \wedge Ga \wedge Fb \wedge Gb$
$Fa \wedge Ga \wedge Fb \wedge \neg Gb$
$Fa \wedge Ga \wedge \neg Fb \wedge Gb$
$Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb$
$Fa \wedge \neg Ga \wedge Fb \wedge Gb$
$Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb$
$Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$
$Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb$
$\neg Fa \wedge Ga \wedge Fb \wedge Gb$
$\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb$
$\neg Fa \wedge Ga \wedge \neg Fb \wedge Gb$
$\neg Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb$
$\neg Fa \wedge \neg Ga \wedge Fb \wedge Gb$
$\neg Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb$
$\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$
$\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb$

Two state descriptions S_1 and S_2 are said to be *permutations* of each other if S_1 can be obtained from S_2 by some permutation of the individual constants. For instance, $Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$ can be obtained from $\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb$ by permuting a and b. Thus, $Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$ and $\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb$ are permutations of each other (in L). A *structure description* in L is a disjunction of state descriptions, each of which is a permutation of the others. In the toy language L, there are the following ten structure descriptions:

$Fa \wedge Ga \wedge Fb \wedge Gb$

$(Fa \wedge Ga \wedge Fb \wedge \neg Gb) \vee (Fa \wedge \neg Ga \wedge Fb \wedge Gb)$

$(Fa \wedge Ga \wedge \neg Fb \wedge Gb) \vee (\neg Fa \wedge Ga \wedge Fb \wedge Gb)$

$(Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb) \vee (\neg Fa \wedge \neg Ga \wedge Fb \wedge Gb)$

$Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb$

$(Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb) \vee (\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb)$

$(Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb) \vee (\neg Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb)$

$\neg Fa \wedge Ga \wedge \neg Fb \wedge Gb$

$(\neg Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb) \vee (\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb)$

$\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb$

Now assign nonnegative real numbers to the state descriptions, so that these sixteen numbers sum to 1. Any such assignment will constitute an unconditional probability function $P(\bullet)$ over the state descriptions of L. To extend $P(\bullet)$ to the entire language L, stipulate that the probability of a disjunction of mutually exclusive sentences is the sum of the probabilities of its disjuncts. Since every sentence in L is equivalent to some disjunction of state descriptions, and every pair of state descriptions is mutually exclusive, this gives a complete unconditional probability function $P(\bullet)$ over L. For instance, since $Fa \wedge Ga \wedge \neg Gb$ is equivalent to the disjunction $(Fa \wedge Ga \wedge Fb \wedge \neg Gb) \vee (Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb)$, one will have:

$$\begin{aligned} &Pr(Fa \wedge Ga \wedge \neg Gb) \\ &\quad = Pr((Fa \wedge Ga \wedge Fb \wedge \neg Gb) \vee (Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb)) \\ &\quad = Pr(Fa \wedge Ga \wedge Fb \wedge \neg Gb) + Pr(Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb). \end{aligned}$$

Now, it is only a brief step to the definition of the conditional probability function $P(\bullet|\bullet)$ over pairs of sentences in L. Using the standard, Kolmogorovian ratio definition of conditional probability, for all pairs of sentences X, Y in L:

$$P(X|Y) = P(X \wedge Y)/Pr(Y), \text{provided that } P(Y) \neq 0.$$

Thus, once the unconditional probability function $P(\bullet)$ is specified for the state descriptions of a language L, all probabilities both conditional and unconditional are thereby determined over L. And, this gives one a logical probability model M over the language L. The unconditional, logical probability functions so defined are typically called measure functions. Carnap (1950) discusses two "natural" measure functions.

The first Carnapian measure function is $m^\dagger$, which assumes that each of the state descriptions is equiprobable a priori: If there are N state descriptions in L, then $m^\dagger$ assigns $\frac{1}{N}$ to each state description. While this may seem like a very natural measure function, since it applies something like the principle of indifference to the state descriptions of L (see below for discussion), $m^\dagger$ has the consequence that the resulting probabilities cannot reflect learning from experience. Consider the following simple example. Assume that one adopts a logical probability function $P(\bullet)$ based on $m^\dagger$ as one's own a priori degree of belief (or credence) function. Then, one learns (by conditionalizing) that an object a is F, that is, Fa. Intuitively, one's conditional degree of credence $P(Fb|Fa)$ that a distinct object b also is F, given that a is F, should not always be the same as one's a priori degree of credence that b is F. That is, the fact that one has observed another F object should (at least in some cases) make it more probable (a posteriori) that b will also be F (i.e., more probable than Fb was a priori). More generally, if one observes that a large number of objects have been F, this should raise the probability that the next object one observes will also be F. Unfortunately, no a priori probability function based on $m^\dagger$ is consistent with learning from experience in either sense. To see this, consider the simple case $Pr(Fb|Fa)$:

$$\begin{aligned} P(Fb|Fa) &= m^\dagger(Fb \wedge Fa)/m^\dagger(Fa) \\ &= \frac{1}{2} = m^\dagger(Fb) = Pr(Fb). \end{aligned}$$

So, if one assumes an a priori probability function based on $m^\dagger$, the fact that one object has property F cannot affect the probability that any other object will also have property F. Indeed, it can be shown (Kyburg 1970, 58–59) that no matter how many objects are assumed to be *F*, this will be irrelevant (according to probability functions based on $m^\dagger$) to the hypothesis that a distinct object will also be *F*.

The fact that (on the probability functions generated by the measure $m^\dagger$) no object's having certain properties can be informative about other objects also having those same properties has been viewed as a serious shortcoming of $m^\dagger$ (Carnap 1955). As a result, Carnap formulated an alternative measure function m^*, which is defined as follows. First, assign equal probabilities to each structure description (in the toy language above, n $\frac{1}{10}$). Then, each state description belonging to a given structure description is assigned an equal portion of the probability assigned to that structure description). For instance, in the toy language, the state description $Fa \wedge Ga \wedge \neg Fb \wedge Gb$ gets assigned an a priori probability of $\frac{1}{20}$ ($\frac{1}{2}$ of $\frac{1}{10}$), but the state description $Fa \wedge Ga \wedge Fb \wedge Gb$ receives an a priori probability of $\frac{1}{10}$ ($\frac{1}{1}$ of $\frac{1}{10}$). To further illustrate the differences between $m^\dagger$ and m^*, here are some numerical values in the toy language *L:*

Measure function $m^\dagger$	Measure function m^*
$m^\dagger(Fa \wedge Ga \wedge \neg Fb \wedge Gb) = \frac{1}{16}$	$m^*(Fa \wedge Ga \wedge Fb \wedge Gb) = \frac{1}{10}$
$m^\dagger((Fa \wedge Ga \wedge \neg Fb \wedge Gb) \vee ((\neg Fa \wedge Ga \wedge Fb \wedge Gb))) = \frac{1}{8}$	$m^*(Fa \wedge Ga \wedge \neg Fb \wedge Gb) = \frac{1}{20}$
$m^\dagger(Fa) = \frac{1}{2}$	$m^*(Fa) = \frac{1}{2}$
$Pr^\dagger(Fa\|Fb) = \frac{1}{2} = m^*(Fa) = Pr^\dagger(Fa)$	$Pr^*(Fa\|Fb) = \frac{3}{5} > \frac{1}{2} = m^*(Fa) = Pr^*(Fa)$

Unlike $m^\dagger$, m^* can model learning from experience, since in the simple language

$$P(Fa \mid Fb) = \frac{3}{5} > \frac{1}{2} = Pr(Fa)$$

if the probability function *P* is defined in terms of the logical measure function m^*. Although m^* does have some advantages over $m^\dagger$, even m^* can give counterintuitive results in more complex languages (Carnap 1952).

Carnap (1952) presents a more complicated framework, which describes a more general class (or "continuum") of conditional probability functions, from which the definitions of $P(\bullet \mid \bullet)$ in terms of m^* and $m^\dagger$ fall out as special cases. This continuum of conditional probability functions depends on a parameter λ, which is supposed to reflect the "speed" with which learning from experience is possible. In this continuum, $\lambda = 0$ corresponds to the "straight rule" of induction, which says that the probability that the next object observed will be *F*, conditional upon a sequence of past observations, is simply the frequency with which *F* objects have been observed in the past sequence; $\lambda = +\infty$ yields a conditional probability function much like that given above by assuming the underlying logical measure $m^\dagger$ (i.e., $\lambda = +\infty$ implies that there is no learning from experience). And setting $\lambda = \kappa$ (where κ is the number of independent families of predicates in Carnap's more elaborate 1952 linguistic framework) yields a conditional probability function equivalent to that generated by the measure function m^*.

But even this λ–continuum has problems. First, none of the Carnapian systems allow universal generalizations to have nonzero probability. This problem was addressed by Hintikka (1966) and Hintikka and Niiniluoto (1980), who provided various alterations of the Carnapian framework that allow for nonzero probabilities of universal generalizations. Moreover, Carnap's early systems did not allow for the probabilistic modeling of analogical effects. That is, in his 1950–1952 systems, the fact that two objects share several properties in common is always irrelevant to whether they share any other properties in common. Carnap's more recent (and most complex) theories of logical probability (1971, 1980) include two additional adjustable parameters (γ and η), designed to provide the theory with enough flexibility to overcome these (and other) limitations. Unfortunately, no Carnapian logical theory of probability to date has successfully dealt with the problem of analogical effects (Maher 2000 and 2001). Moreover, as Putnam (1963) explains, there are further (and some say deeper) problems with Carnapian (or, more generally, syntactical) approaches to logical probability, if they are to be applied to inductive inference generally. The consensus now seems to be that the Carnapian project of characterizing an adequate logical theory of probability is (by his own standards and lights) not very promising (Putnam 1963; Festa 1993; Maher 2001).

This discussion has glossed over technical details in the development of (Carnapian) logical interpretations or theories of probability since 1950. To recapitulate, what is important for present purposes is that Carnap (along with the other advocates of logical probability) was an RV theorist about inductive logic. He identified the concept c(•,•) of inductive strength (or inductive support) with the concept of conditional probability $P(\bullet|\bullet)$. And he thought (partly because of the problems he saw with epistemic interpretations) that in order for an RV account to satisfy desideratum 3, it needed to presuppose a logical interpretation (or theory) of probability. This led him, initially, to develop various logical measures (e.g., the a priori logical probability functions $m^{\dagger}$ and m^*), and then to define conditional logical probability $Pr(\bullet|\bullet)$ in terms of these underlying a priori logical measures, using the standard ratio definition. This approach ran into various problems when it came to the application of $P(\bullet|\bullet)$ to inductive logic. These difficulties mainly had to do with the ability of Carnap's $P(\bullet|\bullet)$ to undergird learning from experience and/or certain kinds of analogical reasoning (for other philosophical objections to Carnap's logical probability project, see Putnam 1963). In response to these difficulties, Carnap began to fiddle directly with the definition of $P(\bullet|\bullet)$. In 1952, he moved to a parameterized definition of $P(\bullet|\bullet)$, which contained an "index of inductive caution" (λ) that was supposed to regulate the speed with which learning from experience is reflected by $P(\bullet|\bullet)$. Later, Carnap (1971, 1980) added γ and η to the definition of $P(\bullet|\bullet)$, as noted above, in an attempt to further generalize the theory and allow for sensitivity to certain kinds of analogical effects. Ultimately, no such theory was ever viewed by Carnap (or others) as fully adequate for the purposes of grounding an RV conception of inductive logic.

At this point, it is important to ask, In what sense are Carnap's theories of logical probability (especially his later ones) *logical*? His early theories (based on the measure functions $m^{\dagger}$ and m^*) applied something like the principle of indifference to the state and/or structure descriptions of the formal language L in order to determine the logical probabilities $P(\bullet|\bullet)$. In this sense, these early theories assume that certain sentences of L are equiprobable a priori. Why is such an assumption *logical*? Or, more to the point, how is *logic* supposed to tell one which statements are equiprobable a priori? Carnap (1955) explains that

> the statement of equiprobability to which the principle of indifference leads is, like all other statements of inductive probability, not a factual but a logical statement. If the knowledge of the observer does not favor any of the possible events, then with respect to this knowledge as evidence they *are* equiprobable. The statement assigning equal probabilities in this case does not assert anything about the facts, but merely the logical relations between the given evidence and each of the hypotheses; namely, that these relations are logically alike. These relations are obviously alike if the evidence has a symmetrical structure with respect to their possible events. The statement of equiprobability asserts nothing more than the symmetry. (22)

Carnap seems to be saying that the principle of indifference is to be applied only to possible events that exhibit certain a priori symmetries with respect to some rational agent's background evidence. But this appears no more logical than Keynes's epistemic approach to probability. It seems that the resulting probabilities $P(\bullet|\bullet)$ will not be logical in the sense Carnap desired (at least no more so than Keynes's epistemic probabilities were), unless Carnap can motivate—on logical grounds—the choice of an a priori probability model. To that end, Carnap's application of the principle of indifference is not very useful. Recall that the goal of Carnap's project (of inductive logic) was to explicate the confirmation relation, which is itself supposed to reflect the evidential relation between premises and conclusions (Carnap 1950 uses the locutions "degree of confirmation" and "weight of evidence" synonymously). How is one to understand what it means for evidence not to "favor any of the possible events" in a way that does not require one to already understand how to measure the degree to which the evidence confirms each of the possible events? Here, Carnap's discussion of the principle of indifference presupposes that degree of confirmation is to be identified with degree of conditional probability. In that reading, "not favoring" just means "conferring equal probability on," and Carnap's unpacking of the principle of indifference reduces directly to a mathematical truth (which, for Carnap, is good enough to render the principle *logical*). If one had independent grounds for thinking that conditional probabilities were the right way to measure confirmation (or weight of evidence), then Carnap would have a rather clever (albeit not terribly informative) way to (logically) ground his choice of a priori probability models. Unfortunately, as will be seen below, there are independent reasons to doubt Carnap's presupposition here that degree of confirmation should be identified with degree of conditional probability. Without that assumption, Carnap's principle of indifference is no longer logical (by his own lights), and the problem of the contingency (nonlogicality) of the ultimate inductive-logical probability assignments returns with a vengeance. There are independent and deep problems with any attempt to consistently apply the principle of indifference to contexts in which hypotheses and/or evidence involve continuous magnitudes (van Fraassen 1989).

Carnap's later theories of $P(\bullet|\bullet)$ introduce even further contingencies, in the form of adjustable parameters, the "proper values" of which do not seem to be determinable a priori (Carnap 1952, 1971, 1980). In particular, consider Carnap's (1952) λ-continuum. The parameter λ is supposed to indicate how sensitive $P(\bullet|\bullet)$ is to learning from experience. A higher value of λ indicates slower learning, and a lower λ indicates faster learning. As Carnap (1952) concedes, no one value of λ is best a priori. Presumably, different values of λ are appropriate for different contexts in which confirmational judgments are made (see Festa 1993 for a contextual Carnapian approach to confirmation). It seems that the same must be said for the additional parameters γ and η (Carnap 1971, 1980). The moral here seems to be that it is only relative to a particular assignment of values to λ, γ, and η that probabilistic (and/or confirmational) judgments are objectively and noncontingently determined in Carnap's later systems. This is analogous to the fact that it is only relative to a (probabilistic) characterization of the agent's background knowledge and complete epistemic state (in the form of a specific epistemic probability model M) that Keynes's epistemic probabilities (or Carnap's measure functions m^* and $m^\dagger$) have a chance of being objectively and noncontingently determined.

A pattern is developing. Both Keynes and Carnap give accounts of a priori probability functions $P(\bullet|\bullet)$ that involve certain contingencies and indeterminacies. They each feel pressure (owing to desideratum 3) to eliminate these contingencies when the time comes to use $P(\bullet|\bullet)$ as an explication of c(•,•). The general strategy for rendering these probabilities logical is to choose some privileged, a priori probability model. Here, both Keynes and Carnap appeal to the principle of indifference to constrain the ultimate choice of model. Carnap is sensitive to the fact that the principle of indifference does not seem logical, but his attempts to render it so (and useful for grounding the choice of an a priori probability model) are both unconvincing and uninformative. There is a much easier and more direct way to guarantee the satisfaction of desideratum 3. Why not just define c from the beginning as a three-place relation that depends on premises, conclusion, and a particular probability model?

The next section describes a simple, general recipe (along the lines suggested by the preceding considerations) for formulating probabilistic inductive logics in such a way that they transparently satisfy desiderata 1–3. This section will also address the following question: Is the RV *materially* adequate as an account of inductive strength or inductive support? This will lead to a fourth material desideratum for measures of inductive support, and ultimately to a concrete alternative to the RV.

Rethinking the Received View

How to Ensure the Transparent Satisfaction of Desideratum 3

The existing attempts to use the notion of probability to explicate the concept of inductive support (or inductive strength) c have foundered on the question of their contingency (which threatened violation of desideratum 3). It may be that these contingencies can be eliminated (in general) only by making the notion of inductive support explicitly relational. To follow such a plan, in the case of the RV one should rather say:

The inductive strength of the argument from $\{P_1, \ldots, P_n\}$ to C relative to a probability model $M = <B, P_M>$ is $P_M(C|P_1 \wedge \ldots \wedge P_n)$.

Relativizing judgments of inductive support to particular probability models fully and transparently eliminates the contingency and indeterminacy of these judgments. It is clear that the revision of RV above satisfies all three desiderata, since:

1. $P_M(C \mid P_1 \wedge \ldots \wedge P_n)$ is maximal and constant when $\{P_1, \ldots, P_n\}$ entails C, and $Pr_M (C \mid P_1 \wedge \ldots \wedge P_n)$ is minimal and constant when $\{P_1, \ldots, P_n\}$ refutes C.
2. The relation of inductive support is defined in terms of the notion of probability.
3. Once the conditional probability function $P_M(\bullet|\bullet)$ is specified (as it is, a fortiori, once the probability model M has been), its values are determined objectively and in a way that is contingent on only certain mathematical facts about the probability calculus. This is, the resulting c-values are determined mathematically by the specification of a particular probability model M.

One might respond at this point by asking, Where do the probability models M come from? and how does one choose an "appropriate" probability model in a given inductive logical context? These are good questions. However, it is not clear that they must be answered by the inductive logician *qua* logician. Here it is interesting to note the analogy between the P_M-relativity of inductive logical relations (in the present approach) and the language relativity of deductive logical relations in Carnap's (early) approach to deductive logic. For the early Carnap, deductive logical (or, more generally, analytic) relations obtain only between sentences in a formal language. The deductive logician is not in the business of telling people which languages they should use, since this (presumably pragmatic) question is "external" to deductive logic. However, once a language has been specified, the deductive relations among sentences in that language are determined objectively and noncontingently, and it is up to the deductive logician to explicate these relations. In the approach to inductive logic just described, the same sort of thing can be said for the inductive logician. It is not the business of the inductive logician to tell people which probability models they should use (presumably, that is an epistemic or pragmatic question), but once a probability model is specified, the inductive logical relations in that model (*viz*., c) are determined objectively and noncontingently. In the present approach, the duty of the inductive logician is (simply) to explicate the c-function—not to decide which probability models should be used in which contexts.

One last analogy might be useful here. When the theory of special relativity came along, some people were afraid that it might introduce an element of subjectivity into physics, since the velocities of objects were now determined only relative to a frame of reference. There was no physical ether with respect to which objects received their absolute velocities. However, the velocities and other values were determined objectively and noncontingently once the frame of reference was specified, which is the reason Einstein originally intended to call his theory the theory of invariants. Similarly, it seems that there may be no *logical ether* with respect to which pairs of propositions (or sentences) obtain their a priori relations of inductive support. But once a probability model M is specified (and independently of how that model is interpreted), the values of c-functions defined relative to M are determined objectively and noncontingently (in precisely the sense Carnap had in mind when he used those terms).

A Fourth Material Desideratum: Relevance

Consider the following argument:

(**P**) Fred Fox (who is a male) has been taking
birth control pills for the past year.
(**C**) Fred Fox is not pregnant.

Intuitively (i.e., assuming a probability model M that properly incorporates one's intuitively salient background knowledge about human biology, etc.), $P_M(\mathbf{C}|\mathbf{P})$ is very high. But does one want to say that there is a strong evidential relation between **P** and **C**? According to proponents of the RV, one should say just that. This seems wrong, because intuitively $P_M(\mathbf{C} \mid \mathbf{P}) = P_M(\mathbf{C})$. That is, $P_M(\mathbf{C}|\mathbf{P})$ is high solely because $P_M(\mathbf{C})$ is high, and not because of any evidential relation between **P** and **C**. This is the same kind of criticism that Skyrms (2000) made against the NIL proposal. And it is just as compelling here. The problem here is that **P** is irrelevant to **C**. Plausibly, it seems that if **P** is going to be counted as providing evidence in favor of **C**, then **P** should raise the probability of **C** (Popper 1954 and 1992; Salmon 1975). This leads to the following fourth material desideratum for c:

- $\mathsf{c}(C, \{P_1, \ldots, P_n\})$ should be sensitive to the probabilistic relevance of $P_1 \wedge \ldots \wedge P_n$ to C.

In particular, desideratum 4 implies that if P_1 raises the probability of C_1, but P_2 lowers the probability of C_2, then $\mathrm{c}(C_1, P_1) > \mathrm{c}(C_2, P_2)$. This rules out $P(C|P_1 \wedge \ldots \wedge P_n)$ as a candidate for $\mathrm{c}(C, \{P_1, \ldots, P_n\})$, and it is therefore inconsistent with the RV. Many nonequivalent probabilistic-relevance measures of support (or confirmation) satisfying desideratum 4 have been proposed and defended in the philosophical literature (Fitelson 1999 and 2001).

One can combine desiderata 1–4 into the following single probabilistic inductive logic. This unified desideratum gives constraints on a three-place probabilistic confirmation function $\mathsf{c}(C, \{P_1, \ldots, P_n\}, M)$, which is the degree to which $\{P_1, \ldots, P_n\}$ inductively supports C, relative to a specified probability model $M = \langle B, Pr_M \rangle$:

$$\mathsf{c}(C,\{P_1,\ldots,P_n\},M) \text{ is } \begin{cases} \text{maximal and } > 0 & \text{if } \{P_1,\ldots,P_n\} \text{ entails } C \\ > 0 & \text{if } P_M(C|P_1 \wedge \ldots \wedge P_n) > P_M(C) \\ 0 & \text{if } P_M(C|P_1 \wedge \ldots \wedge P_n) = P_M(C) \\ < 0 & \text{if } P_M(C|P_1 \wedge \ldots \wedge P_n) < P_M(C) \\ \text{minimal and } < 0 & \text{if } \{P_1,\ldots,P_n\} \text{ entails } \neg C \end{cases}$$

To see that any measure satisfying probabilistic inductive logic will satisfy desiderata 1–4, note that

- the cases of entailment and refutation are at the extremes of c, with intermediate values of support and countersupport in between the extremes;
- the constraints in probabilistic inductive logic can be stated purely probabilistically, and c's values must be determined relative to a probability model M, so any measure satisfying it must use probability as a central concept in its definition;

- the measure c is defined relative to a probability model, and so its values are determined objectively and noncontingently by the values in the specified model; and
- sensitivity to *P*-relevance is built into the desideratum (probabilistic inductive logic).

Interestingly, almost all relevance measures proposed in the confirmation theory literature fail to satisfy probabilistic inductive logic (Fitelson 2001, §3.2.3). One historical measure that does satisfy probabilistic inductive logic was independently defended by Kemeny and Oppenheim (1952) as the correct measure of confirmation (in opposition to Carnap's RV c-measures) within a Carnapian framework for logical probability:

$$\mathsf{c}(C, \{P_1, \ldots, P_n\}, M) = \frac{P_M(P_1 \wedge \ldots \wedge P_n | C) - P_M(P_1 \wedge \ldots \wedge P_n | \neg C)}{P_M(P_1 \wedge \ldots \wedge P_n | C) + P_M(P_1 \wedge \ldots \wedge P_n | \neg C)}.$$

Indeed, of all the historically proposed (probabilistic) measures of degree of confirmation (and there have been dozens), the above measure is the only one (up to ordinal equivalence) that satisfies all four of the material desiderata. The four simple desiderata are thus sufficient to (nearly uniquely) determine the desired explicandum c, or the degree of inductive strength of an argument. There are other measures in the literature, such as the log-likelihood ratio, that differ conventionally from, but are ordinally equivalent to, the above measure (for various other virtues of measures in this family, see Fitelson 2001, Good 1985, Heckerman 1988, Kemeny and Oppenheim 1952, and Schum 1994).

Historical Epilogue on the Relevance of Relevance

In the second edition of *Logical Foundations of Probability*, Carnap (1962) acknowledges that probabilistic relevance is an intuitively compelling desideratum for measures of inductive support. This acknowledgement was in response to the trenchant criticisms of Popper (1954), who was one of the first to urge relevance as a desideratum in this context (see Michalos 1971 for a thorough discussion of this important debate between Popper and Carnap). But instead of embracing relevance measures like Kemeny and Oppenheim's (1952) (and rewriting much of the first edition of *Logical Foundations of Probability*), Carnap (1962) simply postulates an ambiguity in the term "confirmation." He now argues that there are two kinds of confirmation: confirmation as firmness and confirmation as increase in firmness, where the former is properly explicated using just conditional probability (à la the RV) and does not require relevance of the premises to the conclusion, while the latter presupposes that the premises are probabilistically relevant to the conclusion. Strangely, Carnap does not even mention Kemeny and Oppenheim's measure (of which he was aware) as a proper measure of confirmation as increase in firmness. Instead, he suggests for that purpose a relevance measure that does not satisfy desideratum 1 and so is not even a proper generalization of deductive entailment. This puzzling but crucial sequence of events in the history of inductive logic may explain why relevance-based approaches (like that of Kemeny and Oppenheim) have never enjoyed as many proponents as the RV.

BRANDEN FITELSON

References

Bacon, F. (1620), *The Novum Organon*. Oxford: The University Press.

Bayes, T. (1764), "An Essay Towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society of London* 53.

Boole, G. (1854), *An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities*. London: Walton & Maberly.

Carnap, R. (1950), *Logical Foundations of Probability*. Chicago: University of Chicago Press.

——— (1952), *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

——— (1955), *Statistical and Inductive Probability* and *Inductive Logic and Science* (leaflet). Brooklyn, NY: Galois Institute of Mathematics and Art.

——— (1962), *Logical Foundations of Probability*, 2nd ed. Chicago: University of Chicago Press.

——— (1971), "A Basic System of Inductive Logic, I," in R. Carnap and R. Jeffrey (eds.), *Studies in Inductive Logic and Probability*, vol. 1. Berkeley and Los Angeles: University of California Press, 33–165.

——— (1980), "A Basic System of Inductive Logic, II," in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, vol. 2. Berkeley and Los Angeles: University of California Press, 7–155.

Copi, I., and C. Cohen (2001), *Introduction To Logic*, 11th ed. New York: Prentice Hall.

Dale, A. (1999), *A History of Inverse Probability: From Thomas Bayes To Karl Pearson*, 2nd ed. New York: Springer-Verlag.

Festa, R. (1993), *Optimum Inductive Methods*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Fitelson, B. (1999), "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity," *Philosophy of Science* 66: S362–S378.

——— (2001), *Studies in Bayesian Confirmation Theory*. PhD. dissertation, University of Wisconsin–Madison (Philosophy).

Glaister, S. (2001), "Inductive Logic," in D. Jacquette (ed.), *A Companion to Philosophical Logic*. London: Blackwell.

Good, I. J. (1985), “Weight of Evidence: A Brief Survey,” in J. Bernardo, M. DeGroot, D. Lindley, and A. Smith (eds.), *Bayesian Statistics, 2*. Amsterdam: North-Holland, 249–269.

Hájek, A. (2003), “What Conditional Probabilities Could Not Be,” *Synthese* 137: 273–323.

Heckerman, D. (1988), “An Axiomatic Framework for Belief Updates,” in L. Kanal and J. Lemmer (eds.), *Uncertainty in Artificial Intelligence 2*. New York: Elsevier Science Publishers, 11–22.

Hempel, C. (1945), “Studies in the Logic of Confirmation,” parts I and II, *Mind* 54: 1–26 and 97–121.

Hintikka, J. (1966), “A Two-Dimensional Continuum of Inductive Methods,” in J. Hintikka and P. Suppes (eds.), *Aspects of Inductive Logic*. Amsterdam: North-Holland.

Hintikka, J., and I. Niiniluoto (1980), “An Axiomatic Foundation for the Logic of Inductive Generalization,” in R. Jeffrey, *Studies in Inductive Logic and Probability*, vol. 2. Berkeley and Los Angeles: University of California Press.

Hume, D. (1739– 1740), *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*, vols. 1–3. London: John Noon (1739), Thomas Longman (1740).

——— (1758), *An Enquiry Concerning Human Understanding in Essays and Treatises on Several Subjects*. London: A. Millar.

Hurley, P. (2003), *A Concise Introduction to Logic*, 8th ed. Melbourne, Australia, and Belmont, CA: Wadsworth/Thomson Learning.

Johnson, W. E. (1921), *Logic*. Cambridge: Cambridge University Press.

——— (1932), “Probability: The Deductive and Inductive Problems,” *Mind* 49: 409– 423.

Joyce, J. (1999), *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Kemeny, J., and P. Oppenheim (1952), “Degrees of Factual Support,” *Philosophy of Science* 19: 307–324.

Keynes, J. (1921), *A Treatise on Probability*. London: Macmillan.

Kneale, W., and M. Kneale (1962), *The Development of Logic*. Oxford: Clarendon Press.

Kolmogorov, A. (1950), *Foundations of the Theory of Probability*. New York: Chelsea.

Kyburg, H. E. (1970), *Probability and Inductive Logic*. London: Macmillan.

Laplace, P. S. M. d. (1812), *Théorie Analytique des Probabilités.*. Paris: Ve. Courcier.

Layman, C. S. (2002), *The Power of Logic*, 2nd ed. New York: McGraw-Hill.

Maher, P. (1996), “Subjective and Objective Confirmation,” *Philosophy of Science* 63: 149–174.

——— (2000), “Probabilities for Two Properties,” *Erkenntnis* 52: 63–91.

——— (2001), “Probabilities for Multiple Properties: The Models of Hesse and Carnap and Kemeny,” *Erkenntnis* 55: 183–216.

Michalos, A. (1971), *The Popper–Carnap Controversy*. The Hague: Martinus Nijhoff.

Mill, J. (1843), *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London: Parker.

Paris, J. (1994), *The Uncertain Reasoner’s Companion: A Mathematical Perspective*. Cambridge: Cambridge University Press, chap. 12.

Popper, K. (1954), “Degree of Confirmation,” *British Journal for the Philosophy of Science* 5: 143–149.

——— (1992), *The Logic of Scientific Discovery*. London: Routledge.

Putnam, H. (1963), “‘Degree of Confirmation’ and Inductive Logic,” in P. A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*. La Salle, IL: Open Court Publishing, 761–784.

Roeper, P., and H. Leblanc (1999), *Probability Theory and Probability Logic*. Toronto: University of Toronto Press.

Salmon, W. C. (1975), “Confirmation and Relevance,” in G. Maxwell and R. M. Anderson Jr. (eds.), *Induction, Probability, and Confirmation: Minnesota Studies in the Philosophy of Science*, vol. 6. Minneapolis: University of Minnesota Press, 3–36.

Schum, D. (1994), *The Evidential Foundations of Probabilistic Reasoning*. New York: John Wiley & Sons.

Skyrms, B. (1996), “Carnapian Inductive Logic and Bayesian Statistics,” in *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell. IMS Lecture Notes, Monograph Series* 30: 321–336.

——— (2000), *Choice and Chance*. Melbourne, Australia, and Belmont, CA: Wadsworth/Thomson Learning.

Stigler, S. (1986), *The History of Statistics*. Cambridge, MA: Harvard University Press.

van Fraassen, B. (1989), *Laws and Symmetry*. Oxford: Oxford University Press, chap. 12.

INNATE/ACQUIRED DISTINCTION

Arguments about innateness center on two distinct but overlapping theoretical issues. One concerns the explanation of the origin of ideas in the human mind. This ancient question was famously introduced in Plato’s *Meno* and it took center stage in seventeenth- and eighteenth-century debates between rationalists and empiricists. More recently, it has seen a sophisticated revival in arguments about the status of nativism in philosophy of mind and cognitive science (see Chomsky, Noam; Cognitive Science). Nativists believe that a mind cannot learn unless it comes already furnished with certain fundamental ideas and/or representational capacities. The issue remains controversial. Clearly human cognitive architecture has some effect on ways of thinking and the sort of ideas that can be had. Some have wanted to say that this amounts to having innate ideas. Others worry that the notion of innate ideas is really just another (and more confusing) way of discussing cognitive architecture.

The second point at issue concerns the explanation of evolutionary and developmental rigidity in living systems. Evolutionarily rigid traits—such as being five-fingered amongst primates—are resistant to selection pressure. Developmentally rigid traits are resistant to environmental perturbation during ontogeny. Darwin argued that some traits are more advantageous than others and that over time the traits that contributed to more successful organisms would become more common in populations. The power of natural selection in the Darwinian sense comes from the fact that it is cumulative and occurs over very large numbers of generations (see Natural Selection). This explains the existence of complex adaptations such as eyes and brains. For cumulative selection to operate, there must be very reliable inheritance of traits from one generation to the next, otherwise fitness-enhancing traits would "leech out" of populations. Similarly, the development of individual organisms must also be reliable. That is, individuals with similar genomes reared in similar environments must be more likely to develop similar traits (see Heritability). If this were not the case, then biological inheritance could not be reliably fitness enhancing. However, it has long been recognized that traits differ in the extent to which they are rigid. This is true in the development of individuals. Different people might speak different languages or be shorter and taller in stature, but they would not have different numbers of chambers in their hearts. This difference in malleability is also evident when one looks at the evolution of traits in long lineages of organisms. Arthropoda, for example, is a clade of millions of extant species. Its members have evolved a remarkable variety of morphologies and behaviours. Yet despite this diversity, there are still traits that are ubiquitous within the clade. These *diagnostic* traits have become entrenched within the lineage, and they seem highly resistant to selection pressure.

Thus, study of evolutionary theory and developmental biology says that some traits seem to be built into individuals and some seem to be built into evolving lineages. Some biologists have wanted to call such traits innate. But this leads inevitably to the question, What exactly does it mean for a trait to be "built into" an individual or a group of individuals? Is it, for example, the same as saying that the characteristic in question is genetic? Is it the same as saying that the characteristic in question is ubiquitous? Such questions have been asked variously by ethologists, developmental biologists, evolutionary theorists, and, more recently, philosophers of biology. Finally, and most importantly, one might also ask, Is there some clear characterization of innateness that will clarify both of the theoretical issues set out above?

Where Do Thoughts Come From?

The notion that human beings are the bearers of innate ideas has been the servant of many theoretical masters. In the *Meno*, Plato seeks to demonstrate that a slave boy who has received no schooling is, in fact, possessed of knowledge of geometry. Plato's intention is to argue that such knowledge is not learned but rather recollected from a time prior to birth in which humans were in direct contact with the forms. It is thus innate, in the sense of being present at birth.

A much later version of this reasoning seeks to achieve very different programmatic aims. Whereas Plato was concerned to ground a metaphysical system, the Cambridge Platonists attempted early in the seventeenth century to use the same machinery to buttress religious faith. Their claim was that all people have innate knowledge of God's existence and his moral laws. Again, the suggestion was that to be a member of humankind is to have, born within one, a number of undeniable truths (Patrides 1970).

The zenith of this style of argument is not to be found in metaphysics or in theology, but in epistemology. Both Descartes (in the *Meditations*) and Leibniz (in the *New Essays*) employ the notion of innate ideas as a means of staving off a more radical form of skepticism than that which concerned the Cambridge Platonists. As Descartes claims, anything one thinks one knows could in fact be a deception caused by an evil demon. Therefore, if one is to know about the world, then one must do so on the basis of epistemic principles that are not themselves based on the way the world is perceived to be. The solution provided by both Descartes and Leibniz was to defend a foundationalist epistemology that rested upon a God-given bedrock of undeniable truths. These truths are not obtained via perception. They might be thought of as innate, though it should be noted that for something to be known a priori does not in itself imply that the knowledge is innate.

In this tradition, which runs from Plato through to the Rationalists, there is a common denominator. All these systems share the claim that innate ideas are the product of some external metaphysical cause (be it the realm of the forms or an undeceiving God). While this tradition is long-standing, modern philosophical and scientific inquiry has taught that such premises need not underpin arguments for innate ideas.

Nativism

In the latter half of the twentieth century, philosophers of mind and cognitive scientists employed innateness in their explanation of a variety of cognitive phenomena (see Cognitive Science). Noam Chomsky (1965) and his heirs argue that much of the capacity to decipher verbal information is innate; David Marr (1982) defends a similar position with respect to the interpretation of visual information; and Jerry Fodor (1975) argues for a broadly rationalist interpretation of concept acquisition (see Philosophy of Linguistics).

While such positions are now usually categorized as "nativist," they are in some respects direct theoretical descendents of seventeenth-century rationalist arguments (Cowie 1999). Leibniz (1981) argues that a limited modal perspective dictates that one cannot learn necessary truths from observation of the actual world. One simply is not presented with sufficient data to learn facts about the way things are in every possible world (79–80). Thus, he inferred, experience alone cannot explain knowledge of necessary truths. Chomsky also avails himself of arguments based upon poverty of the stimulus. He argues that children are not presented with sufficient input from other language users to explain their acquisition of complex natural languages. More precisely, the grammar they learn is underdetermined by the instances of grammar they encounter. Nor do they receive sufficient negative reinforcement for grammatical errors. Thus, as did Leibniz, Chomsky concludes that the phenomenon in question cannot be explained as the product of experience alone. Famously, he argues that humans are possessed of an innate universal grammar (Chomsky 1965, 47–59). Thus, for him, linguistic development consists not in learning what language is, but rather in learning which language one speaks. For Chomsky, the language faculty is a mental "organ" that develops within humans. It is innate in just the same way that various aspects of morphology are innate. As he puts it, "Language acquisition is not really something that the child does; it is something that happens to the child placed in a certain environment" (Chomsky 1990, 634).

One can draw a similar parallel between Fodor's flagship argument and those of his predecessors. For Descartes, it is impossible that the mechanical process of perception (consisting of the corporeal movement of nerves and resulting flows of vital spirits) could actually result in something as perfect as the idea of blueness. Thus, he concludes that all apparently learned ideas actually come from God. This is usually described as an impossibility argument, for gaining knowledge via perception is impossible, and thus concepts must be supplied by some other means. Fodor (1975 and 1981) also avails himself of an impossibility argument. He argues that concept acquisition is strongly dependent upon a preexisting representational capacity. It would be impossible to learn about the world unless one could form hypotheses about it. But to do that, one must have some system of representation (usually called mentalese) in which to form the hypotheses in the first place. Of course, mentalese could also be learned, but one would have to have formed hypotheses, etc. Ultimately, Fodor argues, if one is to avoid an infinite regress, one must admit the existence of some innate representational vocabulary.

In recognizing that nativism is a true heir of rationalism, it should not be presumed that little has changed since the seventeenth century. It would be unfair to suggest that Fodor's impossibility argument is much the same as those of his philosophical forebears (see the discussion in the following section). Furthermore, Chomsky makes use of a wide variety of naturalistic arguments supporting the idea that human beings inherit domain-specific cognitive capacities. So, for example, Chomsky and other nativists in linguistics also argue for an innate universal grammar on the following grounds:

- All natural languages share a variety of suboptimal features that would be explained by faults in the underlying universal grammar.
- Language learning appears to be tied to biological development in a way that other types of learning are not. Acquiring a language during early childhood is much easier than acquiring one outside this crucial developmental window.
- There is a great breadth of evidence (particularly from cases of brain trauma) suggesting that language learning is modular.
- There is clear evidence that general intelligence and linguistic ability are independent.

Against Nativism

While many see modern arguments for nativism as much more compelling than Leibniz' metaphysics or Socrates' interrogation of the uneducated, it is still the case that all the arguments in favor of nativism are controversial to some degree.

Fodor's flagship argument for nativism is one that many philosophers find unpalatable. Many worry that Fodor's view implies that if humans have born within them concepts of those things that they will later learn to name, then at least some humans must be born with innate concepts of special relativity, cellular phones, and Barbie dolls. But this seems distinctly implausible given that humans have evolved in environments in which at least cell phones and Barbie dolls made no appearance. It should be noted though that Fodor's argument does not imply that *all* the terms used in natural languages must correspond to innate concepts. Fodor admits that one might combine primitive words in mentalese to represent complex thoughts that one then attaches to words in natural languages. So one need not have innate concepts corresponding to all the things that one names using natural languages. One might accept Fodor's basic argument but maintain that relatively few concepts are innate. However, Fodor (1981) rejects that conclusion, arguing that empirical evidence says that most concepts are primitive rather than constructed out of primitives (272–275). Some reject Fodor's argument on the grounds that it wrongly assumes that language acquisition involves learning facts of the form "X_{English} means the same as $Y_{\text{mentalese}}$." This is just to say that anyone who rejects the language of thought hypothesis will also reject Fodor's innateness argument.

Sterelny (1989) argues against Fodor's suggestion that humans have most of their concepts innately and are merely triggered to use them, just as a duckling is triggered to imprint on its mother. Ducklings can be triggered to imprint upon all sorts of things (animals of the wrong species, cuddly toys, and even the odd ethologist). Thus the idea of a trigger implies a certain arbitrariness about what does the triggering. However, suggests Sterelny, concept acquisition appears not to work this way. One's whale concept is caused by whales. One's doorknob concept is caused by doorknobs. Had an individual's doorknob concept been caused by whales, others would be inclined to think that the individual in question had a faulty doorknob concept. But if only doorknobs can cause a doorknob concept, then triggering begins to look a lot more like learning (see Cowie 1999, 69–139, for an extended discussion of this and other objections to Fodorian nativism; see Fodor 2001 for a response to Cowie).

There is similarly a long-standing tradition of close-fought argument against Chomskian nativism. Putnam (1992) claims that suboptimal features common to all languages need not be innate and that first-language learning is much more difficult and time-consuming than nativists assume.

The majority of criticism of Chomskian nativism has been based on empirical studies. However, in one crucial respect the foundations of nativism (Chomskian and otherwise) remain philosophically suspect. This is the suggestion that there is yet no substantive theory of exactly what it means for a concept to be innate (Cowie 1999). One such suggestion is that the putative understanding of innateness is really just a familiarity with certain metaphors (such as Leibniz' veins in marble, etc.). When one asks for a definition of what it means for a behavioral trait to be innate, one finds either a lack of a clear definition or a recognition of a confusing array of definitions.

A similar worry is that while philosophy provides a good characterization of what is meant by the term 'innate,' it does not do so in a way that leads naturally to useful scientific exploration of the putative phenomenon. So, for example, Stich (1975) argues that Descartes' suggestion should be followed to analyze innateness in terms of dispositions:

> A person has a disease innately at time *t*, if and only if, from the beginning of his life to *t* it has been true of him that if he is or were of the appropriate age (or at the appropriate stage of life) then he has or in the normal course of events would have the disease's symptoms. (6)

This characterization remains popular with some philosophers but is of little use to scientists. That is because it gives little advice as to how one might detect innate traits and no explanation as to their cause. How then might nativism respond to the charge that it fails to supply a compelling scientific theory of the nature of innateness? One possibility is to point out that the opponents of nativism are no better off. They claim that some portion of the doxastic furnishings is learned, but despite serious effort in this direction, there is still no good general theory of learning. This view is championed by Fodor (2001). Alternatively, the nativist might suggest that one does not need a general theory of innateness in order to be convinced that some things are indeed innate. After all, lack of a good theory about the nature of inheritance did not stop Darwin from employing inheritance in his theory of natural selection. Finally, the nativist might suggest that actually there already is a good theory of innateness. In this vein, some argue that cognitive science (and philosophy more generally) might avail itself of a biological notion of innateness (Fodor 2001, 102). Of course, this solution to the problem rests on there being a satisfactory biological notion of innateness.

The Biological Notion of Innateness

Innateness and inheritance are biologically distinct and yet closely linked. *Inheritance* is a relation of similarity that holds between generations by virtue of some biological process such as the passing on of genetic structure. *Innateness*, on the other hand, is a claim about a certain type of rigidity within the development of a biological individual (or within a group of individuals). Thus, congenital deformities (such as those caused by the drug thalidomide) are innate but not inherited. Conversely, at least on some accounts of innateness, regional accents in human speech are inherited but not innate.

Having said this, to accept the fact of inheritance is to acknowledge a certain rigidity in biological development, and to accept the fact of cumulative natural selection is to acknowledge a certain rigidity in evolutionary history. Thus one might infer from widespread acceptance of evolution to widespread acceptance of innateness. However, such inference has proved problematic. One stumbling block has been the great variety of ways in which innate traits can be characterized. They can be described variously as:

- Traits that develop in the absence of contact with conspecifics, such as web building in spiders;
- Traits that are characteristic of particular species, such as species-specific birdsong;
- Traits that are evolutionary adaptations, such as “dancing” in bees;
- Behaviors that are unlearned, including so-called reflex actions;
- Behaviors that develop fully formed in animals that have been prevented from practicing them.

While often in agreement, these definitions are not coextensive. Linguistic ability is very much characteristic of the human species, but it does not develop in the rare cases in which human children grow up in the absence of contact with conspecifics. So, some worry that there may be no single property that all purportedly innate characteristics have in common. Certainly there has been no agreed-upon definition in the scientific or philosophical literature despite a considerable amount of work toward this end. Much of this work has been done by ethologists (beginning with Konrad Lorenz) in the 1940s who sought to explain the peculiar character of a class of inherited behavioral characteristics in nonhuman animals. These behaviors have in common that they are performed in very stereotypical ways, are inherited, are triggered by relatively simple proximal stimuli, and can be triggered in circumstances in which their performance is disadvantageous to the animal in question. Ethologists argued that these traits are properly thought of as innate. However, they were well aware of the problem of providing a clear characterization of innateness. In light of the apparent profusion of possible definitions, Lorenz (1950, 261) argued that what these putatively innate behaviors have in common is that they are all genetic.

Are Innate Traits Genetic Traits?

Despite the *prima facie* plausibility of innate traits being the products of genetic structure, this idea has come under fire from a number of developmental biologists as well as philosophers of biology. In part this is because while it is true that all organisms inherit genetic structure from their parent or parents, it is false to suggest that this is the sum total of their inheritance (see Heritability).

Developmental systems theorists (such as Griffiths and Gray [1994]) point out many inherited characteristics are reliably passed from generation to generation via nongenetic channels. Organisms inherit taught hunting behaviors, food sources, and nest sites and even gut microfauna from their parents. None of these are transmitted via the inheritance of their parents’ DNA. Given this, one cannot infer that if innate traits are inherited, they must also be genetic.

If innate traits are to be characterized as genetic, there must first be a robust theory of what it means for a trait to be genetic. However, the idea that such a robust theory will be found has been the subject of considerable skepticism, most famously from Richard Lewontin (1974), who points out that one cannot mean that a “genetic trait” so called is caused exclusively by genes. All traits require both genetic and nongenetic precursors in their development. Furthermore, genes and environment interact in the course of development, so that one cannot determine the extent to which each is a cause of the development of any particular trait. Put more technically—genes and nongenetic developmental resources are typically nonadditive contributors to phenotype, and one therefore cannot partition the variance due to each (although one can partition the additive and nonadditive portions of variance) (see Heritability).

Recent Work on Innateness

Recent work has sought to tie the idea of innateness to particular biological processes. André Ariew (1999) proposes an account of innateness based on C. H. Waddington's notion of developmental canalization. This is the process by which developmental pathways are buffered against environmental perturbation. William Wimsatt (1999) argues that innateness is caused by generative entrenchment. This is the process by which adaptations that have been historically important become locked into a lineage as later adaptations are built atop them and are thus developmentally dependent upon them. But such strategies have the disadvantage of requiring the scientific community to settle on a particular biological process (generative entrenchment, canalization, or some other) as the source of all innateness. This leads back to the original problem with the biological characterization of innateness, namely that there appear to be a variety of processes that give rise to evolutionary and developmental rigidity.

Another alternative is to avoid Lewontin's argument by characterizing genetic traits in terms of genetic information (as did Lorenz, although his description of genetic information was rather sketchy). However this strategy has proved contentious (see Biological Information). Griffiths and Gray (1994) argue that all traits are products of information from both genetic and nongenetic developmental resources. Therefore, attempting to single out sources of developmental information will not provide an explanation of the distinctive nature of innate traits. Indeed, Griffiths has recently argued that the use of the term "innate" ought to be abandoned altogether (Griffiths 2002). However, Maclaurin (2002) suggests that one can nonetheless recognize particular groups of developmental resources (genes among them) as very unequivocal sources of developmental information about particular traits. Thus, innate traits can be characterized as those that are products of information from particular developmental resources that are maintained in populations by a variety of mechanisms. But this is likely to be controversial, as it rejects the idea that innate traits are necessarily genetic and it embraces the idea that innateness is a matter of degree.

The central advantage of making this move is that it focuses the study of innateness on the existence and nature of mechanisms that serve to maintain particular traits in biological populations. In doing so, it avoids the implausible assumption that the development of some traits is entirely ruled by the presence of some particular set of genetic (and perhaps environmental) precursors. This broadening of the notion of innateness is very much in line with recent findings in genetics. A study was begun in 1972 on the lives of 1,037 newborns in the city of Dunedin in the South Island of New Zealand. The most remarkable finding of the study to date has been a gene (monoamine oxidase A) that predisposes young people to violent behavior in later life. Those with low-active versions of the gene did four times the number of rapes, robberies, and assaults as they progressed to adulthood, *but only if they had also been maltreated as children.* Remarkably, the same maltreatment produced no corresponding psychological maladjustment in individuals with high-active versions of the gene (Caspi et al. 2002). No one would have called such environmentally mediated behavior innate in the old Lorenzian sense of the word, and yet here there is a very important and complex set of mechanisms that maintain a cycle of violence in a human community.

As more is learned about such inherited interactions, the study of innateness will increasingly be focused on the enhancement or amelioration of characteristics that are currently innate.

James Maclaurin

References

Ariew, A. (1999), "Innateness Is Canalization: In Defense of a Developmental Account of Innateness," in Valerie Gray Hardcastle (ed.), *Where Biology Meets Psychology*. Cambridge, MA: MIT Press, 117–138.

Caspi A., J. McClay, T. E. Moffitt, J. Mill, J. Martin, I. W. Craig, A. Taylor, and R. Poulton (2002), "Role of Genotype in the Cycle of Violence in Maltreated Children," *Science* 297: 851–854.

Chomsky, N. (1990), "On the Nature, Use and Acquisition of Language," in W. Lycan (ed.), *Mind and Cognition*. Cambridge, MA: Blackwell, 627–646.

——— (1965), *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cowie, Fiona (1999), *What's Within: Nativism Reconsidered*. New York: Oxford University Press.

Fodor, J. (2001), "Doing Without What's Within: Fiona Cowie's Critique of Nativism," *Mind* 110: 99–148.

——— (1981), "The Present Status of the Innateness Controversy," in J. Fodor (ed.), *RePresentations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: Bradford Books/MIT Press, 257–316.

——— (1975), *The Language of Thought*. New York: Crowell.

Griffiths, P. (2002), "What Is Innateness?" *Monist* 85: 70–85.

Griffiths, Paul, and Russell Gray (1994), "Developmental Systems and Evolutionary Explanation," *Journal of Philosophy* XCI: 277–304.

Leibniz, G. E. (1981), *New Essays on Human Understanding*. Translated by P. Remnant and J. Bennett. Cambridge: Cambridge University Press.

Lewontin, R. C. (1974), "The Analysis of Variance and the Analysis of Causes," *American Journal of Human Genetics* 26: 400–411.

Lorenz, Konrad (1950), "The Comparative Method in Studying Innate Behaviour Patterns," *Symposia of the Society for Experimental Biology* 4: 221–268.

Maclaurin, James (2002), "The Resurrection of Innateness," *Monist* 85: 105–130.

Marr, D. (1982), *Vision*. New York: W. H. Freeman.

Patrides, C. A. (ed.) (1970), *The Cambridge Platonists*. Cambridge, MA: Harvard University Press.

Putnam, H. (1992), "What Is Innate and Why: Comments on the Debate," in B. Beakley and P. Ludlow (eds.), *The Philosophy of Mind*. Cambridge, MA: MIT Press.

Sterelny, Kim (1989), "Fodor's Nativism," *Philosophical Studies* 55: 119–141.

Stich, Stephen P. (1975), "Introduction: The Idea of Innateness," in *Innate Ideas*. Berkeley and Los Angeles: University of California Press, 1–22.

Wimsatt, W. C. (1999), "Generativity, Entrenchment, Evolution, and Innateness: Philosophy, Evolutionary Biology, and Conceptual Foundations of Science," in Valerie Gray Hardcastle (ed.), *Where Biology Meets Psychology*. Cambridge, MA: MIT Press, 139–179.

See also **Biological Information; Chomsky, Noam; Cognitive Science; Heritability**

INSTRUMENTALISM

Though John Dewey coined the term *instrumentalism* to describe an extremely broad pragmatist attitude toward ideas or concepts in general, the distinctive application of that label within the philosophy of science is to positions that regard scientific theories not as literal and/or accurate descriptions of the natural world, but instead as mere tools or "instruments" for making empirical predictions and achieving other practical ends. This general instrumentalist thesis has, however, historically been associated with a wide variety of motivations, arguments, and further commitments, most centrally concerning the semantic and/or epistemic status of theoretical discourse (see below). Unifying all these positions is the insistence that one can and should make full pragmatic *use* of scientific theories either without believing the claims they seem to make about nature (or some parts of nature) or without regarding them as actually making such claims in the first place. This entry will leave aside the question of whether the term *instrumentalism* is properly restricted to only some subset of such views, seeking instead to illustrate the historical and conceptual relations they bear to one another and to related positions in the philosophy of science.

Loci Classici

Broadly instrumentalist sentiments concerning scientific theories have a remarkably long intellectual pedigree: indeed, Popper's (1963) famous critique of the position as intellectually sterile counts Andreas Osiander (author of the unsigned preface to Copernicus' *On the Revolutions of the Celestial Spheres*), Cardinal Bellarmino, and Bishop Berkeley as notable early defenders of the view (but cf. Fine 2001), even while resisting Duhem's claim to find its historical antecedents in classical Greek thinkers. Furthermore (as Popper and others note) the influential instrumentalism of nineteenth-century physicist Ernest Mach is rooted in a critique of Newtonian mechanics (and its concepts of absolute space, time, and motion) strikingly similar to Berkeley's own. Mach (1911) also resembles Berkeley in embracing a radical phenomenalism, insisting that what is represented " behind the appearances exists *only* in our understanding, and has for us only the value of a *memoria technica* or formula" (49): He argues that laws of nature (e.g., Snell's law) and theoretical hypotheses (e.g., the atomic hypothesis) are simply conceptual devices for the systematic classification, summary, organization, and coordinated expression and prediction of innumerable particular appearances (Mach [1893] 1960, 582f). Thus, Mach insists that theoretical concepts like 'atoms' are merely "provisional helps" and are ultimately to be dispensed with not because they seek unsuccessfully to describe a reality beyond appearances, but rather because they *successfully* but only *indirectly* describe coordinated and systematized collections of experiences themselves.

The instrumentalist impetus familiar from more recent philosophy of science, however, is rooted more fundamentally in developments within physics at the turn of the century, and in the related logical, epistemic, and historical concerns about the status of scientific theories articulated by thinkers like Pierre Duhem and Henri Poincaré (Worrall 1982) (see Duhem Thesis; Poincaré, Henri). The progress of physical science had by this time begun to suggest that there might be quite genuine cases of differences between actual competing scientific theories that could not possibly be adjudicated by any straightforward appeal to empirical tests or observations: to use a famous example of Poincaré's (though not a case of actual competing theories), any set of measurements of the angles in a triangle marked out by appropriately oriented perfectly rigid rods can be accommodated by the assignment of any number of different combinations of underlying spatial geometries and compensating "congruence relations" for the rods in question; if the sum of the angles differs from 180 degrees, for instance, one may either interpret the underlying geometry as Euclidean and conclude that the distance marked out by each rod varies with its position and/or orientation or assume that the distance marked out by each rod remains constant and conclude that the underlying geometry of the relevant space is non-Euclidean. Poincaré's response to this problem of theoretical underdetermination was *conventionalism*, that is, he regarded such theoretical matters as the assignment of a particular physical geometry to space as a matter of choice or convention to be decided on grounds of greatest convenience (see Conventionalism). And this in turn implied, he suggested, the distinctively instrumentalist conclusion that the quite useful ascription of a particular geometry to space by a theory should not be construed as literally attributing anything (truly or falsely) to nature itself: "[T]he question: Is Euclidean geometry true? . . . has no meaning. We might as well ask if the metric system is true, and if the old weights and measures are false. . . . One geometry cannot be more true than another; it can only be more convenient" (Poincaré [1905] 1952, 50).

To important, distinct worries about theoretical underdetermination Duhem added a further concern about the role played by idealizations in physical theories, and both he and Poincaré noted the long history of repeated and radical discontinuities in the dominant theoretical conceptions of particular domains of nature. But both argued that this history of scientific revolution and wholesale replacement is characteristic only of efforts to "surmise realities hidden under data observable by the senses" (Duhem [1914] 1954, 274). These data name only "the images we substituted for the real objects which Nature will hide forever from our eyes" (Poincaré [1905] 1952, 161). Thus, while both Duhem and Poincaré retained full confidence in the "experimental laws" or generalizations about observable phenomena uncovered by scientific investigations, each denied that such investigations were able to penetrate the actual constitution of nature, or that mathematical theories were able to describe it. Duhem went so far as to consign the explanatory ambitions of theories to the realm of metaphysics rather than science.

Thus recognizing that scientific theories or theorists *aspire* to describe an underlying, inaccessible reality and/or explain observable events by appeal to it, both Duhem and Poincaré simply reject these ambitions as ultimately either unscientific or unsatisfiable in some way. Both ranged at different times and in different works through a wide variety of importantly divergent attitudes (and not all the same ones between them) toward the cognitive, semantic, and epistemic status of theories, including the views that extant scientific theories were not in fact making claims about inaccessible realities behind observable phenomena, that the scientific enterprise need not do so, and that it should not. Moreover, there is reasonable controversy over classifying either thinker as ultimately an instrumentalist in any of these straightforward senses: in his latest work, Poincaré (1999) wholeheartedly embraced the reality of atoms, while Duhem consistently held that scientific theories are able to establish "natural classifications" of the phenomena.

Even this brief excursion through instrumentalist themes in Mach, Duhem, and Poincaré offers some sense of the variety of distinctive further commitments with which the general claim that scientific theories should be understood simply as useful instruments rather than accurate descriptions of inaccessible domains of nature has been conjoined. Among such further commitments are the following suggestions:

- Theoretical discourse is simply a device for organizing or systematizing beliefs about observational experience and its meaning is therefore exhausted by or reducible to any implications it has concerning observable states of affairs (reductive instrumentalism);
- Theoretical discourse has no meaning, semantic content, or assertoric force at all beyond the license it provides to infer some observable states from others (syntactic instrumentalism);
- Even if such discourse is both meaningful and irreducible, it can nonetheless be eliminated from science altogether (eliminative instrumentalism); and
- Even if the literal claims of theoretical science about the natural world are neither reducible, nor meaningless, nor even eliminable, such claims are nonetheless not to be believed (epistemic instrumentalism).

The Language of Science: Reductive, Syntactic, and Eliminative Instrumentalism

It is quite striking that even some of Duhem and Poincaré's explicit reservations about scientific theories have a semantic or linguistic character: Duhem ([1914] 1954) claims that "hypotheses [are] not judgments about the nature of things, only premises intended to provide consequences conforming to experimental laws" (39) and that theoretical propositions "are neither true nor false . . . only convenient or inconvenient" (334), while Poincaré ([1905] 1952) adds that the "object of mathematical theories is not to reveal to us the real nature of things," but "only . . . to co-ordinate the physical laws with which experiment makes us acquainted" (211). Perhaps less surprising is the fact that such a generally linguistic or semantic strategy of analysis was appealing to logical empiricist thinkers.

The logical empiricists' efforts to effect a reduction of all scientific language to a privileged phenomenological or observational basis (a project pursued most notably by the early Carnap, but also influentially by Bridgman) quite naturally grounded an instrumentalism about scientific theories of the sort described above as reductive (see Bridgman, Percy; Carnap, Rudolf). But even after this attempted reduction came to be widely regarded as a failure and such logical empiricists had given up the notion that the semantic content of apparently theoretical discourse was "really" exhausted by its implications concerning collections of observable events or subjective experiences, the distinctively syntactic variety of instrumentalism offered a fallback position: perhaps theoretical claims carry no straightforward ontological commitments regarding unobservable entities, even if they cannot be fully reduced to claims about immediately accessible experiences or states of affairs. More specifically, some logical empiricists suggested that theoretical claims are properly regarded as devoid of any semantic content whatsoever beyond the license they provide to draw inferences from one observable state of affairs to another. In the spirit of Duhem and Poincaré, this view regarded theoretical claims as nonassertoric, that is (appearances to the contrary), as not making claims about what the world is like and not possessing truth values at all.

Of course, this somewhat counterintuitive view of the semantics of theoretical discourse might be evaded by embracing the arguably more natural view (equally in the spirit of Duhem and Poincaré) that such discourse is simply *eliminable* from science altogether. This eliminative form of instrumentalism also gained considerable currency among logical empiricist thinkers, especially following the formulation and proof of an influential theorem by William Craig. Craig's theorem showed that for any recursively axiomatized first-order theory T, given any effectively specified subvocabulary O of T (mutually exclusive of and exhaustive with the remainder of the vocabulary of T), one can effectively construct another theory, T', whose theorems are exactly those of T that contain no nonlogical expressions besides those in O. As Hempel was the first to realize, this theorem implies that if the nonlogical vocabulary of any given scientific theory is partitioned into theoretical and observational components, the theory can be replaced with a "functionally equivalent" Craig transform that preserves all the deductive relationships between observation sentences established by T itself, since (by Craig's theorem) "any chain of laws and interpretive statements establishing [definite connections among observable phenomena] should then be replaceable by a law which directly links observational antecedents to observational consequents" (Hempel [1958] 1965, 186). This implied in turn, Hempel noted, that theoretical terms could be eliminated from theories altogether without losses in the purely observable consequences (deductively) obtainable from them, creating the following "Theoretician's Dilemma":

> If the terms and principles of a theory serve their purpose [of deductively systematizing the theory's observational consequences] they are unnecessary, as just pointed out; and if they do not serve their purpose they are surely unnecessary. But given any theory, its terms and principles either serve their purpose or they do not. Hence, the terms and principles of any theory are unnecessary. (Hempel [1958] 1965, 186)

The apparent feasibility of this eliminative instrumentalist program was further advanced by a related (and earlier, though largely unrecognized at the time) innovation of Frank Ramsey's: He proposed replacing any finitely axiomatized theory with a sentence that existentially generalizes on all the theoretical predicates of that theory. This "Ramsey sentence," he argued, has the same observational consequences as the original theory and therefore captures all the "factual content" of the original.

The significance of Craig's theorem was, however, immediately controversial. Nagel (1961), for instance, famously argued (136–137) that it is of quite limited relevance to the actual eliminability of theoretical discourse from science because:

(i) there is no guarantee that the axioms of T' delivered by Craig's method will not be "so cumbersome that no effective logical use can be made of them";
(ii) in fact, the axioms of T' will be infinite in number, no matter how simple the axioms of T itself, and correspond one-to-one with all the true statements expressible in the language of T', rendering them "quite valueless for the purposes of scientific inquiry"; and
(iii) Craig's method can actually be applied only if one knows, in advance of any deductions made from them, all the true statements in the restricted observational language.

In addition, Glymour (1980, Ch. 2) offers elegant technical objections to Ramsey's proposal, most importantly that as a theory of truth it fails to respect even the most elementary forms of demonstrative inference: For example, the Ramsey sentence of a conjunction may be necessarily false while the Ramsey sentences of each conjunct is individually true.

More recently, however, the profound differences between actual scientific theories and the sorts of artificial formal systems to which tools such as Craig's theorem and Ramsey's technique apply have led these formal results to be regarded as increasingly irrelevant to the genuine prospects for instrumentalism. More specifically, philosophers of science have become increasingly convinced that

(i) there is no strict, principled, or systematic division of the vocabulary of a theory into observational and theoretical parts;
(ii) the parts of a theory bear important logical, epistemic, and cognitive relations to one another that go far beyond what is captured by mere deductive systematization; and
(iii) scientific theories may not be best regarded as axiomatic formal systems in any case.

Thus, at least part of the solution to the Theoretician's Dilemma, as Hempel himself recognized, is to reject the claim that the only function of theoretical terms is to deductively systematize a theory's observational consequences.

Credibility and Belief: Epistemic Instrumentalism

Even as the philosophical fortunes of these distinctive semantic and eliminativist theses have declined, interest has remained strong in the broader instrumentalist conception of theories as tools for pursuing practical ends rather than accurate descriptions of nature itself. The most influential recent approaches have pursued this conception by exchanging the logical empiricists' reductive, syntactic, and eliminativist commitments for epistemic alternatives, that is, more recently influential forms of instrumentalism grant both the assertoric force and the ineliminability of theoretical claims but insist that such theories should simply be *used* for prediction of experimental outcomes and other practical goals without a requirement of *belief* in the claims they in fact make about nature itself (or some parts thereof). A further recent trend has been to make the case(s) for instrumentalism piecemeal, arguing that quite specific features of a given scientific theory (e.g., quantum mechanics, evolutionary biology) either require or recommend an instrumentalist stance toward that theory.

One prominent example of general instrumentalism of this epistemic variety is constructive empiricism of Bas van Fraassen (1980). Like Duhem and Poincaré, van Fraassen appeals to the underdetermination of theories by evidence to challenge the conclusion that empirically successful scientific theories describe what inaccessible domains of nature are really like, and he insists that even a reflective endorsement of the actual inferential and other practices of science itself requires only a cognitive attitude of *acceptance* toward theories, rather than belief. He argues that it is epistemically supererogatory to believe any more of scientific theories than that they are *empirically adequate*, that is, that what they say about *observable* phenomena is true, and he insists that epistemic prudence recommends agnosticism regarding even the most successful theories' claims about unobservables. Thus, constructive empiricism regards scientific theories as reliable tools for anticipating how observables will behave, while it resists the conclusion that such theories describe what unobservable domains of nature are really like—but on epistemic rather than semantic grounds.

Of course, constructive empiricism still relies fundamentally on an extremely controversial distinction (albeit itself naturalized) between observables and unobservables, so it is important to note that the distinctively epistemic form of instrumentalism *need* not rely upon any such distinction: As Fine (1991) argues the guiding commitment of instrumentalism is simply to the *reliability* of a causal story, which "treats all entities (observable or not) perfectly on par":

> Of course if the cause happens to be observable, then the reliability of the story leads me to expect to observe it (other things being equal). If I make the observation, I then have independent grounds for thinking the cause to be real. If I do not make the observation or if the cause is not observable, then my commitment is just to the reliability of the causal story, and not to the reality of the cause. (86)

Perhaps the most fully developed form of epistemic instrumentalism that eschews any important distinction between observables and unobservables is the historically oriented variety inspired by thinkers like Thomas Kuhn and pursued more recently by Larry Laudan. Like Duhem and Poincaré, these thinkers draw centrally on the history of repeated fundamental changes over time in the descriptions of nature offered by dominant scientific theories, in support of a skeptical attitude toward the claims of the dominant scientific theories of the present day. Kuhn ([1962] 1996) not only appeals to this history to undermine the notion that contemporary science is in possession of any final theoretical truth about a stable natural world, but also famously claims that the very "notion of a match between the ontology of a theory and its 'real' counterpart in nature now seems to me illusive in principle"; nonetheless he insists that scientific theories have improved over time "as instruments for puzzle-solving" (206). Laudan (1981a) argues that the long historical record of successful but ultimately rejected scientific theories undermines any justification for inferring even the approximate truth of contemporary scientific characterizations of nature (observable or not) from their dramatic empirical successes, but insists nonetheless not only that such theories can and should be used to tackle and solve a wide variety of empirical and conceptual problems, but also that there is a clear sense in which cumulative progress in this regard has been achieved over time, by attaining with the theoretical instruments of science an ever larger and more various set of effective solutions to such problems (Laudan 1977 and 1981b) (see Kuhn, Thomas).

As these influential formulations of the view illustrate, epistemic instrumentalism seems committed to some distinction between believing a theory to be true and accepting or using it *without* believing what it says. Perhaps unsurprisingly, the cogency of this distinction has itself been the target of recent influential criticisms of epistemic instrumentalism, on the grounds that these cognitive attitudes simply cannot be distinguished in the way that one or more forms of instrumentalism require. Horwich (1991) points out, for example, that some accounts of belief itself simply identify it as the mental state responsible for use, while Blackburn (1984) argues that there is no room for a distinction between merely "accepting" a statement with a truth condition and simply believing it to be true (see Fine 1986, espectially sec. 4). By contrast, Sober (2002) defends the distinction, pointing out not only that idealized models known to be false are often accepted or used as the basis for accurate predictions across a range of phenomena, but also that recent work in model selection theory shows why models (statements containing adjustable parameters) known to be false will routinely serve as the basis for more accurate predictions of new data than competitors known to have higher likelihood conferred on them by the available data or even by competitors known to be true. Thus, he argues, not only is there a genuine difference between the goal of seeking instrumental or predictive reliability and the goal of seeking truth, but this distinction is respected within scientific practice itself, which typically chooses models (with adjustable parameters) with the former and fitted models (once parameters have been adjusted) with the latter goal in mind.

In a related vein, Nagel (1961, 139) famously argues that there is a "merely verbal" difference between the instrumentalist contention that a theory offers satisfactory techniques of inference and the realist contention that it is true. More recently, Stein (1989) has argued that the dispute between realism and instrumentalism is not well joined: There would be no appreciable difference (or no difference that makes a difference) between the two positions once

(a) realism becomes sophisticated enough (as Stein suggests it must) to (i) give up its pretensions to metaphysically transcendent theorizing, (ii) eschew aspirations to noumenal truth and reference, and (iii) abandon the idea that a property of a theory might somehow explain its success in a way that does not simply point out the use that has been made of the theory; and

(b) instrumentalism becomes sophisticated enough (as Stein suggests it must) to recognize the scope of a theory's role *as an instrument* to include not just calculating experimental

> outcomes, but also adequately representing phenomena in detail across the entire domain of nature and providing resources for further inquiry.

Thus, Stein argues for a convergence between the appropriately restricted ambitions a sophisticated realism holds out for theories and the appropriately expanded ambitions a sophisticated instrumentalism holds out for them, and indeed, that in the work of the deepest scientists (his examples are Maxwell, Newton, and Einstein) the two attitudes are present together in such a way that the alleged contradiction between them simply vanishes. Thus, even as instrumentalism persists as a viable and influential position in the contemporary philosophy of science, its comparative merits and even the coherence of its formulation remain the subject of deservedly intense controversy.

P. KYLE STANFORD

The author acknowledges the helpful input of Arthur Fine, Bas Van Fraassen, Bill Demopoulos, Elliott Sober, Larry Laudan, David Malament, Aldo Antonelli, Jeff Barrett, Stathis Psillos, and Philip Kitcher. This material is based upon work supported by the National Science Foundation under Grant No. SES-0094001. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

References

Blackburn, Simon (1984), *Spreading the Word*. Oxford: Clarendon Press.

Duhem, Pierre ([1914] 1954), *The Aim and Structure of Physical Theory*. Translated by Philip P. Wiener. Princeton, NJ: Princeton University Press.

Fine, Arthur (1986), "Unnatural Attitudes: Realist and Instrumentalist Attachments to Science," *Mind* 95: 149–179.

——— (1991), "Piecemeal Realism," *Philosophical Studies* 61: 79–96.

——— (2001), "The Scientific Image Twenty Years Later," *Philosophical Studies* 106: 107–122.

Glymour, Clark (1980), *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Hempel, Carl ([1958] 1965), "The Theoretician's Dilemma: A Study in the Logic of Theory Construction," in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Horwich, Paul (1991), "On the Nature and Norms of Theoretical Commitment," *Philosophy of Science* 58: 1–14.

Kuhn, Thomas S. ([1962] 1996), *The Structure of Scientific Revolutions*, 3d ed. Chicago: University of Chicago Press.

Laudan, Larry (1977), *Progress and Its Problems*. Berkeley and Los Angeles: University of California Press.

——— (1981a), "A Confutation of Scientific Realism," *Philosophy of Science* 48: 19–49.

——— (1981b), "A Problem-Solving Approach to Scientific Progress," in Ian Hacking (ed.), *Scientific Revolutions*. New York: Oxford University Press, 144–155.

Mach, Ernest ([1893] 1960), *The Science of Mechanics*, 6th ed. Translated by T. J. McCormack. La Salle, IL: Open Court.

——— (1911), *History and Root of the Principle of the Conservation of Energy*. Translated by P. E. B. Jourdain. Chicago: Open Court.

Nagel, Ernest (1961), *The Structure of Science*. New York: Harcourt, Brace and World.

Poincaré, Henri ([1905] 1952), *Science and Hypothesis*. New York: Dover.

Popper, Karl. R. (1963), *Conjectures and Refutations*. London: Routledge and Kegan Paul, chap. 3.

Psillos, Stathis (1999), *Scientific Realism: How Science Tracks Truth*. New York: Routledge.

Sober, Elliott (2002), "Instrumentalism, Parsimony, and the Akaike Framework," *Philosophy of Science* 69: S112–S123.

Stein, Howard (1989), "Yes, But . . . Some Skeptical Remarks on Realism and Anti-Realism," *Dialectica* 43: 47–65.

van Fraassen, Bas C. (1980), *The Scientific Image*. Oxford: Oxford University Press.

Worrall, John (1982), "Scientific Realism and Scientific Change," *Philosophical Quarterly* 32: 201–231.

See also **Conventionalism; Empiricism; Logical Empiricism; Phenomenalism; Realism; Theories**

INTENTIONALITY

Some things are *about,* or are *directed on,* or *represent* other things. For example, the sentence 'Cats are animals' is about cats (and about animals); this entry is about intentionality; Emanuel Leutze's most famous painting is about Washington's crossing of the Delaware; lanterns hung in Boston's

North Church were about the British; and a map of Boston is about Boston. In contrast, #a$b, a blank slate, and the city of Boston are not about anything. Many mental states and events also have "aboutness": the belief that cats are animals is about cats, as is the fear of cats, the desire to have many cats, and seeing that the cats are on the mat. Arguably some mental states and events are not about anything: Sensations, like pains and itches, are often held to be examples. Actions can also be about other things: Hunting for the cat is about the cat, although tripping over the cat is not. This (rather vaguely characterized) phenomenon of aboutness is called intentionality. Something that is about (directed on, represents) something else is said to "have intentionality" or to be an "intentional mental state."

This medieval terminology was reintroduced by the Austrian philosopher, Franz Brentano ([1874] 1995), in his book *Psychology from an Empirical Standpoint*, although Brentano himself did not use the word "intentionality." (For a brief history of the terminology, and further references, see Crane 1998a; for an account of Brentano's thought, see Moran 2000, Ch. 2.) In a famous passage, Brentano ([1874] 1995) claimed that every mental state/event has intentionality:

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction towards an object (which is not to be understood here as meaning a thing), or immanent objectivity. In presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired, and so on. (88)

Brentano's use of "intentional inexistence" is liable to confuse. Brentano did not mean that mental states are about peculiar nonexistent objects, but was rather referring to the (admittedly obscure) sense in which the object of a mental state is "in" the mind. The terminology of intentionality can also be confusing, for at least two reasons. First, intentionality has nothing in particular to do with intending, or intentions. Intentions—for instance, the intention to adopt a cat—are just one of many types of intentional mental states. Second, inten*t*ionality must be sharply distinguished from inten*s*ionality (Searle 1983). Mental states are not intensional, only sentences (and, for some sentences, other linguistic entities). A sentence *S* is intensional, or is an *intensional context*, just in case substitution of some expression *a* in *S* with some coreferring expression *b* yields a sentence with different truth value from the truth value of *S*. So, for example, "Necessarily, the number of planets is nine" and "Hegel believed that the number of planets is seven" are intensional. Substituting "nine" for the coreferential "the number of planets" turns the first false sentence into the true sentence "Necessarily, nine is nine" and the second true sentence into the false sentence "Hegel believed that nine is seven." As the first example indicates, a sentence can be intensional and yet have nothing to do with intentionality. Conversely, sentences that report intentional mental states/events need not be intensional (Crane 1998a). For example, "Berkeley heard the coach" is (arguably) not intensional: If that sentence is true, and if "the coach" and "Locke's favorite carriage" refer to the same thing, then "Berkeley heard Locke's favorite carriage" is true.

Paradoxes of Intentionality

As informally explained above, an intentional mental state (for example) is "about" something. The belief that Brentano is Austrian is about Brentano. The object that the state is about is called the intentional object of the state. (Intentional objects are sometimes taken to include states of affairs as well as particulars like Brentano: The belief that Brentano is Austrian could be said to be about Brentano's being Austrian.) So there should be a relation of aboutness that holds between a mental state and an object just in case the state is about the object—"the intentional relation," in Brentano's terminology.

Thinking of intentionality in this way, as a relation to intentional objects, leads to three classic "paradoxes of intentionality" (Thau 2002). The first paradox is that the intentional object need not exist (at any time). The belief that the fountain of youth is in Florida bears the intentional relation to the fountain of youth, and the fountain of youth does not exist. But if *a* is related to *b*, then there is such a thing as *a* and such a thing as *b*. One rather extreme solution, famously proposed by Brentano's student Alexius Meinong, is to hold that there are objects that do not exist. In this view, there *is* a fountain of youth, and the belief that the fountain of youth is in Florida bears the intentional relation to that (nonexistent) object. (This is Meinong's view, but not his terminology. Meinong used "subsists" to mean *exists*, and "exists" to mean something like *spatiotemporally exists*. Thus, in Meinong's terminology, Mont Blanc exists, the number 7 subsists but does not exist, and the fountain of youth neither exists nor subsists.)

The second paradox is that a mental state can bear an intentional relation to something without there being any particular thing that the state bears the relation to. If one wants a cat, but has no particular cat in mind, then one's state of wanting a cat bears the intentional relation to an object—a cat, presumably—yet there is no particular cat that the state bears the intentional relation to. But if *a* is related to something, then there is a particular object that *a* is related to.

The third paradox is that a mental state can bear the intentional relation to *a,* but not bear the intentional relation to *b*, even though *a* is *b.* The belief that the first postmaster general was a United States president is about the first postmaster general, but not about the inventor of bifocals, even though the inventor of bifocals was the first postmaster general, namely Benjamin Franklin. But if *a* bears a certain relation to *b*, and $b = c$, then *a* is related by the same relation to *c.*

In *Psychology from an Empirical Standpoint* Brentano ([1874] 1995) himself did appear to think that a mental state was always related to an intentional object, but in an appendix he insisted that the "only thing which is required by mental reference is the person thinking. The terminus of the so-called relation does not need to exist in reality at all" (272).

The moral of the paradoxes of intentionality is that thinking of intentionality in terms of the intentional relation is a bad idea. A better way involves drawing a distinction between the *representational content* of a mental state (or some other thing that has intentionality) and the objects (if any) the mental state is about. So, for example, the belief that the fountain of youth is in Florida has as its content the proposition that the fountain of youth is in Florida, and there is no object that the belief is about—at any rate, not the fountain of youth (the belief is about Florida). To believe that the fountain of youth is in Florida is to stand in the belief-relation to the proposition that the fountain of youth is in Florida. This proposition exists whether or not the fountain of youth does (it does not contain the fountain of youth as a constituent), and this proposition is true just in case there is such a thing as the fountain of youth and it is located in Florida. Similarly, the desire that one have a cat has as its content the proposition that one has a cat, and there is again nothing that the belief is about—at any rate, no particular cat. Finally, the belief that the first postmaster general was a United States president and the belief that the inventor of bifocals was a United States president are both about the same object, namely Benjamin Franklin. However, there is some truth behind the original mistaken claim that the two beliefs are about different objects. This can be brought out by noting that the contents of the two beliefs are true at different possible worlds (of course, the contents are both false at the actual world). Specifically, the first proposition, but not the second, is true at a possible world in which the first postmaster general became president and the inventor of bifocals never entered politics. The truth of the first proposition at a world depends on the political fortunes of whomever is the first postmaster general at that world—whether or not that individual invented bifocals.

Brentano's Two Theses

Brentano ([1874] 1995) proposed two theses that form the basis of contemporary discussions of intentionality:

1. No "physical phenomenon" has intentionality.
2. Intentionality is *the mark of the mental:* All and only mental states/events have intentionality.

> [I]ntentional in-existence is characteristic exclusively of mental phenomena. No physical phenomenon exhibits anything like it. We can, therefore, define mental phenomena by saying that they are those phenomena which contain an object intentionally within themselves. (89)

Brentano's examples of physical phenomena were not, say, brain processes, but were (chiefly) perceptible properties, like "color, sound and warmth" (92). Nonetheless, mainly through the influence of Roderick Chisholm, Brentano came to be associated with the doctrine that intentionality is not reducible to the physical—in the contemporary sense of 'physical' (see Physicalism). Although quite dubious as an interpretation of Brentano (Moran 1996), it started a debate that continues to this day. Chisholm himself argued for the irreducibility of intentionality by first transforming this thesis into one about the sort of language adequate for psychology. Thus recast, the thesis of the irreducibility of intentionality becomes one about the ineliminability of intensional contexts, like "Revere believes that the British are coming," in the language of a scientific psychology. (Chisholm called such sentences "intentional sentences.")

Chisholm's (1957) reformulation of "a thesis resembling that of Brentano" is:

> [W]e do not need to use intentional language when we describe non-psychological phenomena; we can express all of our beliefs about what is merely 'physical' in sentences which are not intentional. But … when we

> wish to describe perceiving, assuming, believing, knowing, wanting, hoping, and other such attitudes, then either (a) we must use language which is intentional or (b) we must use terms we do not need to use when we describe nonpsychological phenomena. (172–173)

Chisholm argued for his "linguistic version" of Brentano's first thesis as opposed to various behavioristically inspired analyses of "intentional language." Chisholm did not conclude that the failure of reduction impugned the reality of intentional mental states, but Quine ([1960] 1998) famously did:

> One may accept the Brentano thesis as either showing the indispensibility of intentional idioms and the importance of an autonomous science of intention, or as showing the baselessness of intentional idioms and the emptiness of a science of intention. My attitude, unlike Brentano's, is the second. (221)

Many philosophers are not so pessimistic, and there are many suggestions for providing a physicalistic or naturalistic reduction of intentionality. This is discussed in the following section.

Brentano's first thesis is true but has been (fruitfully) misinterpreted. Brentano's second thesis, on the other hand, has been correctly interpreted but seems obviously false, because of examples given in the first paragraph of this article. However, as discussed in the final section, Brentano's second thesis is in better shape than initial appearances suggest.

Reducing Intentionality

Many philosophers hold that there must be a physicalistic/naturalistic reduction of intentionality—at least if intentionality is a genuine phenomenon. Fodor (1987) is a prominent example:

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin, charm,* and *charge* will perhaps appear on their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. . . . If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or maybe of their supervenience on?) properties that are themselves *neither* intentional *nor* semantic. If aboutness is real, it must be really something else. (97)

There are many different approaches to providing the reduction of intentionality that Fodor says we need. Most adopt some kind of divide-and-conquer strategy. First, a distinction is made between *original* intentionality and *derivative* intentionality (Haugeland 1998; see Searle 1992 for a similar distinction between *intrinsic* and *derived* intentionality). A thing has derivative intentionality just in case the fact that it represents such-and-such can be explained in terms of the intentionality of something else; otherwise it has original intentionality. Often the intentionality of language and other sorts of conventional signs is said to be derivative. Language, in this view, inherits its intentionality from that of mental states, specifically from the intentions and conventions adopted by language users (Grice 1989). This is an attractive and plausible claim, although it is not obvious, and has been denied (see e.g., many of the essays in Davidson 1985). However, if it is correct, then the problem of reducing intentionality is itself reduced to the problem of reducing the intentionality of the mental.

Theories that attempt to provide a physicalistic reduction of intentionality fall into three broad groups. The first group comprises causal covariational theories (Stampe 1977; Dretske 1981; Stalnaker 1984; Fodor 1990). The basic idea is that mental states represent in much the same way that tree rings represent. The number of rings on a tree represents the tree's age, because the fact that the tree's age is n years old causes the tree to have n rings, or (a refinement) would cause the tree to have n rings in optimal conditions. A simple example of a causal covariational theory is this: A belief state S represents that p (that is, has propositional content that p) if and only if the fact that p would cause a subject to be in S. (This formulation takes the notion of a *belief state* for granted; a physicalistically acceptable version of the theory would have to provide a further reduction of a belief state.)

The second group comprises teleological theories (Papineau 1987; Millikan 1993; Dretske 1995). The basic idea is to explain the intentionality of mental states in terms of their biological functions, which might in turn be given a reductive account in terms of evolutionary history. A simple example is this: A belief state S represents that p if and only if in conditions in which a subject's cognitive system is functioning as it is designed to function by evolution, the subject would be in S when and only when it is the case that p (see Function).

The third group comprises functional role theories. Here the basic idea is that a representation or symbol means what it does because of its *functional role*—its causal interaction with other representations. A simple example (for public language): A two-place sentence connective * means *and* if and only if the acceptance of sentence P and sentence Q is disposed to cause the acceptance of the sentence $P \wedge * \wedge Q$ (i.e., P concatenated with * concatenated with Q), and the acceptance of $P \wedge * \wedge Q$ is disposed to cause the acceptance of P and the acceptance of Q. If this is to be an account of thought rather than language, then there must be an appropriate range of neural representations—perhaps words in a "language of thought." In "long-armed" theories, functional roles are taken to include causal interactions with the environment (Harman 1999); "short-armed" functional role theories exclude such causal interactions, and for that reason are often taken to be accounts of the so-called "narrow content" of mental states (Block 1986).

Two other notable approaches should be mentioned. One is Dennett's (1987) instrumentalism, which attempts to vindicate intentional notions from a physicalistic perspective without providing explicit reductions of the sort just illustrated. The other is Brandom's (1994) inferentialism, which attempts to reduce intentionality to normativity, in particular to norms governing inferential practices.

Intentionality as the Mark of the Mental

Brentano's second thesis is independent of (the misinterpretation of) his first, that intentionality cannot be given a physicalistic reduction. The irreducibility of intentionality does not imply that all and only mental states/events are intentional. Searle is an example of a philosopher who holds that intentionality is irreducible, yet that sensations are not intentional. Neither does the converse implication hold: if intentionality is the mark of the mental, it might still be reducible. Tye and Dretske, whose views are mentioned below, think that intentionality is the mark of the mental and that it can be given a physicalistic reduction.

Brenanto's second thesis divides into two parts:

- Intentionality is *sufficient* for mentality and
- intentionality is *necessary* for mentality.

The sufficiency claim is false—at least if 'intentionality' is used in the broad and loose contemporary way, to include nonmental entities like sentences, paintings, and maps (see the beginning of this article). However, the sufficiency claim might be amended as follows: Original intentionality is sufficient for mentality. According to the revised sufficiency claim, the mental is the source of all intentionality. This revised claim still faces problems. First, if the indication of a tree's age by the number of its rings is an example of intentionality at all, then it is presumably original intentionality. And if it is original intentionality, the sufficiency claim is false. Again, the sufficiency claim is false if the intentionality of language does not derive from the intentionality of the mental. But these are controversial issues, and there is at least some prospect of defending a modified version of Brentano's sufficiency claim.

Matters might seem even less promising with the other part of Brentano's second thesis, the claim that intentionality is necessary for mentality. At any rate, some philosophers think that sensations are obviously nonintentional. However, the claim that bodily sensation is a form of perception of one's own body was defended in the 1960s by D. M. Armstrong (1968) and has been revived today by a number of philosophers including Dretske (1995), Lycan (1996), and Tye (1995). And if this thesis is correct, then because perceptions have intentionality, bodily sensations are not counterexamples to the claim that intentionality is necessary for mentality. (See Brentano [1874] 1995, 82–85, for his account of the intentionality of pain, which anticipates many modern discussions.)

More problematic cases are provided by certain "objectless" emotions, like forms of anxiety or depression, where one is hard put to say what one is anxious or depressed about (Searle 1983). For defenses of Brentano's second thesis against this sort of example, see Tye (1995, 128–131) and Crane (1998b).

Assuming that every mental state/event is intentional, a further issue arises, whether the representational content of a mental state determines "what it's like" to be in the state—the state's qualia. Dretske, Lycan, and Tye, among others, endorse this determination claim. Such an "intentional theory of qualia" is controversial and has been widely discussed in the literature on consciousness.

Alex Byrne

References

Armstrong, David (1968), *A Materialist Theory of the Mind*. London: Routledge.

Block, Ned (1986), "Advertisement for a Semantics for Psychology," *Midwest Studies in Philosophy* 10: 615–678.

Brandom, Robert (1994), *Making it Explicit*. Cambridge, MA: Harvard University Press.

Brentano, Franz ([1874] 1995), *Psychology from an Empirical Standpoint*. Translated by Antos C. Rancurello, D. B. Terrell, and Linda L. McAlister. London: Routledge.

Chisholm, Roderick (1957), *Perceiving: A Philosophical Study*. Ithaca, NY: Cornell University Press.

Crane, Tim (1998a), "Intentionality," in Edward Craig (ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge.

——— (1998b), "Intentionality as the Mark of the Mental," in Anthony O'Hear (ed.), *Current Isssues in Philosophy of Mind* (Royal Institute of Philosophy Supplement 43). Cambridge: Cambridge University Press, 229–251.

Davidson, D. (1985), *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Dennett, Daniel (1987), *The Intentional Stance*. Cambridge, MA: MIT Press.

Dretske, Fred (1981), *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

——— (1995), *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Fodor, Jerry (1987), *Psychosemantics*. Cambridge, MA: MIT Press.

——— (1990), *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Grice, Paul (1989), *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, part 1.

Harman, Gilbert (1999), "(Nonsolipsistic) Conceptual Role Semantics," in *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press, 206–231.

Haugeland, John (1998), "The Intentionality All-Stars," in *Having Thought*. Cambridge, MA: Harvard University Press, 127–170.

Lycan, W. G. (1996), *Consciousness and Experience*. Cambridge, MA: MIT Press.

Millikan, Ruth (1993), *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Moran, Dermot (1996), "Brentano's Thesis," *Proceedings of the Aristotelian Society, Supplementary Volume* 70: 1–27.

——— (2000), *Introduction to Phenomenology*. London: Routledge.

Papineau, David (1987), *Reality and Representation*. Oxford: Blackwell.

Quine, Willard ([1960] 1998), *Word and Object*. Cambridge, MA: Harvard University Press.

Searle, John (1983), *Intentionality*. Cambridge: Cambridge University Press, chap. 1.

——— (1992), *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Stalnaker, Robert (1984), *Inquiry*. Cambridge, MA: MIT Press.

Stampe, Dennis (1977), "Towards a Causal Theory of Linguistic Representation," *Midwest Studies in Philosophy* 2: 42–63.

Thau, Michael (2002), *Consciousness and Cognition*. Oxford: Oxford University Press, chap. 2.

Tye, Michael (1995), *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

See also **Function; Naturalism; Physicalism; Searle, John; Teleology**

IRREVERSIBILITY

Chunks of ice melt in warm water, but warm water never spontaneously forms chunks of ice. Gases expand to fill their containers, but uniformly spread gasses never contract into one corner of their (isolated) containers. These examples illustrate an incredibly pervasive regularity. Certain types of processes proceed only in one direction: they occur, but their time reverses do not. And this appears to be a matter of law. These processes are said to be irreversible. (This notion of irreversibility should be distinguished from a similarly termed notion that appears in classical thermodynamics: A system is said to undergo a "reversible change" when it changes so slowly that it remains very near thermal equilibrium throughout [Uffink 2001].)

Irreversible processes typically involve temperature-difference equalization, diffusion, the completion of chemical reactions, and certain phase transitions. More generally, isolated systems (or, rather, systems that can be treated as isolated) tend to move toward states of equilibrium. What explains this regularity? The simplest hypothesis is that fundamental laws of nature explicitly require that isolated systems tend toward equilibrium, and the science of thermodynamics provides a system of postulates that yields just this constraint.

Thermodynamics introduces a physical quantity, *thermodynamic entropy*, which is defined for systems at equilibrium. The thermodynamic entropy of a system measures how much of the system's energy is available for conversion into useful work (the higher the entropy, the less energy is available). The second law of thermodynamics says that the entropy of an isolated system never decreases (Sklar 1993, 21).

Irreversible processes—for example, a chunk of ice melts in warm water, a gas spreads to fill its container—involve increases in entropy. The second law rules out the time reversals of these processes because this would involve *decreases* in entropy. Despite the elegance and practical indispensability of classical thermodynamics, its postulates do not have the character of fundamental dynamical laws, which are thought to govern the detailed motions of the microscopic constituents of matter.

The natural next suggestion is that the fundamental dynamical laws themselves are time asymmetric and this asymmetry helps explain irreversibility. For example, particle physics has produced evidence that there are *T-symmetry violations*—interactions that have (slightly) different chances than their time reverses. So it is believed that the dynamical laws are not time symmetric. However, these slight differences in chances do not have a significant effect on the manner in which macroscopic systems undergo thermodynamic change. So this time asymmetry in the laws does not help explain irreversibility (Sklar 1993, 248).

Thus the following question is left open: How does the sort of irreversible behavior captured by the postulates of thermodynamics arise from the fundamental dynamical laws?

Statistical Mechanics

It is simplest to introduce statistical mechanics in the context of classical mechanics. As a simple example, consider a number of billiard balls, undergoing perfectly elastic collisions on a frictionless table (Figure 1). Suppose that at some initial time the balls are concentrated in the upper-left corner of the table and that they spread out over the course of a minute. Now perform a thought experiment. At the end of the minute, stop time, reverse the velocities of the balls, and start time again. The balls will retrace their original paths and return to the corner of the table in which they began.

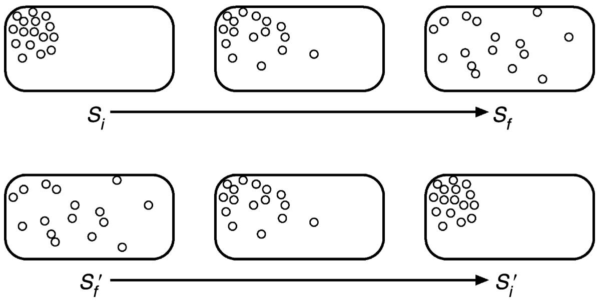


Fig. 1. The initial state of the table (S_i), in which the balls are concentrated, evolves into a final state (S_f), in which the balls are dispersed. Reversing the velocities of the balls in S_f results in a state S_f', which evolves into a state S_i' in which the balls are concentrated (Goldstein 2001).

This thought experiment illustrates a striking fact about classical mechanical laws of motion: Whenever a process is allowed by the laws, so is the time reverse of that process. So, given just these laws, there is no hope of showing that entropy-decreasing processes are downright disallowed. The best that can be hoped for is a statistical argument that entropy-decreasing processes are highly improbable. This observation is known as the reversibility objection (famously put to L. Boltzmann by J. Loschmidt [Sklar 1993, 35]).

There are many approaches to producing such an argument (Sklar 1993). One approach introduces the notion of *Gibbs entropy.* Gibbs entropy is a quantity defined not for individual systems, but for probability distributions over phase space (measuring the extent to which such distributions are spread out). This approach seeks to explain irreversibility by deriving certain facts about how such probability distributions evolve under the laws. It has been objected that this sort of result does not address the phenomenon to be explained, *viz.,* that *individual* isolated systems tend to increase in entropy (Lebowitz 1993; Goldstein 2001; Maudlin 1995). *Interventionism* attempts to avoid the reversibility objection by observing that thermodynamic systems interact with their environments. This observation is correct but does not avoid the difficulty, for one can shift attention to a larger system that is not subject to such interference.

The above difficulties are avoided by a highly influential approach to explaining irreversibility, which has its roots in the work of Ludwig Boltzmann.

Phase Space

The *phase space* of a system is a set of points, each of which is a dynamical state for the system to be in at a time. In the case of a number of classical point particles, each point of phase space determines the position and momentum of each particle. Phase space is also equipped with some additional geometric structure, including a measure that determines the volume of each of its regions. (Since the total energy of a classical system remains constant over time, attention can be restricted to that portion of phase space associated with some particular fixed total energy of the system.)

Phase space is carved up into disjoint sets, called *macrostates* (Figure 2). Points of phase space (or *microstates*) that are in the same macrostate are alike with respect to macroscopic parameters. For

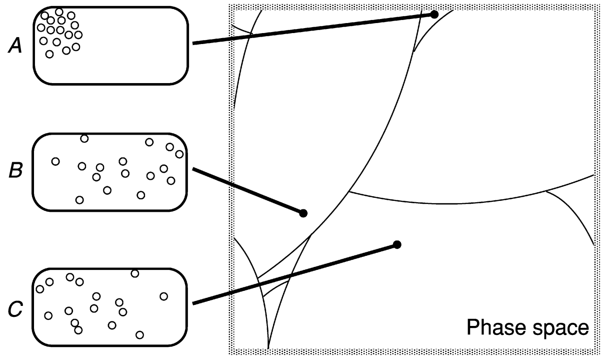


Fig. 2. Schematic representation of the phase space of a system of hard spheres. The space is divided into macrostates, which vary greatly in size. The space is dominated by equilibrium states such as *B* and *C*, which are shown in the left-hand column. Only a tiny proportion of phase space consists of far-from-equilibrium states such as *A*.

example, they have roughly the same temperature and pressure distributions.

Boltzmann noticed that for combinatorial reasons, macrostates vary greatly in size, and the variation is systematic. If all else is equal, macrostates in which the particles are spread out over physical space are bigger than ones in which they are clumped together, and macrostates in which the particles are moving with a large variety of momenta are bigger than ones in which, say, all have the same momentum.

That observation motivates the following definition: The statistical-mechanical entropy (also known as the *Boltzmann entropy*) of a point in phase space is a measure of the size of the macrostate to which it belongs—the bigger the macrostate, the greater the entropy. More precisely: the statistical-mechanical entropy of a point in phase space is proportional to the logarithm of the volume of the macrostate to which it belongs. For macroscopic systems, the imbalance in sizes of macrostates is overwhelming. Virtually all of phase space consists of points representing states in which the system is at equilibrium. This vast imbalance in size of macrostates can be used to explain why irreversible processes are so common.

A Statistical Explanation of Irreversibility

Consider a system consisting of a (nearly isolated) gas confined to a box. Take a particular low-entropy macrostate *L*, in which the gas is concentrated in the left half of the box. Notice that the phase space for this system is dominated by points whose entropies are *higher* than that of *L*. In other words, phase space is dominated by points in which the gas is *more spread out* than it is in *L*. So it is reasonable to think that practically all of the states in *L* have futures in which entropy *increases*. In other words, it is reasonable to think that practically all of the states in *L* have futures in which the gas *spreads out.*

So the following *appears* to be a statistical explanation of why gases spread out: the vast proportion of microstates compatible with a given gas-is-in-the-left-half macrostate have futures in which the gas spreads out. But as it stands, this explanation is defective. The trouble is that exactly analogous reasoning shows the following: The vast proportion of microstates compatible with a given gas-is-in-the-left-half macrostate have histories in which the gas *was more spread out in the past!*

And it is certainly *not* true that gases that are at one time concentrated in the left half of their boxes tend to have recent pasts in which they were dispersed throughout those boxes. A promising way out of this difficulty is to introduce an assumption concerning the initial state of the universe, *viz*., that the universe started in a state of extremely low entropy. For example, one might posit an additional law of nature that has the effect of constraining the initial state of the universe in this way (Penrose 1993). Given such a law, the following modified explanation is available: almost all of the microstates compatible with a given gas-is-in-the-left-half macrostate *and compatible with the low-entropy constraint on the initial state of the universe* have futures in which the gas spreads out. Furthermore, the modified explanation (unlike the original one) does not lead to the incorrect retrodictions about the past history of the gas.

The global picture is this: judging purely by size of regions of phase space, one would expect for the universe to start (and stay) in global equilibrium. But restricting attention to just those regions of phase space compatible with a low-entropy initial condition, one would expect for global entropy to start low and increase over time. And it is reasonable to think that in a world with global increase of entropy, irreversible processes abound.

Entropy in Contemporary Physics

The explanation given above worked under the assumption of classical mechanics. Whether an explanation of the same type is consistent with contemporary physics remains to be seen. It is a reasonable (but as yet unproven) hypothesis that a definition of statistical-mechanical entropy in terms of phase space volume can be given in the context of general relativity (Bekenstein 2001 appraises this hypothesis).

It is expected that a general treatment of relativistic statistical-mechanical entropy would require a quantum-mechanical theory of gravitation (Wald 1998). Such a theory has been notoriously elusive. Nevertheless, the study of gravitational entropy is an active area of research. For example, theorists have offered statistical-mechanical measures of the entropy of black holes, from both the standpoint of quantum field theory (Sorkin 1998) and the standpoint of string theory (Horowitz 1998).

ADAM ELGA

References

Albert, David (2001), *Time and Chance*. Cambridge, MA: Harvard University Press.

Bekenstein, Jacob D. (2001), "The Limits of Information," *Studies in History and Philosophy of Modern Physics* 32: 511–524.

Callender, Craig (2001), "Taking Thermodynamics Too Seriously," *Studies in History and Philosophy of Modern Physics* 32: 539–553.

Goldstein, Sheldon (2001), "Boltzmann's Approach to Statistical Mechanics," in Jean Bricmont, Detlef Durr, Maria C. Galavotti, Giancarlo Ghirardi, Francesco Petruccione, and Nino Zanghi (eds.), *Chance in Physics: Foundations and Perspectives, Lecture Notes in Physics* 574. New York: Springer-Verlag.

Horowitz, Gary T. (1998), "Quantum States of Black Holes," in R. Wald (ed.), *Black Holes and Relativistic Stars*. Chicago: University of Chicago Press.

Lebowitz, Joel L. (1993), "Macroscopic Laws, Microscopic Dynamics, Time's Arrow and Boltzmann's Entropy," *Physica A* 194: 1–27.

Maudlin, Tim (1995), "Review of L. Sklar's Physics and Chance and Philosophy of Physics," *British Journal of the Philosophy of Science* 46: 145–149.

Penrose, Roger (1993), *The Emperor's New Mind*. Oxford: Oxford University Press.

Price, Huw (2002), "Boltzmann's Time Bomb," *British Journal for the Philosophy of Science* 53: 83–119.

Sklar, Lawrence (1993), *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

Sorkin, Rafael D. (1998), "The Statistical Mechanics of Black Hole Thermodynamics," in R. Wald (ed.), *Black Holes and Relativistic Stars*. Chicago: University of Chicago Press.

Wald, Robert M. (1998), "Black Holes and Thermodynamics," in Wald (ed.), *Black Holes and Relativistic Stars*. Chicago: University of Chicago Press.

Uffink, Jos. (2001), "Bluff Your Way in the Second Law of Thermodynamics," *Studies in History and Philosophy of Modern Physics* 32: 305–394.

See also **Classical Mechanics; Reductionism; Thermodynamics; Time**

K

KIN SELECTION

See **Natural Selection**

KINETIC THEORY

Kinetic theory explains the properties and behavior of physical systems on the basis of the hypothesis that they consist of a great number of particles (e.g., molecules or atoms) in motion (Greek: *kinesis* = motion). Its most important application is the kinetic theory of gases, but it can be applied to liquids and collections of subatomic particles as well. This discussion will be restricted to the kinetic theory of gases, as this suffices to clarify the nature of kinetic theory and to highlight its philosophical aspects.

The kinetic theory of gases proceeds from two principles:

1. *An ontological principle*: Gases are composed of freely moving particles subject to the laws of classical mechanics (atomism, mechanism); and
2. *A methodological principle*: The behavior of gases is analyzed not by tracing the trajectory of every individual particle but by applying statistical methods to the collection of particles as a whole (see Classical Mechanics).

The theory is historically and conceptually related to the theories of statistical mechanics and thermodynamics. Statistical mechanics is a generalization of kinetic theory that emerged in the course of the latter's development by Ludwig Boltzmann and J. W. Gibbs. Thermodynamics is a phenomenological theory that accounts for the behavior of gases without a hypothesis about their microscopic constitution (see Thermodynamics).

Therefore, thermodynamics can be regarded as a competitor to kinetic theory, a relation that is a topic of philosophical debate. The kinetic theory of gases has played a significant role in the shaping of modern physics and has been relevant for the philosophy of science in a variety of ways, having had an impact on the development of its ideas. The kinetic theory of gases is regularly employed in philosophical discussions.

The Kinetic Theory of Gases in Historical Perspective

An essential feature of the kinetic theory of gases is the identification of temperature with molecular motion, specifically with the mean kinetic energy of molecules: $T \sim <\frac{1}{2}mv^2>$, where T is the temperature and m and v are the mass and velocity, respectively, of an individual molecule. In the second half of the nineteenth century, this kinetic view of heat replaced alternatives such as the caloric theory, which assumed that heat was a substance ("caloric"), and the wave theory of heat, which took heat to be vibrations of an ether. Before 1850, some kinetic theories of gases were advanced, which attempted to explain Boyle's law that at constant temperature the product of pressure and volume of the gas is constant (known since 1662). These theories did not have much impact because Newton had already explained Boyle's law by means of a static molecular theory of gases. The earliest kinetic theory was Daniel Bernoulli's ([1738] 1965) in which pressure is proportional to molecular velocity squared. It did not yet have a temperature scale, however. Bernoulli derived the formula (still accepted in modern kinetic theory)

$$PV = \frac{1}{3}N<mv^2>,$$

where P is the pressure, V is the volume, N is the number of molecules, and m and v are the mass and velocity, respectively, of an individual molecule. Kinetic theories were proposed by John Herapath in 1820 and J. J. Waterston in 1845 in which temperature was related to molecular velocity: Herapath (wrongly) supposed that T was proportional to v; Waterston took T to be proportional to v^2. Both theories were ignored (for historical references, see Brush 1965–1972 and 1976).

Around 1850, the scientific scene turned favorable for kinetic theories: Joule and others established the law of conservation of energy (or convertibility of heat and work), which gave essential support to the kinetic view of heat. This paved the way for the important kinetic theory of Rudolf Clausius ([1857] 1965). Like earlier theories, Clausius's statistical hypotheses were of a simple kind: calculations were based on a random variation in the *direction* of particle velocities; their *magnitude* was represented by one average value. In response to an objection by Buys Ballot, Clausius introduced the "mean free path" of molecules and calculated that this path is so small that gases mix quite slowly despite the high velocities of individual molecules.

In the 1860s, Clausius's theory was further developed by James Clerk Maxwell, who at first regarded his work on gas theory as an "exercise in statistics"; he did not (yet) believe in the atomistic view of matter. Maxwell's theory explained many properties of real gases; most important were its predictions regarding transport phenomena (heat conduction, diffusion, viscosity). On the basis of Maxwell's approach, Boltzmann later developed a general transport equation (the Boltzmann equation), which is still employed today. Maxwell refined Clausius's statistical analysis: Instead of using merely one average velocity, he introduced a statistical distribution function $f(v)$ for molecular velocities (a Gaussian curve), where v is the velocity of individual molecules. Boltzmann generalized Maxwell's distribution law for situations in which external forces are present (leading to what is now called the *Maxwell–Boltzmann distribution law*):

$$f(v) = N\sqrt{\frac{2m^3}{\pi k^3 T^3}}v^2 e^{-(\frac{1}{2}mv^2 + V[x])/kT}$$

where m and v are the mass and velocity, respectively, of an individual molecule; k is Boltzmann's constant; and $V[x]$ is a potential due to an external force depending on the position x. This law applies only to equilibrium situations; in order to explain the tendency toward equilibrium, Boltzmann advanced the H-theorem (see below).

A fundamental element of kinetic theory is the equipartition theorem, which states that every degree of freedom of the system takes up an equal part of the total kinetic energy. The theorem had an important anomalous consequence: Its prediction for the specific heat ratio gases is at odds with the experimentally obtained value for many ordinary gases (e.g., oxygen, nitrogen). This "specific heat anomaly" was discovered by Maxwell (1875), who called it "the greatest difficulty which the molecular theory has yet encountered." Boltzmann's proposed solution—the "dumbbell model" of diatomic molecules—was controversial at the time because it disregarded spectral evidence for internal atomic structure (see de Regt 1996). In his famous 1900 lecture, William Thomson (Lord Kelvin) (1904) labeled the equipartition problem as one of two "nineteenth-century clouds over the dynamical theory of heat and light." Today Boltzmann's model is accepted as an idealization: nineteenth-century objections to the model have dissolved since quantum mechanics has separated internal atomic structure (spectra) from mechanical degrees of freedom (see Quantum Mechanics).

In the early twentieth century, kinetic theory was incorporated into statistical mechanics, due to the work of J. W. Gibbs. Kinetic theory itself did not witness serious changes anymore; the most important twentieth-century contributions consisted in methods for solving the Boltzmann equation, notably by S. Chapman and D. Enskog (Brush 1976, chap. 12). Kinetic theory was fruitfully applied to other topics, such as radiation transfer, ionization, chemical reactions, evaporating liquids, and neutron transport. While kinetic theory remains a paradigm nineteenth-century theory, it has had a profound influence on twentieth-century physics, especially on the genesis and development of quantum theory: Boltzmann's work was a key element of Planck's solution of the problem of black-body radiation, marking the beginning of quantum theory in 1900.

Kinetic Theory: Atomism and Scientific Realism

Kinetic theory is based on an atomistic ontology. The unobservability of atoms led to philosophical disputes over their existence, and kinetic theory played a pivotal role in debates between scientific realists and their opponents.

Around 1850, atoms were regarded as fictional entities. However, due to the impressive successes of the kinetic theory in later years, more and more scientists took a realist stance toward the theory. But while the existence of atoms was gradually accepted by the scientific community, their structure was still open to debate and speculation. In the second half of the nineteenth century, various models of atomic structure were proposed. Maxwell treated atoms as hard elastic spheres and later as centers of force. By contrast, Kelvin's vortex theory represented atoms as spinning rings of a homogeneous, frictionless, incompressible fluid. Meanwhile, estimations of molecular sizes were made, first by J. Loschmidt (1865), who also calculated the number of molecules per volume unit (today this number, which is the same for every gas at standard temperature and pressure, is known as Avogadro's number). Subsequently, J. D. van der Waals replaced the ideal gas law ($PV = nRT$, where P is the pressure, V the volume, n the number of moles of the gas, R the ideal gas constant, and T the temperature) by an equation of state for real gases, containing correction terms depending on intermolecular forces and the size of the molecules:

$$\left(P + \frac{an^2}{V^2}\right)(V - nb) = nRT,$$

where a and b are constants depending on the type of molecule, n is the number of moles, and R is the molar gas constant.

These scientific successes firmly but only temporarily established the atomistic worldview and an accompanying realistic view of scientific theories. After 1880, kinetic theory lost momentum, and in the final decade of the nineteenth century the theory was strongly attacked by anti-atomists who based their objections on a positivist philosophy of science. A notable early example is J. B. Stallo's ([1884] 1960) criticism of kinetic theorists for having "faith in spooks" and for wasting "their efforts upon a theory so manifestly repugnant to all scientific sobriety" (151). The most important anti-atomist was Ernest Mach, whose radical empiricist movement (to which the young Max Planck adhered) influenced twentieth-century philosophy of science, particularly the logical positivism of the Vienna Circle (see Mach, Ernest; Vienna Circle).

The tide turned once again in the early twentieth century. In the course of his work on black-body theory, Planck was converted to atomism and became a staunch opponent of Mach. In 1905, Einstein published an explanation of Brownian motion—the observable irregular motion of small particles suspended in fluids, discovered by Robert Brown in 1828—on the basis of kinetic theory: Brownian motion results from the impact of surrounding molecules on the particle, and Einstein ([1905] 1965) derived a prediction of the mean displacement of Brownian particles. This prediction was successfully tested by Jean Perrin (1913), who subsequently made it into a strong case for the reality of atoms in his book *Les Atomes.* As such, Einstein's explanation (independently developed by Smoluchowski) was the final vindication of kinetic theory and atomism.

Today, the existence of atoms is uncontested among scientists, but the case of atomism and kinetic theory is still used in philosophical debates about the status of unobservable entities, particularly (not surprisingly) by realists. For example, Salmon (1984, 213–227) returns to Perrin's argument in order to defend scientific realism; he argues that from the fact that there are many independent methods of determining Avogadro's number, which all arrive at the same result, one must conclude that this number describes something real and that molecules thereby exist. (For an illuminating analysis of the debate between realists and instrumentalists from the perspective of the development of kinetic theory, see Gardner 1979.)

Kinetic Theory: Explanation and Reduction

Kinetic theory is often cited as a paradigmatic example in philosophical discussions about explanation and reduction. In the context of his deductive-nomological model of explanation, Carl Hempel cited kinetic-theoretical explanations of phenomenological gas laws (such as Boyle's law) as exemplary cases of "theoretical explanation" (see Explanation; Hempel, Carl Gustav). Crucial in Hempelian theoretical explanations (where the explanans refers to theoretical, that is, unobservable, entities) are the so-called bridge principles connecting the theoretical and the observational level. In the case of kinetic theory, the bridge principles relate macroscopic features such as temperature and diffusion rate with microscopic properties such as velocity and kinetic energy of the gas molecules (Hempel 1966, 73).

Contemporary philosophers who reject Hempel's model of explanation feel nonetheless obliged to give alternative interpretations of how kinetic theory explains gaseous behavior. Apparently, the fact that kinetic theory provides scientific explanations is undisputed: Any respectable theory of explanation should be able to account for the explanatory power of kinetic theory. Thus, the presently influential unificationist conception of explanation argues that kinetic theory provides explanations because it unifies many (sometimes seemingly unrelated) facts about nature: It accounts not only for Boyle's law, but for many other phenomenological gas laws as well, and also relates gaseous behavior to other natural phenomena governed by the laws of mechanics (Friedman 1974, 14–15). Alternatively, the causal conception of explanation claims that it is the causal-mechanical features of the kinetic theory that do the explanatory work (Salmon 1984, 227–228).

A related philosophical issue is that of intertheoretic reduction. According to Nagel (1961, 342), the relation between kinetic theory and the phenomenological theory of thermodynamics is "a classic and generally familiar instance of such a reduction" (see 338–345 for an account). *Pace* Nagel, the reduction of thermodynamics to kinetic theory is not completely unproblematic: kinetic theory appears to be unable to account for the tendency toward equilibrium described by the second law of thermodynamics. This inconsistency was already observed by Maxwell (1872), in his famous thought experiment known as Maxwell's Demon: Maxwell imagined a microscopic but "very observant and neat-fingered being" manipulating molecules in such a way as to make heat flow from a cold to a hot body, thereby contradicting the second law of thermodynamics (see Irreversibility). Boltzmann's ([1872] 1966) H-theorem was intended as a microphysical analogue to the second law. Boltzmann defined a function H on $f(v^2)$ and proved that H always decreases. As such, H can be regarded as a microphysical counterpart of entropy S. Boltzmann's proof required an extra statistical hypothesis, the *Stosszahlansatz* (molecular chaos): there is no statistical correlation between colliding molecules before and after the collisions. However, if the system behaves according to deterministic Newtonian mechanics (as kinetic theory presupposes), the *Stosszahlansatz* cannot be absolutely true.

This incompatibility between mechanical laws (read: kinetic theory) and thermodynamics was made explicit in the reversibility objection (Thomson [1874] 1966; also known as Loschmidt's [1876] *Umkehreinwand*): If one considers a process and reverses the velocity of every molecule, the resulting process will be physically possible as well. This contradicts the second law, and the experience that irreversible processes exist in nature. Boltzmann responded to the objection with his famous equation $S = k.\log W$, relating the entropy S of a macroscopic state to the number (W) of possible microscopic states corresponding with the macroscopic state in question (in other words, to its relative probability). This equation, which lies at the basis of statistical mechanics, implies that entropy decrease is not impossible but only very improbable: Because there are many more microstates corresponding with a macrostate of high entropy (disorder), the probability that the system develops into a state of higher entropy is much greater than vice versa. In contrast to Boltzmann's response, however, some authors take an antireductionist approach by claiming that the second law has absolute (ontological) validity and that attempts at reducing thermodynamics to mechanics are misguided (e.g., Prigogine 1980). A more detailed overview of philosophical issues related to kinetic theory can be found in Sklar (1993).

Henk W. de Regt

References

Bernoulli, D. ([1738] 1965), "On the Properties and Motions of Elastic Fluids, Especially Air," in Stephen G. Brush (ed.), *Kinetic Theory*, vol. 1. Oxford: Pergamon Press, 57–65.

Boltzmann, L. ([1872] 1966), "Further Studies on the Thermal Equilibrium of Gas Molecules," in Stephen G. Brush (ed.), *Kinetic Theory*, vol. 2. Oxford: Pergamon Press, 88–175.

Brush, Stephen G. (1976), *The Kind of Motion We Call Heat*. Amsterdam: North-Holland.

——— (ed.) (1965–72), *Kinetic Theory*, 3 vols. Oxford: Pergamon Press.

Clausius, R. ([1857] 1965), "The Nature of the Motion Which We Call Heat," in Stephen G. Brush (ed.), *Kinetic Theory*, vol. 1. Oxford: Pergamon Press, 111–134.

de Regt, Henk W. (1996), "Philosophy and the Kinetic Theory of Gases," *British Journal for the Philosophy of Science* 47: 31–62.

Einstein, A. ([1905] 1956), *Investigations on the Theory of Brownian Movement*. New York: Dover.

Friedman, Michael (1974), "Explanation and Scientific Understanding," *Journal of Philosophy* 71: 5–19.

Gardner, Michael (1979), "Realism and Instrumentalism in 19th-Century Atomism," *Philosophy of Science* 46: 1–34.

Hempel, Carl G. (1966), *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall.

Loschmidt, J. (1865), "Zur Grösse der Luftmolecüle," *Sitsungsberichte, K. Akademie der Wissenschaften in Wien, Math.-Naturwiss, Kl.* 52: 395–413.

——— (1876), "Über den Zustand des Wärmegleichgewichtes eines Systemes von Körpern mit Rucksicht auf die Schwerkraft," *Sitsungsberichte, K. Akademie der Wissenschaften in Wien, Math.-Naturwiss, Kl.* 73: 128–142.

Maxwell, J. C. (1872), *Theory of Heat*. New York: Appleton.

——— (1875), "On the Dynamical Evidence of the Molecular Constitution of Bodies," *Nature* 11: 357–359, 374–377.

Nagel, Ernest (1961), *The Structure of Science*. London: Routledge and Kegan Paul.

Perrin, Jean (1913), *Les Atomes*. Paris: Alcan. English translation: *Atoms*. New York: Van Nostrand, 1923.

Prigogine, Ilya (1980), *From Being to Becoming*. San Fransisco: Freeman.

Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Sklar, Lawrence (1993), *Physics and Chance*. Cambridge: Cambridge University Press.

Stallo, John B. ([1884] 1960), *The Concepts and Theories of Modern Physics*. Cambridge, MA: Belknap Press.

Thomson, William (Lord Kelvin) ([1874] 1966), "The Kinetic Theory of the Dissipation of Energy," in Stephen G. Brush (ed.), *Kinetic Theory*, vol. 2. Oxford: Pergamon Press, 176–187.

——— (1904), "Nineteenth-Century Clouds over the Dynamical Theory of Heat and Light," in *Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light*. London: C. J. Clay and Sons, 486–527.

See also **Classical Mechanics; Explanation; Irreversibility; Mechanism; Quantum Mechanics; Probability; Realism; Reductionism**

THOMAS KUHN

(18 July 1922–17 June 1996)

Thomas S. Kuhn was the most widely read, and most influential, philosopher and historian of science of the twentieth century. *The Structure of Scientific Revolutions*, first published in 1962, challenged then-dominant philosophical views of science regarding progress, rationality, observation, theories, and language. The book has been continuously in print for forty years; it has been translated into more than twenty languages, and the various editions have sold over a million copies. Unlike all other books in the history or philosophy of science, *Structure* was, and is, still widely read outside of the philosophical community.

Kuhn was also the author of *The Copernican Revolution*, and in 1978, *Black-Body Theory and the Quantum Discontinuity, 1894–1912*, as well as numerous essays, most of which were reprinted in two collections, *The Essential Tension* (Kuhn

1977) and *The Road since Structure* (Kuhn 2000). Although the monographs each made important contributions, respectively, to contemporary understanding of the Copernican revolution and of the early stages of the quantum revolution, none of the other work has attracted nearly as much attention as *Structure*, especially within philosophy of science. If Kuhn had written only these two monographs and the articles, he would have merited a minor footnote as a historian of science in the twentieth century.

Although Kuhn was the most influential philosopher of science of the twentieth century, his formal training was in physics, and his next career turn was to the history of science. He was trained as a physicist, receiving a B.S. in 1943, an M.A. in 1946, and a Ph.D. in 1949 from Harvard in that discipline. Moreover, his first teaching position, at University of California, Berkeley, from 1956 to 1964, and his second, at Princeton University (1964–1979), were in history departments and in the history and philosophy of science program at Princeton, but not in philosophy. Only when he moved to Massachusetts Institute of Technology in 1979 did he become a member of a philosophy department (though it was the *Linguistics and* Philosophy Department). He was elected president of the Philosophy of Science Association in 1989, after having been a member of that association for only a few years.

Revolutions and Two Kinds of Science

The main thesis of *Structure* is that the development of the natural sciences and their subfields proceed through alternations of two kinds of scientific development: *normal* and *revolutionary* science (see Scientific Revolutions). A period of normal science produces cumulative progress in understanding of the domain of that field and involves the application and refinement of generally accepted theories to the unresolved questions in a domain according to an agreed understanding both of what constitutes a reasonable question and on the criteria used to adjudicate answers. Revolutionary change involves rejection of a significant portion of the theories, methods, and criteria for problem solution, and their replacement by new ideas. In revolutionary change at least some of the previously "solved" problems are rejected or reopened. One of the most controversial claims of the book was that there was incommensurability between a revolutionary theory and the one it supplanted (see Incommensurability).

The book received both widespread praise and condemnation. In addition to the claims about "revolutions," which were very controversial, critics argued that Kuhn's account of science made it neither rational nor objective. Kuhn regarded this as a misinterpretation of his views. In addition to the claim about incommensurability, he claimed that when a scientific revolution occurs, "the world changes," and both of these claims provoked philosophical outrage in many critics (see Shapere 1964 for an example). An elaboration of the basic views of the book is presented first; discussion of controversies and consequences follows. For an extensive discussion of the meanings and history of the phrase 'scientific revolution' (see Scientific Revolutions).

Understanding Kuhn's account is complicated by the fact that throughout his career he was in the process of refining his positions to clarify them, to meet objections and to eliminate misunderstandings. *Structure* was only published in its actual form because Kuhn had agreed to write an entry for the Encyclopedia of Unified Science series and the editors pressed him to produce a manuscript (see Unity of Science Movement). His personal preference would have been to continue to develop the ideas and to relate them in more detail to the existing tradition in philosophy of science, with which he was then acquainting himself. In the Preface, he apologizes for leaving out many topics and more precise references because of lack of space. After the somewhat negative reception of the book, he suggested a major terminological change in the second edition (1970), which he did not incorporate in the text generally, but only in a postscript, which has been largely ignored by readers and critics. He continued to work throughout his life on a clearer and more definitive formulation of his views and he died in 1996 without having completed that project. Thus any evaluation must be of work still in progress.

Structure

In the first edition of *Structure*, Kuhn defined the two kinds of scientific development in terms of *paradigms.* Normal science involves the articulation and refinement of a paradigm that is shared by the relevant scientific community; in revolutionary scientific change, one paradigm is rejected and another takes its place. One reason for the widespread influence of the book outside of the community of philosophers and historians is that the conception of a group or community guided by a *paradigm* seemed to have explanatory value in many settings. This use of the term has become firmly entrenched as a standard expression in English and appears in cartoons and business management courses, although most of its contemporary users have no notion of its source.

However useful the term ‘paradigm’ has proven to be in the general culture, it was the cause of considerable criticism in the reception of the book because critical readers perceived that he was using the term very variously and loosely. One critic presented a taxonomy of twenty-two distinguishable senses of the term in *Structure* (Masterman 1974). Kuhn disagreed with the precise count, but was sufficiently persuaded by Masterman’s critique and those of many other critics that clarification was required. Many of the criticisms were aired at two important conferences that focused heavily on his work, in London in 1965 and Champaign, Illinois, in 1969. The proceedings of these were eventually published as *Criticism and the Growth of Knowledge* (Lakatos and Musgrave 1974) and *The Structure of Scientific Theories* (Suppe 1977). Kuhn was also conducting graduate seminars at Princeton that were attended by philosophers of science on the faculty, as well as historians and graduate students from both disciplines. As a result of these influences and further reflection, in the postscript to the second edition of *Structure* in 1970, he expressed a desire to replace the term ‘paradigm’ with two new terms, ‘disciplinary matrix’ and ‘exemplar,’ which he believed expressed the two main distinct uses he had made of “paradigm.” Similar qualifications of the view in *Structure* are expressed in his contributions to those two conferences.

The six elements that Kuhn intends to include in a *disciplinary matrix* are:

(i) equations or other symbolic representations,
(ii) instruments,
(iii) standards of accuracy and experimental repeatability,
(iv) metaphysical assumptions,
(v) the domain of inquiry, and
(vi) exemplars.

The *domain of inquiry* includes the problems that workers in the field regard as relevant but unsolved. An *exemplar* is, as we have seen, the second meaning of ‘paradigm,’ and Kuhn emphasizes that these are *concrete* examples of problem solutions. One of the most crucial points in his emphasis on exemplars is that they give guidance to future research by example; these examples are implicitly constrained by rules and are instances of a method, but neither the rules nor the method is explicit in them. A scientific field or specialty is given its coherence partly by the shared examples, but it is given its diversity of approaches by the possibility of researchers interpreting those examples somewhat differently from one another. Researchers can all agree that they want to do for their field what Newton did for his, but they may disagree fairly radically about what that was, and therefore on what they intend to achieve. Some of the confusion in interpreting *Structure* was due to the fact that one sense of paradigm, that is, exemplars, are an element of the other sense (disciplinary matrices).

Disciplinary Matrices and “The Scientific Revolution”

In the Ptolemaic/Aristotelian scheme that dominated scientific thought in the Western world for almost two millennia, one of the fundamental metaphysical assumptions is that there are unbridgeable differences between terrestrial and celestial phenomena, while the Copernican/Newtonian view assumes the uniformity of laws throughout the universe. The former emphasizes qualitative explanations; the latter, quantitative predictions. Famously, the telescope, particularly in the hands of Galileo, was a critical instrument in the arguments against the static Ptolemaic/Aristotelian view of the heavens. While Kuhn was still a graduate student, James Conant, a chemist who was then president of Harvard, asked him to assist in preparing a historically oriented physics course for non-science majors. In the process of preparing for this course, Kuhn, who had read little or no history of science previously, spent an extensive period of time reading Aristotle.

He describes how as he read Aristotle he discovered that Aristotle had known almost no mechanics, if one understands mechanics as the system discovered by Galileo, Newton, and others. This baffled Kuhn because Aristotle’s contribution to logic remained of central importance at least until the twentieth century, and Kuhn believed that Aristotle’s observations in biology provided models that were instrumental to the emergence of the modern biological tradition. If Aristotle had been both a keen observer and the epitome of reasoning, how could he be so mistaken (Kuhn 2000, 16)?

His conceptual difficulty led Kuhn to reflect that perhaps Aristotle’s (translated) words did not mean quite the same to the modern reader as they had to Aristotle. This thought, together with continued concentrated immersion in the texts, led to an abrupt revelation:

> Suddenly the fragments in my head sorted themselves out in a new way, and fell into place together. My jaw dropped, for all at once Aristotle seemed a very good

> physicist indeed, but of a sort I'd never dreamed possible. Now I could understand why he had said what he'd said, and what his authority had been. Statements that had previously seemed egregious mistakes, now seemed at worst near misses within a powerful and generally successful tradition. (Kuhn 2000, 16)

This experience initiated the intellectual development that led to *Structure* and his position on revolutionary change of worlds and worldviews, a position that raised problems that he would continue to struggle with until his death in 1996. The careful reader can detect this theme of sudden revision already in Kuhn's first book, *The Copernican Revolution*, published in 1956, although it is probably only with hindsight that one can see the importance of the idea. Writing that monograph cemented many of the major themes of Kuhn's later work in its detailed description of the complex transformation from the world as described by Ptolemaic astronomy and Aristotelian physics (or perhaps medieval neo-Aristotelian physics) to the worldview that developed through Copernicus, Kepler, and Galileo, to culminate in Newton.

In assessing this sudden transformation and the significance of the process of writing this book on Kuhn's subsequent views, it is essential to recall the unique character of the Copernican-Newtonian revolution. Indeed, for many historians and philosophers, it is called *the* Scientific Revolution. Before this revolution, humans saw themselves as situated on a motionless Earth in the center of a relatively small finite universe. Terrestrial substances were divided into four kinds, and the motions of objects depended on the substances composing them; each kind had a natural motion defined in terms of the center of the universe located at the center of the Earth. Celestial substances were a different matter—or rather were *not* matter—and followed circular paths.

By the end of the Scientific Revolution, humans were on an Earth that was not only rotating at 1,000 miles per hour but also one that was traversing an orbital path around the sun at an even greater velocity. They were not at the center of the solar system, but on the third planet from the sun. Nor is the sun at the center of the universe, for the universe is infinite and there is no center at which to be located. Terrestrial and celestial objects were now subject to the same governing laws, and those laws were abstract, quantitative, and mathematical rather than qualitative and teleological.

Notice that two abrupt changes are involved here. One is the change in the scientific worldview—the Ptolemaic/Aristotelian view that had evolved little over almost two millennia was suddenly replaced by the Copernican/Newtonian view. The second is that Kuhn's understanding of Aristotle's worldview and the relation between Aristotle's views and post-Newtonian views underwent an instantaneous change when his jaw dropped. Until much later (Kuhn 2000), Kuhn did not distinguish these two kinds of changes, one of which is a personal psychological transformation, and the second a social and epistemological change in a community. Neither the scale nor the processes of these two kinds of change are identical, and some of the lack of clarity of his earlier views is due to failure to make this distinction.

Progress

The question of scientific "progress" is a complicated one in Kuhn's thought. He clearly believed that there is directionality to scientific change, that, for example, scientific disciplines frequently split into subdisciplines that become fields in their own right, and that this process is never reversed. And at the most general level, as mentioned above, he thought that one can distinguish pre-paradigm from paradigm-driven fields. But directionality does not mean that it is movement toward some ultimate end rather than simply solutions to current problems.

As indicated above, many of Kuhn's later views are already discernible in *The Copernican Revolution*, but there are other points at which his ideas had clearly developed in the six-year period between that work and *Structure.* In the conclusion to the former book, he discusses the fact that progress in scientific concepts is not cumulative. "But though the achievements of Copernicus and Newton are permanent, the concepts that made those achievements possible are not" (Kuhn 1957, 264–265). This is clearly consistent with his more elaborated view in *Structure* that science does not make cumulative progress with respect to the underlying structure of the world, and is closely connected with his controversial views about how the "world changes" during evolutionary periods.

However, in *The Copernican Revolution* he still holds that the list of solved problems is cumulative across even the Scientific Revolution. "Only the list of explicable phenomena grows; there is no similar cumulative progress for the explanations themselves" (265). By the time he wrote *Structure*, Kuhn had ceased to think that even the list of explained phenomena was cumulative. The beginning of the evidence for this was already in *The Copernican Revolution*, but he had not seen it as such.

For example, in the Ptolemaic/Aristotelian view of the world, since the Earth was (almost) at the center of the celestial sphere of fixed stars, there was a trivial explanation of why there was no apparent parallax in the position of the fixed stars over a period of six months. On the other hand, given the Copernican view of the Earth as orbiting around the sun, one would expect to see the stars appear at slightly different angles when observed at intervals of six months because the Earth is at opposite sides of its orbit. Since no parallax was observed, this was taken by many as evidence for the Ptolemaic/Aristotelian view. The Copernican explanation of the lack of observed parallax was to infer that the distance to the fixed stars was so great that the angle of parallax was less than the limit of observation (Kuhn 1957, 159).

In this instance, what was an unproblematic observation in the Ptolemaic system was "explained" by an assumption that was seen as ad hoc by traditional astronomers when propounded. The detection of parallax remained an issue for Copernican and successive astronomical views even after the invention of the telescope, and it was not until 1838 that the first measurements demonstrating parallax were made (163).

One of the exemplars of the Copernican/Newtonian view was the pendulum. In this case the universal laws of gravitation, force, and acceleration combine to give a derivation of a precise quantitative law stating that the period of a pendulum depends only on its length. The place of the pendulum in the development of Copernican/Newtonian physics is very important. The properties of a simple ideal pendulum are easy to establish, but more complex approximations to actual physical instantiations pose important puzzles for normal science. For the Ptolemaic/Aristotelian view, a heavy object suspended from a string or chain is of no theoretical interest because it is an example of constrained motion—the string or chain prevents the object from pursuing its natural motion toward the center of the Earth and universe. The example of the pendulum was of central importance for some of Kuhn's most controversial views on world change during revolutions, a topic that will be discussed later.

Normal Science

According to Kuhn, normal science consists of periods of cumulative progress in which scientists apply generally accepted theories to the unresolved questions in a domain according to shared assumptions about what constitutes the important problems and what would count as a solution. He characterizes "normal science" as a very sophisticated form of puzzle-solving that can require great ingenuity but occurs within a stable framework of tradition.

Kuhn's characterization of "normal science" was criticized from at least two directions. Popper argued that if Kuhn were correct, then normal science was in fact not science at all, because scientists under those conditions were presupposing the cognitive elements of the disciplinary matrix and not testing them (Lakatos and Musgrave, 1974) (see Popper, Karl Raimund). According to Popper, the Kuhnian characterization of science requires that the laws and theories under consideration be falsifiable and that experiments be attempts at falsification. Kuhn did not accept this criticism—his view was that the presuppositions were necessary in order to make progress. Continual reexamination of what is taken as basic knowledge would impede the process of extending knowledge. He also held that it was only through the vigorous pursuit of normal science that further revolutions could be achieved.

A second criticism was that Kuhn's account made so-called normal science, which is quantitatively the large majority of scientific activity, uninteresting and routine. Although his use of the term "puzzle solving" to describe normal-scientific activity may have made it sound more routine, Kuhn certainly did not think that normal science was uninteresting and routine. The phrase "puzzle solving" may have been unfortunate, because it tends to suggest crossword puzzles or jigsaw puzzles, challenges that have been created by humans with explicit rules and a solution that is known in advance before the puzzle solvers enter the activity.

In contrast, the "puzzles" of normal science are posed by scientists, but the answers are not known by anyone in advance, and the rules for solution are given at best implicitly by previous exemplars. Both features—the difficulty of solution and the bridge to the next revolution—can be illustrated by many examples, but the refinement of the Newtonian theory of the solar system is a particularly clear one.

The textbook accounts say that Newton derived Kepler's laws from his own laws of motion, but this is a derivation in the sense of physicists, not of philosophers or logicians. Kepler's first law states that the planets move in elliptical orbits with the sun at one of the foci of the ellipse. However, the derivation of an elliptical orbit for one body revolving around a second body holds only when no other forces are acting on these bodies than the mutual force proportional to the inverse square of their distance apart. Thus, this derivation is possible only if the effects of all the remaining planets on the orbit are ignored, and it is clear from Newton's law of gravitation that there is a non-null effect. So the "derivation" involves a deliberate simplification, though one that was unproblematic at the time of Newton.

However, as telescopic observation improved, the discrepancy from the Keplerian model was observationally recorded, and one of the puzzles of the astronomical tradition became to provide a more exact mathematical analysis that did not oversimplify as much. For example, by 1770 it was noted that there were deviations in the motions of Jupiter and Saturn from predictions. Laplace established in 1775 that these deviations were periodic (with a period of 929 years!) and showed that the deviations followed from a mathematical analysis that included the gravitational attraction between those two planets. There is a crucial mathematical fact that intrudes here, *viz.,* that the problem of solving the differential equations for two bodies acting under the influence of gravity has a general analytic solution, but in general there is no closed analytic solution for equations involving three or more bodies. Thus significant new mathematical tools were required to solve the puzzle. And the puzzle led to the major mathematical discovery of the mathematical fact just cited, so the process influenced the development of not only astronomy but mathematics as well.

The better mathematical approximations fit well with observation until about the 1840s, when it became clear that the predictions for Uranus did not fit the data. The possibility of an undiscovered planet was one potential explanation, though others were also offered, including the possible failure of the inverse square law for gravitation. However, the standards of astronomy at that post-Keplerian point would not permit as a "solution" the mere postulation of such a planet, but required calculations to determine the location of the hypothetical planet, and observations of the planet and its positions.

When the relevant calculations were made and the appropriate portion of the sky observed, astronomers saw Neptune. But not for the first time! Once the existence of Neptune was known and its orbit calculated backward, astronomers learned that on several occasions the planet had been seen and noted, but that the observers (including Galileo) did not discern that it was a planet. This episode provided (though it was hardly necessary) more confirmation of the universal applicability and validity of the Newtonian equations.

A similar situation arose late in the nineteenth century with respect to the planet Mercury, whose orbit likewise did not fit predictions according to the standards of the time. Astronomers hypothesized another inner planet, calculated its orbit and mass, and even named it: Vulcan. However, nature was not so cooperative in this instance and Vulcan was never observed. The anomaly of Mercury's orbit later became one of the important phenomena predicted correctly by general relativity theory and was a crucial part of the overthrow of Newtonian theory. Rigorous pursuit of increasingly precise predictions within the normal-science tradition of Newtonian astronomy provided the data that gave observational leverage to the overthrow of that tradition.

Rationality

Perhaps the most common and outraged criticism of Kuhn's work was that it denies the rationality of science, because the acceptance or rejection of a new theory depends not only on "scientific" factors but also on social factors. Kuhn's response was indignation and perplexity. The first sentence of *Structure* promises (or warns) that "History, if viewed as a repository for more than anecdote or chronology, could produce a decisive transformation in the image of science by which we are now possessed." One important respect in which Kuhn wanted to correct the image of science was in the general understanding of what constitutes its rationality.

The generally accepted philosophical view of scientific rationality in the 1950s was that science rested on a basis of *observation statements*—these statements were thought to be neutral with respect to the various theories to be compared and to be unproblematic with respect to verification (see Logical Empiricism; Verifiability). While there were disagreements about the character of observation statements (whether they were subjective reports of the agent's perceptions or reports of external states that were intersubjectively agreed upon) and about whether their verification was absolute or not, the assumption of a basis of observation sentences was shared by Carnap, Hempel, and almost all others working at the time (See also Carnap, Rudolf; Hempel, Carl Gustav; Observation; Physicalism; Protocol Sentences). So the first element of scientific rationality was the observational base.

The second component of scientific rationality consisted of the method of establishing verification, confirmation or falsification of theories on the basis of observation sentences (see Confirmation Theory; Observation). Significant controversies raged among those who shared this framework over such issues as whether the basic concept should be probability (see Popper, Karl Raimund; Reichenbach, Hans), confirmation, or refutation, but all shared the overarching program of articulating exactly how the logical relation between theory and observation was to be analyzed. The guiding motivation for this program was the success that had supposedly been obtained in the foundations of mathematics. Mathematical logic had proved to be an enormously successful tool for representing mathematical statements and for analyzing the proof relations among the statements. The second stage in securing the foundations of science was to be a comparable analysis of scientific reasoning. In particular since the surprising success of mathematical logic consisted largely in showing how semantic conceptions, such as entailment, could be rendered into equivalent syntactical forms, such as derivability in a formal system, the goal was to provide a syntactic characterization of the process of theory evaluation.

Summarizing, the two principles are that the basis of evaluating scientific theories was a shared collection of observation sentences and that evaluation is to be done according to a single algorithm, which was to be proved optimal. Given these conditions, it follows that if two scientists disagree about theory choice, then either

- they are using different evidential bases; or
- (at least) one of them is not using the proper algorithm for theory evaluation.

Thus, if Carnap's original program, to prove that there is a unique correct quantitative confirmation function, could have been carried out, any scientific controversy must result from ignorance or inductive error; scientific controversies are an inefficiency in the progress of science. Furthermore, the scientific community, or at least the ideal one in which all scientists follow the algorithmic ideal, has no discernible function except to make the process of scientific development move faster by having more hands to make the tasks lighter. In the ideal scientific community, in this view, there would be no disagreements about theory evaluation. If Carnap's program does not succeed, then one must make a reevaluation of the nature of scientific controversies (see Carnap, Rudolf; Inductive Logic).

One consequence of Kuhn's work and the subsequent discussion has been to encourage alternative approaches to rationality. The Bayesian approach to theory evaluation would undoubtedly have commanded some attention in any event, but interest in this alternative to the traditional approaches was undoubtedly increased by the desire to find ways to accommodate Kuhnian insights in a more formal structure. The view that apparent incommensurability stems from scientists using very different prior probabilities is explored in interesting depth in Salmon (1990) (see Bayesianism).

Rationality and the Social

Since finding a formal analog for inductive logic to the very successful formal deductive foundation of almost all of mathematics was taken as key to demonstrating the rationality of the scientific method, it is not surprising that Kuhn was accused of advocating irrationality when he questioned the possibility of this project. On this point he was very clear in his own mind that the traditional attempts to underwrite the rationality of science misunderstood the history and character of the scientific process. For him, the process of theory choice by the scientific community was the touchstone of rationality, and his goal was to better understand that process. In other words, the scientific process is not rational because the community is embodying an independently specifiable algorithm for theory choice, but the process is rational because it is being carried out by the community.

There are two importantly different ways of understanding this claim. The first is based on an idea from information theory and signal processing. A classical engineering problem was how to improve signal detection given either imperfect detectors or noisy signal channels. Early researchers in the field showed that with imperfect signal detectors one could obtain an arbitrarily good improvement in the accuracy of signal detection by combining a sufficiently large number of the imperfect detectors and taking the result from the majority of the detectors. The application of this analogy is to think of scientists as imperfect detectors of the signals sent by nature, perhaps under experimental questioning, regarding the best of a set of available theories.

This model depends on (at least) two crucial assumptions. One assumption is that the scientists/detectors are more likely to choose the better theory or to detect the correct signal, than the alternative theory/signal. Lest the reader think that "better theory" presupposes too much, note that for these current purposes, one can take *better theory* in a limited way to mean the better theory from an available specified set for the purposes of fitting experimental results, making correct predictions, and cohering with other theories in the near future. This argument can be formulated in a way that is agnostic about metaphysical truth and realism—indeed, both realists and anti-realists can adopt this assumption. If there are more than two theories in contention, then the assumption need only be that the better theory is more likely to be chosen than any individual competitor. One does not need the stronger assumption that the better theory is more likely to be chosen than the disjunction of the competitors. And the difference in probability of the better theory being chosen does not have to be great; even small differences can be leveraged with enough agents.

Secondly, the scientists/detectors must be appropriately independent of one another. Otherwise they will all simply reproduce the same errors. Elaboration of this assumption is a complicated and subtle matter, and only an approximation can be given here. The detectors cannot be statistically independent, for if each is more likely to detect the true theory than not, then their conclusions will be at least weakly correlated. The proper statement of the assumption is that their correlation is entirely due to their propensity to detect the true theory. If the detectors are all made in the same factory, and thus are subject to the same biases and will produce the same result, true or false, then they are not independent in this sense. If the scientists are all trained by the same narrow-minded dissertation advisor and thus will produce the same result, true or false, regardless of the facts of the matter, then they are not independent in this sense.

This emphasis on the independence of judgment suggests an atomistic view of the scientists that is at odds with the emphasis Kuhn places on the weighty role played by the social community of scientists. As Kuhn discussed in *The Essential Tension* (Kuhn, 1977), the process of being trained as a scientist in a particular specialty is simultaneously a process of making the trainee conform to the standard acceptable canons of the discipline while trying to preserve the freedom and independence of judgment that will enable the new specialist to explore and evaluate alternative approaches to anomalies in the discipline. The process of training ranges inclusively from training in specific experimental or mathematical fields to absorption of the cultural tales of the field (Traweek, 1988).

This model of the scientific community is oversimplified in many ways, but it brings out important points. If not exclusively, at least largely, the community must consist of inquirers who are responsive to information about the world—they must be rational inquirers as individuals. But there are also issues of optimality at the level of the community, and one can speak of rationality at the level of the structure of the community.

Kitcher (1993) develops a version of this approach. Since there is no theory-choice algorithm, a point he takes as given, it is important to have a diversity of approaches to theory testing and evaluation. If everyone agreed on and pursued the most plausible direction, then potentially promising theories might well be ignored, often at a considerable loss to the community. One could add to his point that a diversity of cognitive styles and variability in willingness to take risks would also be beneficial. In these matters there remains the difficult task of maintaining the right balance in the essential tension Kuhn noted, the balance between having sufficient communal agreement on enough matters to define a field while leaving room for vigorous debate on others.

An alternative, bolder interpretation is that the rationality is present only at the level of the community. As Longino puts it: "Objectivity, then, is a characteristic of a community's practice of science rather than of an individual's, and the practice of science is understood in a much broader sense than most discussions of the logic of scientific method suggest" (Longino 1990, 74). In her analysis, there are four conditions that a community must satisfy in order to qualify as rationally developing scientific knowledge:

1. There must exist within the community recognized and approved forums or avenues for the criticism of theories, evidence, experiments, assumptions, and inferences.
2. The criticism must be effective in that the community at times changes its belief and practices in response to it. Criticism is not merely tolerated and ignored.
3. As a background for criticism, the community must have publicly recognized and shared standards that provide the criteria for evaluating theories, hypotheses, experiments, data analysis, etcetera. These establish standards for the quality and relevance of criticism.
4. Communities must be characterized by appropriate levels of equality of intellectual authority. This does not require that all members of the community have equal influence, but that any disparities be due to past accomplishments or training and not to political, economic, or other factors not directly

> related to the epistemological task at hand (Longino 1990, 76–81).

Whether communities or individuals are the fundamental basis of rationality in science is still under debate and development, but there is no question that participants such as Kitcher and Longino are continuing a debate opened by *Structure.*

Contexts: Discovery, Justification, Development

Perhaps the most important break with the standard philosophy of science tradition was Kuhn's rejection of the distinction, due to Reichenbach, between the context of discovery and the context of justification. At the very beginning of *Experience and Prediction*, Reichenbach distinguishes the descriptive task of epistemology, *viz.,* "giving a description of knowledge as it really is," from the main evaluative task of epistemology of considering "a logical substitute rather than real processes" (Reichenbach 1938, 5) He concludes that "it will therefore never be a permissible objection to an epistemological construction that actual thinking does not conform to it" (6). Reichenbach (and many other logical empiricists) saw the task of philosophy of science as providing a "rational reconstruction" of scientific processes. These reconstructions substituted abstract logical operations for the actual psychological processes. Probably Kuhn's most important contribution was to bring both history of science and psychology into contact with philosophy of science.

At the time Kuhn was writing *Structure*, almost all of philosophy of science focused on the context of justification, understood as analyzing the relation between a theory axiomatized as a set of sentences in a formal language and evidential sentences from an agreed-on base also represented in a formal language. Kuhn's use of history was frequently seen as dealing with the context of discovery and thus as irrelevant to philosophy of science. Kuhn's insight, which is now generally accepted among philosophers of science, is that theories and evidence undergo significant transformations during the period between the time the theory is first formulated and the time it reaches the final form that is enshrined in textbooks. Most of the process of scientific development occurs between the time of discovery and the time at which a formal justification is developed, and formalization is ill-suited to represent those processes of development.

To develop Kuhn's ideas, it is essential to distinguish a third context, that of *development.* This is the stage in which an embryonic, or perhaps fetal, scientific theory is nurtured and developed so as to analyze its implications. The most famous and dramatic example is the Copernican revolution discussed above. Copernicus' theory of planetary motion described the motion of the planets in terms of circular motion with epicycles on the main cycles, and attributed the movement to a rather mystical force emanating from the sun but pushing the planets around in their orbit. It took more than a half century after Copernicus for Kepler to develop the description of the elliptical orbits, and another half century before Newton provided the dynamics to explain the approximately elliptical orbits in terms of a force attracting the planets to the sun. Although Kuhn sometimes describes scientific revolutions as sudden, he had documented in his first book the century and a half that was required for the Ptolemaic/Newtonian revolution to unfold.

The Scientific Revolution

The issue of the context of discovery is closely related to the character of normal science. As indicated earlier, in Kuhn's view, the adoption of a new disciplinary matrix by a community requires an extended period that looks instantaneous only from a rather distant perspective. For a new disciplinary matrix to be taken seriously by the community, it must solve some of the outstanding problems that have eluded solution with the older disciplinary matrix. But the new disciplinary matrix at that stage has not been developed sufficiently to resolve many of the outstanding problems or to provide new alternative solutions to previously solved problems. Thus much of the process of normal science is the development of new scientific ideas, instruments, mathematics, and experimental methods. The formalized theories on which logical empiricism focused are an end product and do not represent the character of most of the activities of scientists (see Logical Empiricism).

Gestalts and World Changes

One of the significant oversimplifications in *Structure* is that Kuhn uses the same vocabulary to describe the process of change in a scientific community and in an individual—both are characterized as "Gestalt switches," or instantaneous changes of perspective, familiar from such psychological examples as the Necker cube: a two-dimensional configuration of lines that can be seen in either of two three-dimensional orientations. (The relevance of Gestalt psychology to philosophy of science was also argued by Hanson [1958] in his *Patterns of*

Discovery. Although Hanson published these ideas earlier and Kuhn acknowledges Hanson's work in *Structure*, Kuhn had been familiar with the Gestalt examples from his time at Harvard. There is also a question of the scale of revolutions and various perspectives on them. The Copernican–Newtonian revolution took over a century, but this looks relatively abrupt compared with the millennium and a half of domination by the Aristotelian–Ptolemaic disciplinary matrix before the Scientific Revolution, and the two and a half centuries subsequently by the Copernican–Newtonian.

The contrast between the psychological processes by which an individual comes to accept one theory rather than another and the social processes by which the consensus of the scientific community shifts can be illustrated by examples. Galileo quickly adopted the Copernican framework but was part of a long and tragic struggle within (and without) the scientific community. The Gestalt character of how perception can change is illustrated by Kuhn's discussion of Planck's discovery of the constant that bears his name. In his earliest paper, Planck uses a quantity that is equal to the constant, but his conceptualization of it at that time was not that it was a constant, but was simply another quantity that emerged in accounting for the data. It was not until considerably later that Planck conceived of it as a fundamental constant (Kuhn 1978).

Two of the passages that most disturb critics of *Structure* are Kuhn's comments that before Galileo there were no pendula and that when scientists change their disciplinary matrix, as we have seen, the world changes. However, one way of interpreting Kuhn's pendulum comment less dramatically can be derived from recent work on the view that science consists primarily of modeling (Giere 1988 and 1999; Cartwright 1983). If a pendulum is to be understood as described in physics textbooks as a *point mass* suspended from a *massless*, completely *inelastic* string with *no resistance* to the medium in which it moves, then there are no pendula in the physical world, neither before nor after Galileo. However, from Galileo on, the worlds of natural philosophers and physicists include abstract pendula of all various masses and string length. And with these in the mental space of the scientist, the world looks very different because many physical objects approximate the properties of an ideal pendulum so that prediction is possible and fruitful.

Later Work

The main focus of Kuhn's later work was on language, attempting to understand and explicate precisely the nature and causes of incommensurability. This direction began in the 1960s while he was at Princeton, where Quine's ideas about radical translation were a dominant theme of discussions (see Quine, Willard Van). In the postscript to the second edition of *Structure* and subsequent work Kuhn often used the term "translation" and was interested in the parallels between language learning and scientific development. For example, he claimed repeatedly that Aristotelian "physics" cannot be translated into Newtonian terms. A frequent criticism of this claim was that it implied falsely that no one who knew Newtonian mechanics could understand Aristotle, whereas Kuhn himself explicated Aristotle. His response was that the process of learning Aristotle's science was a process of second-language learning, and the ability to speak two languages does not guarantee that what is said in one can be translated without remainder into the other.

In discussing "incommensurability" claims, it is important to bear in mind what Kuhn meant and not overinterpret the claim. He used the term in its historical mathematical sense, in which a mathematician says that the square root of 2 is incommensurable with any rational number. This means that no rational number is a square root of 2. On the other hand, one can approximate the square root of 2 as precisely as you wish by rationals. It is not clear whether Kuhn was committed to this latter consequence, that one could approximate Aristotelian claims as precisely as one wishes by Newtonian statements, but it is at least evident that his terminology did not imply that no comparison could be made between the two theories.

It is unfortunate that Kuhn did not more systematically develop his conception of disciplinary matrix, because it might have led him to recognize that in addition to the semantic incommensurability that most concerned him, other sources and kinds of incommensurability also abound. The Millikan–Ehrenhaft controversy over the electron illustrates how metaphysical and experimental commitments can produce incommensurable analyses of experimental results.

Around 1910, many physicists believed that there was a fundamental particle, now known as the electron, which had the minimal unit of negative electrical charge. In other words, all electrical charges were multiples of the charge of the electron. But many others were still not persuaded of the existence of such quantized particles. Millikan (1911), believing strongly in the existence of the electron, embarked on the research project of determining more precisely the value of the charge.

His method was to produce small oil drops that were ionized by radiation so that the drops were charged. He then observed their behavior in both the presence and the absence of an electrical field.

Meanwhile, Ehrenhaft, who was opposed to ideas of quantization, and perhaps even of atomism, was performing similar experiments to test whether the basic unit of charge existed. Ehrenhaft used small particles of metal rather than oil drops. These had the advantage that they were even smaller than the oil drops and were not susceptible to evaporation, as the oil drops were. Ehrenhaft's experimental procedure produced data that led to calculations of various charges smaller than the unit charge that Millikan reported (Ehrenhaft 1941).

Millikan, being an atomist, believed that Ehrenhaft's particles were so small that they were subject to irregularities in their motion caused by encounters with individual atoms and that Ehrenhaft's results were thus unreliable.

Ehrenhaft rejected Millikan's claims and pointed out that Millikan's process of analyzing data included omission of microscopic observations that produced data inconsistent with his atomic hypothesis. Ehrenhaft also argued that his own data were more reliable because his particles were not subject to any evaporation and could be produced as exact spheres, whereas oil drops were only approximately spherical. Ehrenhaft was never persuaded of Millikan's results, but Millikan (1965) received the Nobel Prize in 1924 for this work.

The differing metaphysical assumptions—continuity versus atomism—led to differing experimental approaches and to divergent interpretations of the reliability of data. Since those metaphysical assumptions were not only in the background, but also the subject at issue for Ehrenhaft, the result was incommensurability. (For a much more detailed discussion of the controversy and issues, see Holton 1978.)

Influence

Assessments of Kuhn's influence vary enormously, and although most philosophers of science would agree that his influence was very large, some would not; and among those who do agree, there is disagreement about whether it was positive or negative. Those who argue that his influence was slight, point to the fact that almost all of the elements of *Structure* can be found in other philosophers of science writing at the time—Hanson (1958), Hesse (1966), Feyerabend (1993), Toulmin (1961), and others. However, the constellation of ideas in *Structure* and the rhetorical tone caught readers' attention in a way that produced much more dramatic results did than any of the others. Feyerabend made more radical claims than Kuhn and was largely dismissed or ignored, and the others made less sweeping claims and attracted less attention. For complex reasons that are not fully understood, *Structure* struck a resonant chord and transformed philosophy of science.

Some of the disagreement stems from unclear aspects of the book. As discussed earlier, although Kuhn changed his mind about the central term of the book, "paradigm," he did not rewrite the book to reflect that change. This decision was the result of Kuhn's recognition that reworking *Structure* was not a very good option, since he was still in the midst of reworking many of his views, and so the "postscript" strategy was a stopgap measure until he could reach the stage where a new and more thorough book was prepared. During the 1960s and 1970s Kuhn gave frequent graduate seminars on *Structure* and his further thoughts, as well as giving lectures and publishing intermediate elaborations. In 1977, he published *The Essential Tension*, a collection of his essays ranging from reprintings of pre-*Structure* papers to items that appeared for the first time in that volume. The "essential tension" referred to is that between the desire to assimilate all data/observations within the current paradigm and the desire to find revolutionary new solutions.

The characterization of revolutionary science has attracted the most attention—both positive and negative—from readers of *Structure.* Much of the popularity of the book outside the community of philosophers and historians derived from the conceptual tool it provided to analyze change and often to attempt to bring about a "change of paradigm." The book especially attracted interest from the social sciences because, in addition to the two types of science discussed earlier, Kuhn also described the "pre-paradigm" periods of the physical sciences before they achieved maturity. Whether any changes in these sciences are due to *Structure* or whether they are positive is mostly outside the scope of this entry. In particular, the widespread use of "paradigm" to mean something like a holistic worldview has its origin in the first edition of *Structure* and is rather distinct from the official approach of the second edition, which replaces "paradigm" with the disciplinary matrix with various explicit dimensions.

Although *Structure* was one of the major final blows to the logical positivist program, Kuhn's relations with at least two of the major figures in that program were cordial. *Structure* was originally published in the Encyclopedia of Unified Science series, which was the primary publishing format for logical empiricism (see Logical Empiricism). Carnap, one of the editors, was enthusiastic about publishing *Structure.* Although it was demonstrated in a paper by English (1978) that Carnap's own account of the relation between theory and evidence leads to incommensurability, since the meaning of theoretical terms is given implicitly by the relation of the theoretical terms to observation terms, this has not been generally noted. Also, in Carnap's account of probability and changes in probability assignments, he recognizes that in addition to conditionalization on new evidence, sometimes probability assignments are made in a global way that do not rely on evidence. This distinction is not exactly that between normal-science belief revision and revolutionary belief revision, but it is also not entirely foreign to it (see Carnap, Rudolf).

Hempel was a colleague of Kuhn's at Princeton and they were frequent interlocutors. Both influenced each other's views, and by the late 1970s their positions on many issues were very similar, although they had arrived at those positions from different directions. Kuhn came to emphasize more that revolutionary changes were made for good reasons, though he continued to assert that particularly at early stages of revolutionary change there were also good reasons against the eventual successful theory. And he continued to emphasize that there was no formal algorithm for theory evaluation that could be appealed to (see Hempel, Carl Gustav).

Some examples of Kuhn's influence within philosophy of science have already been given. Kitcher (1993) refers to Kuhn as one of the two most significant influences on his work and thought (the other is Hempel). Longino (1990) does not delineate her debts so explicitly, but there are far more references in her work to Kuhn than to any other philosopher of science. Kuhn's inclusion of values in the disciplinary matrix fostered the possibility of feminist philosophy of science, which studies, among other topics, the extent to which assumptions about gender influence the choice of research topics, funding, and scientific prestige, as well as biases with regard to theory selection (see Feminist Philosophy of Science). Another consequence of note was the development by Joseph Sneed, Wolfgang Stegmuller, and others of a formalization of theories inspired by Kuhn (for more details, see Theories).

One of the effects of *Structure* was that researchers in a number of fields beyond philosophy of science cited it as justification for emerging new domains in the study of science or for transformations of old ones. Sociologists of science, which had been dominated by a model of scientific development emphasizing the rational and cumulative character of scientific knowledge, was emboldened to investigate issues of authority and power and other issues at the level of the scientific group (Crane 1972). Kuhn's introduction of Gestalt psychology into discussions of scientific change helped to encourage both psychologists and philosophers to investigate these further. Productive examples can be found in Brewer and Chinn's (1994) research on anomalous data and Giere's *Explaining Science* (Giere 1988). Anthropology of science, which did not exist in any significant way prior to *Structure*, began to study scientific communities in the same ways in which other esoteric cultures were investigated (e.g., Traweek 1988). Some of the influence, and debates about it, continue. For example, Duschl (1990) and others are currently arguing for transformation in science education on the basis of what they perceive as a Kuhnian understanding of the process of scientific development.

In conclusion, the influence of Kuhn in philosophy of science is difficult to gauge because a great deal of what he argued for is now taken as part of the underlying assumptions. Studying history of science, having more realistic accounts of scientific development, and appreciating the relevance of theories of cognition and of social processes are all accepted and valued by the mainstream of philosophy of science. Thus what is attributed to Kuhn are primarily the claims about incommensurability and the dichotomous nature of scientific change, which are the less plausible parts of his views with the hindsight of fifty years. Kuhn's work has achieved a transformation of views of science that makes his most valuable contributions invisible to many current philosophers of science.

RICHARD GRANDY

References

Brewer, W. F., and C. A. Chinn (1994), "Scientists' Responses to Anomalous Data: Evidence from Psychology, History, and Philosophy of Science," *PSA* 1: 304–313.

Cartwright, Nancy (1983), *How the Laws of Physics Lie.* Oxford: Oxford University Press, Oxford.

Crane, Diana (1972). *Invisible Colleges: Diffusion of Knowledge in Scientific Communities.* Chicago: University of Chicago Press.

Cushing, James T. (1994), *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony.* Chicago: University of Chicago Press.

Duschl, Richard A. (1990), *Restructuring Science Education: The Importance of Theories and Their Development.* New York: Teacher's College Press.

Ehrenhaft, Felix (1941), "The Microcoulomb Experiment: Charges Smaller than the Electronic Charge," *Philosophy of Science* 8: 403–457.

English, Jane (1978), "Partial Interpretation and Meaning Change," *Journal of Philosophy* 75: 57–76.

Feyerabend, Paul (1993), *Against Method* (3rd ed.). New York: Verso.

Galison, Peter (1987), *How Experiments End*. Chicago: University of Chicago Press.

Giere, Ronald N. (1988), *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

——— (1999), *Science Without Laws*. Chicago: University of Chicago Press.

Hanson, Norwood Russell (1958), *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge, UK: Cambridge University Press.

Hesse, Mary (1966), *Models and Analogies in Science*. Notre Dame, IN: University of Notre Dame Press.

Holton, Gerald (1978), "Subelectrons, Presuppositions, and the Millikan-Ehrenhaft Debate," *Historical Studies in the Physical Sciences* 9: 161–224.

Horwich, Paul (ed.) (1993), *World Changes: Thomas Kuhn and the Nature of Science*. Cambridge, MA: MIT Press.

Hoyningen-Huene, Paul (1993), *Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science*. Translated by Alexander T. Levine. Chicago: University of Chicago Press.

Kitcher, Philip (1993), *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. New York: Oxford University Press.

Kuhn, T. S. (1957), *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*. Cambridge, MA: Harvard University Press.

——— (1977), *The Essential Tension: Selected Essays in Scientific Tradition and Change*. Chicago: University of Chicago Press.

——— (1978), *Black-Body Theory and the Quantum Discontinuity, 1894–1912*. Oxford: Oxford University Press.

——— (1996), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

——— (2000), *The Road since Structure: Philosophical Essays, 1970–1993 (with an autobiographical interview)*. Edited by James Conant and John Haugeland. Chicago: University of Chicago Press.

Lakatos, Imre, and Alan Musgrave (eds.) (1974), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

Longino, Helen E. (1990), *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

Masterman, Margaret (1974), "The Nature of a Paradigm," in Imre Lakatos and Alan Musgrave (eds), *Criticism and the Growth of Knowledge*. Cambridge, UK: Cambridge University Press, 59–89.

Millikan, Robert A. (1911), "The Isolation of an Ion, a Precision Measurement of its Charge, and the Correction of Stokes's Law," *Physical Review* 32: 350–397.

——— (1965), "The Electron and the Light-Quantum from the Experimental Point of View," *Nobel Lectures—Physics 1922–41*. Amsterdam: Elsevier.

Nersessian, Nancy J. (1984), *Faraday to Einstein: Constructing Meaning in Scientific Theories*. Boston: Kluwer Academic Publishers.

Nickles, Thomas (ed.) (2003), *Thomas Kuhn*. New York: Cambridge University Press.

Pickering, Andrew (ed.) (1992), *Science as Practice and Culture*. Chicago: University of Chicago Press.

Reichenbach, Hans (1938), *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press.

Salmon, Wesley, (1990), "Rationality and Objectivity in Science, or Tom Kuhn meets Tom Bayes." *Scientific Theories, V. 14, Minnesota Studies in Philosophy of Science*. Minneapolis: University of Minnesota Press, 175–204.

Shapere, Dudley (1964), "The Structure of Scientific Revolutions," *Philosophical Review* LXXIII: 383–394.

Suppe, Fredrick (ed.) (1977), *The Structure of Scientific Theories*. Urbana: University of Illinois Press.

Suppes, Patrick (1969), "Models of Data," *Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969*. Dordrecht, Netherlands: D. Reidel, 24–35.

Toulmin, Stephen (1961), *Foresight and Understanding*. New York: Harper and Row.

Traweek, Sharon (1988), *Beamtimes and Lifetimes: The World of High Energy Physicists*. Cambridge, MA: Harvard University Press.

See also **Carnap Rudolf; Feyerabend, Paul; Hanson, Norwood Russell; Hempel, Carl Gustav; Lakatos, Imre; Incommensurability, Observation; Popper, Karl Raimund; Protocol Sentences; Scientific Change; Scientific Progress; Scientific Revolutions**

L

IMRE LAKATOS

(5 November 1922–2 February 1974)

Lakatos was born Imre Lipsitz to Jewish parents in Budapest on November 5, 1922. He was not sent to a Jewish school, and when his family moved to Debrecen in eastern Hungary (in 1932) he attended the local *realgymnasium* (a secondary school with an emphasis on the sciences). Having excelled at school Lakatos went on to study mathematics, physics, and philosophy at the University of Debrecen, from which he graduated in 1944. During the war years he gravitated to the Marxist left. When the Nazi occupation of 1944 placed Hungarian Jews in mortal danger, he used a false identity to escape the labor gangs and the deportations, unlike many of his family and friends, including his mother and grandmother, who died in Auschwitz. During the occupation he belonged to an underground Marxist group, and late in 1944 he adopted the name "Lakatos" ("Locksmith"). After the war he continued his education, now at the prestigious Eötvös College in Budapest. At this time he wrote his first published works, on politics and its relation with science, written from a Marxist perspective. In 1947, he was awarded a Ph.D. for his dissertation *On the Sociology of Concept Formation in Natural Science.* Meanwhile, the Moscow-backed communist government in Hungary suffered from a shortage of dedicated Marxists to fill its offices. Thus it was that in 1947, the young Lakatos was attached to the Ministry of Education with responsibility for the "democratic reform of higher education." In practice this meant a rapid expansion in student numbers, together with brutal measures to bring independent intellectual centers under Party control—including Eötvös College, where Lakatos is still remembered for his role as denouncer-in-chief. At this time he was also a research student of the Hegelian-Marxist philosopher György Lukács, and he traveled to Moscow University in 1949. On his return to Hungary in the spring of 1950, he was arrested, charged with "revisionism," and imprisoned for almost four years (including a period in solitary confinement), one of many Party members caught up in the Stalinist purges. On his release after Stalin's death Lakatos returned to academia. Between 1954 and 1956, he worked on probability-and-measure theory under the mathematician Alfred Rényi. Crucially, one of his tasks was to translate into Hungarian György

Pólya's *How to Solve It*, a book on mathematical heuristics (Pólya, working in the United States, wrote *How to Solve It* for his undergraduates). At the same time, Lakatos began to question the entire edifice of Marxist thought. When demonstrations against the government erupted in October 1956, Lakatos was at the forefront of the student movement demanding academic and intellectual freedom. His conviction that science and philosophy should suffer no external control was by now firmly established. The uprising was quashed by Soviet troops, and Lakatos left Hungary for Vienna in late November 1956.

In 1957, he secured a Rockefeller fellowship to King's College, Cambridge, England, where (supervised by R. B. Braithwaite) he wrote a Ph.D. thesis, *Essays in the Logic of Mathematical Discovery,* a later version of which was eventually published as *Proofs and Refutations.* On taking his doctorate in 1960 he joined the London School of Economics (LSE), where he remained until his untimely death. In 1969, he was appointed professor of logic. During this time, his philosophical interests broadened to include physical science, and he developed his Methodology of Scientific Research Program. On the political front, the LSE was the British center of the student uprising around 1968. Remembering the damage done to Hungarian intellectual life by political interference both from the state and from student activists, Lakatos urged the university to resist demands for a student role in policymaking. In public he insisted on clear distinctions: between logic and psychology; between science and pseudo-science; and in this case, between the "constructive" student demand for the right to criticize the university and the "destructive" demand to take part in its decision-making processes. In private his tone was playful and his logic supple. In his publications he sharply distinguished his position from that of "irrationalists" such as Feyerabend (see Feyerabend, Paul); in his lectures he recommended Feyerabend to his students; and, in letters, he and Feyerabend argued for so long and took each other so seriously that in the end their positions were in danger of collapsing into one another. This combination of public dogmatism and private openness was motivated by an acute sense of the fragility of intellectual liberty. The enemies of free inquiry, whether they be totalitarian governments or ideologically blinkered students, cannot be resisted by the force of argument alone. Defenders of freedom, Lakatos thought, must stand ready to use whatever combination of rhetoric and military force the occasion demands.

In the early 1970s, Lakatos was full of plans for books and papers on mathematics and science, but he was dogged by ill health. He died after a heart attack on February 2, 1974, aged fifty-one.

Proofs and Refutations

Lakatos' earliest published works are Hungarian book reviews written in 1946–1947 when he was finishing his first Ph.D., *On the Sociology of Concept Formation in Natural Science.* These book reviews range over politics, science, and literature, but they invariably criticize their subjects from a Marxist perspective. For example, one author claims that as the result of scientific progress, modern man is alienated from nature—but, contends Lakatos, this is the case only under capitalism. In another review Lakatos complains that undialectical thinking makes a mystery of the fact that yesterday's progressive social class can become tomorrow's empty husk. Lakatos' concern in these reviews and in his work for the Ministry of Education was to place the resources and achievements of bourgeois intellectual life at the disposal of the new, postwar, communist reality. No doubt he reasoned that this was the only hope for their survival and growth. If intellectual culture remained bourgeois and "idealist," it would die for lack of relevance.

Lakatos had abandoned this Marxist framework by the time he began his Cambridge Ph.D. a decade later. Nevertheless, he retained a dialectically oriented study of the emergence of novel concepts. Indeed, two themes—the growth of knowledge and the relation between intellectual life and politics—dominated his writing from his first postgraduate publications of 1946 to his death in 1974. At the beginning of his Cambridge thesis, he declared: "The three major—apparently incompatible—'ideological' sources of [this] thesis are Pólya's mathematical heuristic, Hegel's dialectic and Popper's critical philosophy" (Lakatos [1962] 1978, 70n). The thesis is a case study in the growth of informal mathematics, that is, mathematics that have not been translated into the language of a system of formal logic. The case in hand is the Descartes–Euler formula, $V - E + F = 2$, where V is the number of vertices of a polyhedron, E the number of edges, and F the number of faces. A cube, for example, has 8 vertices, 12 edges, and 6 faces: $8 - 12 + 6 = 2$. The formula holds for the Platonic solids and many other polyhedra. Obvious questions arise: Does it hold for all polyhedra, or only for a special class of them? What exactly is a polyhedron anyway? Formal logic (that is, the logic expressed in such systems as the predicate calculus) offers no answers to these questions, though it may help to clarify answers found by some other means. If, therefore, one supposes that the rationality of mathematics lies entirely in its use of formal logic, then there can be no rational means of addressing these questions. One can only hope to stumble on an answer by a lucky guess, intuitive leap, or stroke of genius. This is where Pólya enters the picture.

Pólya's books on mathematical heuristics began as teaching aids. Some of his tips on mathematical research are altogether general (e.g., Do you know how to solve a problem similar or related to the one in hand?), while others are topic specific (e.g., If you want to discover new theorems about solids, remember that theorems in plane geometry often have three-dimensional analogues). These research strategies are fallible, but they do offer students and researchers in mathematics a more productive approach than simply tinkering with the material and hoping that luck or inspiration will supply an insight. Pólya's heuristics also suggest a philosophical account of the growth of mathematical knowledge (psychologistic talk of "inspiration" or "genius" is no account at all). Pólya suggested that Lakatos should use the Descartes–Euler conjecture as a case study. Philosophically, Pólya contributed three further thoughts to Lakatos' thesis:

1. Natural science and mathematics have many heuristic patterns in common. Pólya's work provides examples in pure mathematics of enumerative induction, inference to the best explanation, and the testing of general hypotheses by checking that their logical consequences are true.
2. The mathematical operations used to test a conjecture may eventually form the kernel of a proof. For example, one wishes to test the Descartes–Euler formula. One could try lots of different polyhedra in turn, but this would be hopelessly slow. One can, however, generate a vast collection of polyhedra out of the Platonic solids by "roofing" (building a pyramid taking one side of an existing polyhedron as its base) and "slicing" (cutting off a corner). It is easy to check that roofing and slicing do not change the alternating sum $V - E + F$. One has, therefore, checked the hypothesis for all polyhedra generated from the Platonic solids by roofing and slicing. But, by rearranging the elements, one also has the means to prove it for this class of solids.
3. Proofs and tests may suggest definitions and theorems. Pólya provides cases in which, rather than first seizing on definitions and conjectures by luck or insight and then later finding a suitable proof, mathematicians tailor their definitions and theorems to suit a promising proof idea. For example, the roofing-and-slicing idea suggests a definition: Let polyhedra constructed from the Platonic solids by finite iterations of roofing and slicing be called "P-constructable." The earlier argument is now the proof of a theorem: For all P-constructable polyhedra, $V - E + F = 2$.

The introduction of new definitions leads to Hegel. From Hegel, Lakatos learned that the arrival of new concepts in science is not usually announced by an explicit definition. Rather, existing concepts are developed in use. This may be a surreptitious expansion occasioned by consideration of a new sort of object. For example, in Lakatos' case study, the concept 'polyhedron' is quietly stretched as mathematicians contemplate solids with holes and solids formed by joining simple polyhedra at a vertex or along an edge. Lakatos also shows how a new proof can reinvent an entire field of study. Cauchy suggested a proof of the Descartes–Euler formula that involved removing one side of a polyhedron and flattening the remaining figure onto a plane. This is to treat a polyhedron as a closed surface rather than as a solid, and thereby to shift the problem from geometry to topology. Such changes, from minor concept-stretching to revolutions in the very subject matter, may not be apparent when they happen. They may come to light only later, when the growth of knowledge is rationally reconstructed. This explains the unusual literary structure of Lakatos' essay. The main text is a dialogue in a fictional mathematics class, while the historical sources are supplied in footnotes. Philosophical rational reconstruction is not concerned with accuracy in historical detail. Rather, the point is to make known the subterranean conceptual shifts masked by the use of old words for new ideas.

A consequence of Pólya's heuristic is that one may end up proving something deeper and more interesting than first intended. Lakatos' "dialectical" interest in conceptual change shows that as a result of trying to prove a theorem, one may end up with a whole new theoretical language. Thus, Lakatos' second Hegelian inheritance was a distinction between formal logic, which analyzes and rearranges the conceptual resources already available, and a heuristic rationality that develops new concepts out of old. A movement to a new language L_2 may introduce contradictions and refutations that were not present in the conceptual order associated with the old language L_1. Formal logic would recommend sticking with L_1, but this may be overridden if there are good heuristic reasons to prefer L_2. Indeed, the contradictions and refutations thus introduced may require improvements to the mathematical theory that might not otherwise have been made. Hegel expressed this distinction in psychologistic terms inherited from Kant: Formal logic is the natural tool of the Understanding, while "dialectical" or "speculative" thinking is the business of Reason. Lakatos preferred to distinguish between "language statics" and "language dynamics," but the point is the same. The logical analysis of concepts alone can only clarify and entrench the present conceptual order, when the real philosophical task is to understand the progress from one stage of conceptual development to the next.

Lakatos' third Hegelian lesson is that there is no general method for the development of concepts. Hegel is associated with the thesis–antithesis–synthesis schema, but one ought not to expect to apply this formula mechanically. In Hegel's work the development of any concept comes in three "moments," the third being in some sense a return to or rediscovery of the first, suitably modified by passage through the second. However, the details in any given case cannot be anticipated. There are no laws of dialectical growth as there are laws of formal logic. "Formalism," for Hegel, is the false supposition that there is a universal scientific method that may be grasped abstractly in advance of any particular inquiry. In his Cambridge thesis, Lakatos explicitly denied the possibility of a general system of heuristic rules, though this passage vanished from published versions of the text.

After Pólya and Hegel, Lakatos' third source was Popper and his critical philosophy (see Popper, Karl Raimund). Popper argued that empirical science is not built up by establishing true laws of nature. Rather it develops by the refutation of conjectures. Popper's view depends on the asymmetry between proof and disproof: No argument can conclusively prove a general statement because no argument can rule out altogether the possibility that an exception to the proposed law will be discovered in the future. On the other hand, a single counterexample is sufficient to refute a general statement. Because Popper put refutation rather than confirmation at the core of his model of science, he had no need for a nondeductive logic of induction. Pólya's thought that mathematics shares heuristic patterns with natural science enabled Lakatos to import Popperian ideas about science into his thinking about mathematics. Specifically, he developed Pólya's observation that a theorem and its proof may evolve together so that the final theorem is carefully tailored to capture the results of an insightful proof strategy. Lakatos gave this a Popperian gloss: Mathematics evolves through a process of conjecture and refutation. What is more (following Pólya), the same thought experiment can be both proof and test. This is most obvious when all the steps in a proof are reversible. However, a proof with nonreversible steps in it can also function as a source of counterexamples because it decomposes a theorem into its logical dependencies. One theorem may depend on many lemmas, and by finding a counterexample to a lemma, one may refute the theorem itself (alternatively one may learn that the lemma is stronger than it need be, and so one can improve the proof). Mathematics, in this view, is not about proving theorems from self-evident axioms. It is a matter of offering conjectures for refutation—but the refutations are found in proofs and other thought experiments rather than in empirical experiments and observations. This, then, is Lakatos' first Popperian thesis.

The second Popperian lesson is that formal logic is indispensable to the development of concepts. In Hegelian dialectic, no conception is simply false. A concept is found to be one-sided, partial, or otherwise inadequate. This inadequacy is exposed by contemplation of its dialectical twin. Finally, a new conception supersedes them both, repairing the inadequacy but preserving what was true in the original. Hegel thought that this process of conceptual evolution had to be separate from the work of formal logic. Formal logic requires that terms keep their meanings unchanged from start to finish, otherwise one commits the fallacy of equivocation. On the other hand, the very point of dialectic is to develop the meanings of terms. Hegel attempted to exhibit a dialectical logic that made no appeal to the notion of contradiction found in formal logic precisely because he understood the nature of formal rigor. However, this separation is artificial. Mathematical practice does not divide into formal and dialectical phases. Mathematical thought experiments, whether they function as proofs or tests, are structured by formal logic. The criterion for a successful proof is still given by the formal definition of a valid argument.

Lakatos' third Popperian claim is that mathematical heuristic does not begin inductively from facts, but from guesses at solutions to problems and questions. Neither science nor mathematics can start from bare facts because facts do not spontaneously order themselves into inductive tables. They can be so ordered only in the light of a conjecture, a problem, or some other governing idea. This is as true of mathematical facts about the parts of polyhedra as it is of empirical data. Indeed (still following Popper), there are no bare facts, in either science or mathematics. In particular, there are no self-evident mathematical axioms. Hence Lakatos is anti-foundationalist and fallibilist in his philosophy of mathematics.

Some of the fictional dialogue in Lakatos' Ph.D. thesis concerning the Descartes–Euler formula was published in four parts in the *British Journal for the Philosophy of Science* (Lakatos 1963–1964). This treatment is confined to the early history of the formula, during which the proofs and refutations made appeal to geometric intuition. It stops short of the absorption of the formula into the relatively abstract systems of modern algebra. A common criticism was that Lakatos' view did not apply to these more abstract parts of modern mathematics. Lakatos' plans to publish a more comprehensive account were cut short by his death in 1974, but selections from his thesis were published posthumously as *Proofs and Refutations* (Lakatos 1976). This material extended the Descartes–Euler story to include algebraic treatments, and included studies of topics from analysis and set theory. Lakatos was in no doubt that his heuristic view applied to the whole of mathematics. In the introduction to *Proofs and Refutations*, Lakatos described the book as an argument against "formalism," which he defined as the tendency to identify mathematics with its formal axiomatic abstraction. This, he argued, caused the philosophical neglect of everything about mathematics not captured in fully formalized axiom systems of meta-mathematics. In particular, the history and heuristics of mathematical practice vanish from sight. When one comes to assess Lakatos' philosophy of mathematics, it is well to distinguish this anti-formalism from the claim that the growth of mathematics is marked by dramatic refutations of a Popperian sort—for, in the case of advanced mathematics, the former claim is rather more plausible than the latter.

Transition to Research Programmes

Lakatos observed that his three "ideological" sources appeared to be inconsistent. In fact, the inconsistency is more than apparent. Popper held that epistemology can have no interest in the process of conjecture production and that philosophers ought to confine their attention to the logic of conjecture evaluation. Pólya, on the other hand, showed that there can be a fallible logic of conjecture production (his heuristic). Moreover, though he separated them in thought, Pólya's studies showed that the "contexts of discovery and justification" overlap in practice, because the same thought experiment can function as a test and then later as a proof. Unsurprisingly, the deepest contradictions are between the Popperian and Hegelian elements. Though Lakatos argued that theorems develop by responding to counterexamples, he quickly introduced a distinction between "logical" counterexamples (that is, counterexamples in the ordinary Popperian sense) and "heuristic" counterexamples (cases that, while not strictly inconsistent with the theorem in hand, show it to be conceptually deficient in some respect). Logical counterexamples belong to language statics (in Hegelian terms, the rigid, formal logic of the Understanding). Heuristic counterexamples belong to language dynamics (in Hegelian jargon, the dialectic of Reason). Popper was contemptuous of any philosophy that paid attention to fine distinctions of meaning. In his view, scientists use explicit, stipulative definitions to establish their terms with as much precision as required for the task in hand, and philosophers ought to do likewise. He had no sympathy with the Hegelian project of revealing the subtle shifts of meaning hidden in the evolving use of a single word. This, though, is precisely what Lakatos did with the central terms of his dialogue ('polyhedron,' 'proof,' etc.). Therefore, much of Lakatos' achievement in *Proofs and Refutations* could never be absorbed into Popperian philosophy.

The final contradiction in *Proofs and Refutations* between Popper and Hegel concerns the possibility of a general method. Hegel, though he organized some of his works into triples within triples, insisted that there is no general logic of scientific or philosophical progress. Each episode in the intellectual history of humanity has its own character, which must be traduced if it is forced into some all-purpose mould. Popper, for his part, believed himself to have discovered a general logic of science. Science, in his view, has a characteristic logical process (philosophically articulated as critical rationalism) that distinguishes it from other activities in general and from pseudo-science in particular. *Proofs and Refutations* vacillates between these extremes. On the Popperian side, Lakatos offers highly general heuristic patterns that might find application outside mathematics. There is no reason in principle why lemma incorporation or "monster adjustment" (to take two examples) should not be found in the development of legal arguments (for example) (*monster adjustment* is the redescription of an outlandish counterexample in such a way that it ceases to be outlandish and thereby ceases to be a counterexample.) On the Hegelian side, Lakatos offers no *general* solution to the problem of identifying ad hoc-ness and degeneration, preferring instead to appeal to the judgment of persons with refined mathematical taste. Indeed, he argues that any of his heuristic patterns can lead to triviality and degeneration if pursued mindlessly (in this sense *Proofs and Refutations* is an anarchist tract in the sense of Feyerabend). These tensions between the Hegelian and Popperian elements of Lakatos' early views became acute as he tried to move beyond the piecemeal studies of his Ph.D. thesis.

As it turned out, Lakatos tended to land on the Popperian side of dilemma. In 1965 (the year after he published part of his thesis in the *British Journal for the Philosophy of Science*), he wrote a brief paper titled "A Renaissance of Empiricism in the Recent Philosophy of Mathematics?" (Lakatos 1978b). In this paper he developed a contrast between "Euclidean" and "quasi-empirical" theories. In a so-called Euclidean theory, truth flows down from self-evident axioms to derived theorems. In a quasi-empirical theory, falsehood is transmitted up from theorems to axioms. Such theories are *quasi*-empirical because the falsifying criticism need not come from empirical observation or experiment. He went on to claim that mathematics is quasi-empirical in this sense. In other words, he stripped the Hegelian elements from his earlier work, leaving a straightforwardly Popperian philosophy of mathematics that required him to claim that the growth of mathematics is marked by Popperian refutations.

The Methodology of Scientific Research Programs

While Lakatos was working on his philosophy of mathematics, the Popperian school had identified the alleged relativism and irrationalism of Thomas Kuhn as the principal philosophical threat to free enquiry (see Kuhn, Thomas). Lakatos agreed that Kuhn's philosophy had to be resisted, but by the late 1960s, he was convinced that Popper's account of scientific method was not adequate. To meet the need, he developed his own Methodology of Scientific Research Programs (Lakatos 1978a) (see Research Programs). In his Cambridge Ph.D. dissertation, he had suggested that theorems and proofs ought not to be regarded as separate entities. A theorem and its proof evolve together, and this common evolution makes sense of the terms employed in stating the theorem and the lemmas deployed in the proof. The natural unit of philosophical appraisal is thus not the theorem, but the theorem–proof pair. Similarly, in *Methodology of Scientific Research Programmes*, he replaced the single theory as the unit of appraisal with the "program," a temporal sequence of theories. In both cases the point was to allow philosophers to understand the growth of knowledge over time.

Two features distinguish the Methodology of Scientific Research Programs from Lakatos' earlier work on mathematics. One is the almost total absence of any discussion of shifts in meaning. The distinction between language statics and language dynamics that motivated the search for hidden conceptual changes was quietly dropped. The other is the specification of criteria by which progress and degeneration might be judged. The early Lakatos agreed with Hegel (and Feyerabend) that there can be no general rule for judging scientific progress, because each moment in the growth of knowledge has its own inner logic, which it is the task of philosophy to exhibit. In the absence of a general rule, it is a matter of judgment whether this or that development represents progress or degeneration. Feeling the threat of Kuhn's "irrationalism," the later Lakatos tried to articulate explicit criteria to do the job that he had previously left to good scientific taste. Indeed, he dismissed the exercise of taste as "elitism," which is ironic given his determination to exclude students from university decision making. At the same time, Feyerabend, in conversation and correspondence, persuaded Lakatos that a rigidly mechanical rule would fail to respect the special merits of individual programs. Therefore, he found himself arguing that his Methodology of Scientific Research Programs is *both* sufficiently exacting to distinguish progressive from degenerative programs *and* sufficiently supple to account for particular cases.

After Lakatos' death, there were sporadic attempts to develop his Methodology of Scientific Research Programs and to apply it more widely than Lakatos was able to in his lifetime (e.g., Howson 1976). Some philosophers attempted to develop his papers on mathematics into a Methodology of Mathematical Research Programs (e.g., Hallett 1979). None of these efforts succeeded in rebutting the charges of rigidity and arbitrariness leveled against Lakatos' later work, and the supply of would-be heirs eventually dried up. On the other hand, there has lately been a rising tide of historically oriented philosophy of mathematics, for which *Proofs and Refutations* is often cited as an inspiration. *Proofs and Refutations*, with its three ideological sources, is subtle and rich, while *Methodology of Scientific Research Programmes* is relatively rigid and ideologically uniform. Moreover, Kuhnian relativism was a temporary threat, while formalism is a permanent menace. For these reasons, one may expect *Proofs and Refutations* to be the most durable part of Lakatos' contribution to philosophy.

Brendan Larvor

References

Hallett, Michael (1979), "Towards a Theory of Mathematical Research Programmes," *British Journal for the Philosophy of Science* 30: 1–25, 135–159.

Howson, Colin (ed.) (1976), *Method and Appraisal in the Physical Sciences*. Cambridge: Cambridge University Press.

Lakatos, I. ([1962] 1978) "Essays in the Logic of Mathematical Discovery," in Worrall and Currie (eds), *Mathematics, Science and Epistemology (Philosophical Papers volume 2)*. Cambridge: Cambridge University Press.

——— (1963–4), "Proofs and Refutations," *British Journal for the Philosophy of Science* [BJPS] 14: 1–25, 120–139, 221–245, 296–342.

——— (1976), *Proofs and Refutations*. Edited by Worrall and Zahar. Cambridge: Cambridge University Press. (Consisting of the BJPS article plus additional material from the Ph.D. thesis)

——— (1978a), *The Methodology of Scientific Research Programmes (Philosophical Papers volume 1)*. Edited by Worrall and Currie. Cambridge: Cambridge University Press.

——— (1978b), "A Renaissance of Empiricism in the Recent Philosophy of Mathematics?" in Worrall and Currie (eds), *Mathematics, Science and Epistemology (Philosophical Papers volume 2)*. Cambridge: Cambridge University Press, 24–42.

See also **Feyerabend, Paul; Kuhn, Thomas; Popper, Karl Raimund; Scientific Change; Scientific Progress**

LAWS OF NATURE

The main difficulty in developing an account of laws is to distinguish laws from accidental truths without leaving mysterious how scientists could ever gain knowledge of laws. Those accounts that adequately distinguish laws from accidental truths fail to make clear how scientists could ever determine which general truths are laws; and those that make knowledge of laws feasible seem unable to adequately distinguish laws from accidental generalizations.

Contrast the following generalizations:

All spheres of gold have a mass of less than 100,000 kg.
All spheres of U^{235} have a mass of less than 100,000 kg.

Assuming both are true, the former is merely accidentally true, whereas the latter expresses a lawful relation; the critical mass of U^{235} makes it impossible for there to be spheres of U^{235} of such a large mass. The question is what makes for the difference. A simple regularity view of laws is inadequate, since both generalizations describe regularities. Moreover, the distinction cannot be drawn based on features of the generalizations themselves, as Hempel and Oppenheim (1948) attempt to do; both generalizations are universal in form, are of unlimited scope, make no essential reference to particulars, and involve purely qualitative predicates. Therefore, either there must be something outside the regularity (and the generalization used to describe it) that makes one a law and the other not, or law claims must be in some way stronger than universal generalizations. The former option is the route taken by sophisticated regularity accounts, which Dretske (1977) labels "universal truth + *X*." According to such views, laws are simply regularities (leaving aside probabilistic laws for the moment), but some regularities are not laws. There is nothing about the regularities themselves that distinguishes laws from accidental regularities. Law statements are universal generalizations that satisfy some additional external criterion. There are as many such accounts as there are functions for laws. For example, laws might be those generalizations that are used to make predictions (e.g., Goodman 1954), are resilient (Skyrms 1980), function in explanations (e.g., Braithwaite 1953), are integrated into the best systematization of the facts (e.g., Lewis 1973), and so on. Alternatively, many have argued that laws are not mere regularities, but are ontologically stronger. There are two main accounts that follow this route: those that conceive of laws as nomically necessary regularities (e.g., Pargetter 1984) and those that conceive of laws as relations between universals (Dretske 1977; Armstrong 1983; Tooley 1987).

Possible-Worlds Accounts

Perhaps the most natural way to distinguish laws from accidental regularities is to conceive of them as nomically necessary truths, since laws seem to involve some sort of natural (as opposed to logical) necessity. "It is a law that α" is equivalent to "It is nomically necessary that α"; and the latter is true if and only if α is true in all nomically accessible worlds. Accidental generalizations, as distinguished from laws, are not true in all nomically accessible worlds.

Four problems arise for such accounts. First, they are committed to metaphysically dubious entities. Bigelow and Pargetter (1990) have attempted to address this problem by conceiving of possible worlds as complex structural universals. However, one might still find this objectionable, since it requires a commitment to uninstantiated structural universals, else there would be only one possible world—the complex structural universal that is instantiated by the actual world. In fact, in order to use possible worlds to make sense of logical necessity, such an account requires the existence of nomically impossible universals. Otherwise, "It is a law that α" would be equivalent to "It is logically necessary that α," since all possible worlds would be nomically accessible.

This leads to a second difficulty, that of providing an account of nomic accessibility that does not rely on a prior understanding of lawfulness. It would be circular, for example, to analyze "It is a law that α" is true if and only if α is true in all nomically accessible worlds and then require an understanding of "It is a law that α" to make sense of nomic accessibility. Pargetter's response is analogous to Lewis's (1986a) argument that his analysis of possible worlds is not circular. Lewis argues that possible worlds exist and understanding the notion of a possible world does not rely on a previous understanding of possibility, since possible worlds are like the actual world, differing only in what occurs in them. Pargetter (1984) argues likewise that the nomic accessibility relation exists and that one need not have a prior understanding of lawfulness to understand this accessibility relation. Nomic accessibility does not require that all accessible worlds have the same laws; it merely requires that the appropriate generalizations be true in these worlds. Moreover, what this involves is easily understood, since it is analogous to a generalization being true in different parts of the universe. The central problem with this response is that it is uninformative, since it simply treats the accessibility relation as a primitive. (For alternative accounts, see Vallentyne 1988 and Mormann 1994.)

The third difficulty is whether such accounts can make sense of probabilistic laws. Tritium has a half-life of 12.26 years, which means that tritium atoms have a probability of $\frac{1}{2}$ of decaying in 12.26 years. There are typically two ways to understand this probability in terms of possible worlds: One might consider one particular tritium atom and understand the probability for this atom as the proportion of nomically possible worlds (with the same history until now) in which that atom (or its counterpart) decays in the next 12.26 years; alternatively, one might consider *all* tritium atoms in all nomically possible worlds and understand the probability as the proportion of all nomically possible tritium atoms that decay in 12.26 years. In order for such accounts to even get off the ground, an assumption must be made about the likelihood of each possible world—typically that each is equally probable. Of course, if probabilities are grounded in possible worlds, it is unclear what could ground this probability assignment. It might be taken as a primitive, but that would not answer which primitive probability assignment should be used—should they be treated as equally probable, or should some be treated as more likely than others? Moreover, as van Fraassen (1989) has argued, if there are a countable infinity of possible worlds, they cannot be treated as equally likely, and "if they are infinitely many and form a continuum (surely the most plausible idea) then it literally makes no sense to say: the objective chance is *the* measure that treats them all as equally likely" (79).

Perhaps the most serious difficulty faced by possible-worlds accounts is how scientists could ever gain knowledge of which generalizations are laws. It would require knowing not only what goes on in other possible worlds, but also which possibilities are nomically accessible. Knowledge of logic might yield answers to the former, but knowledge of the latter is more problematic. Consider Bigelow and Pargetter's (1990) account that treats possible worlds as complex structural universals. The question is how one could know which complex structural universals are nomically possible and which are not. Perhaps scientists could run experiments to determine which lower-level structural universals are possible. Experiments allow scientists to instantiate possibilities that might otherwise not have existed; and since actuality implies possibility, this would provide some access to which structural universals are possible. However, since it is impossible to instantiate more than one most complex structural universal, that is, the one that is instantiated by the entire actual world, there is at least one level at which it is in fact impossible to instantiate other possibilities. Moreover, to determine which generalizations are nomically necessary, scientists would have to know what all the nomic possibilities are, or which structural universals are nomically impossible, and it is unclear how instantiating some of the possibilities yields knowledge of which are impossible.

The problem for probabilistic laws seems even worse, since it is unclear what the proportion of possible worlds with a certain type of event implies about the proportion of those events in the actual world (van Fraassen 1989), much less what the proportion of a type of event in the actual world implies about the proportion of events in other possible worlds, which is what knowledge of probabilistic laws would need to rely on. To resolve these difficulties, one might argue that scientists can simply use the methods of inference already used to figure out which generalizations are laws. However, such a response is inadequate, since the question is precisely how these methods could yield reliable knowledge laws, *if* laws are understood as nomically necessary truths. One could argue that this simply shows that laws ought not to be understood as nomically necessary truths, since it is possible for scientists to gain knowledge of which generalizations are laws, and they could not have such knowledge were laws nomically necessary truths.

Universals Accounts

The three main proponents of the universals account (Dretske 1977; Armstrong 1983 and 1997; Tooley 1987) have slightly different formulations, but the general idea is to conceive of laws as relations between universals. Law statements, rather than being universally quantified statements ranging over individuals, are singular statements about the relations between universals. However, not all relations between universals are laws, since accidental generalizations can be described as second-relations of extensional inclusion between universals. Tooley (1987) and Armstrong (1997) differ in their accounts of what makes laws distinct. According to Tooley, statements expressing nomological relations are contingent, irreducibly and purely of order two or higher, and logically entail the appropriate first-order generalization. Probabilistic laws are contingent, irreducible *probabilification* relations between universals, where the relation of probabilification between universals F and G makes it the case (due to logical probability) that, given that x has F, it is probable to degree k that x has G. According to Armstrong, laws are irreducible relations of necessitation (1983) or causation (1997) between universals, and probabilistic laws are probabilistic relations between universals that specify the probability of an instance of one universal necessitating or causing an instance of a second.

Both accounts successfully distinguish laws and accidental generalizations, since the relation of extensional inclusion is reducible to a first-order generalization and is neither a necessitation nor a causal relation between universals. This comes, however, at a price. In particular, it is unclear how such claims could entail anything about what happens in particular instances. Moreover, this problem infects accounts of both deterministic and probabilistic laws. (See van Fraassen 1989 for a discussion of this and other problems, and Armstrong 1997 for a reply.)

Even if this problem can be solved, another difficulty remains. If laws are relations between universals, how could scientists gain knowledge of which generalizations correspond to lawful relations and which are merely accidental? Armstrong argues that scientists can observe causal relations in single instances, draw causal generalizations from these, and then use inference to the best explanation (IBE) to infer that the causal regularity holds due to a lawful relation between the universals instantiated in the causal instances. Leaving aside problems with observing causation in single instances (see Causality), this would not allow scientists to discriminate laws and accidental regularities. For Armstrong, there are two kinds of accidental regularities: causal and noncausal regularities. Assuming scientists could observe causation in single instances, the latter could be excluded. However, there might be accidental causal regularities. The question is how IBE could discriminate an accidental causal regularity from a lawful one.

Similar problems arise for Tooley. Tooley argues that when $(x)(Px \rightarrow Qx)$ survives potential falsification, there is good reason to infer that there is a nomological relation between universals P and Q. In fact, it is only if laws are relations between universals that surviving potential falsification can provide support for the lawful status of a generalization. Otherwise, assuming that the universe is potentially infinite, assigning a nonzero probability to a universal generalization would be unjustifiable. Of course, since generalizations are supposed to follow from law claims, the law claim cannot be better confirmed than the generalization (Woodward 1992). Moreover, scientists would be unable to discriminate a generalization that is accidental from one that has a nomological relation. Either no accidental generalization can survive falsification (but then Tooley needs to clarify why) or scientists need to determine, prior to (or independently of) the attempted falsification, which generalizations are nomological and therefore ought to be assigned nonzero prior probabilities.

An alternative account that falls broadly into the universals category treats laws as metaphysically necessary truths, much like "Water is H_2O," rather than as contingent relations between distinct universals. The central motivation for such views is the idea that properties are individuated, or perhaps even constituted, by the causal powers they have and, therefore, by the lawful relations in which they stand (e.g., Shoemaker 1980; Swoyer 1982).

Sophisticated Regularity Accounts

There have been numerous attempts to develop sophisticated regularity accounts that conceive of laws as universal truths satisfying some additional functional requirement. The most fully developed of these is the best systems account, according to which laws are those generalizations that function in the appropriate way in the best systematization of the facts. There are various versions of this account that differ in significant respects (Ramsey 1928; Kitcher 1986 and 1989; Lewis 1973) (see also Ramsey, Frank Plumpton). However, the focus here is on Lewis's later (1994) version, according to which laws are those generalizations that are axioms or theorems in the true deductive system that achieves the best balance of simplicity and strength. Simplicity and strength often conflict, so the best system must balance the two. The use of simplicity also requires that the predicates used in the axioms refer to natural kinds. Otherwise, one could use gerrymandered predicates to artificially alter the simplicity of the system. Moreover, there might be different criteria used to measure strength, simplicity, and the balance between them. Lewis's hope is that nature will be kind and there will be only one system that will be best on any standards of strength, simplicity, and balance. If nature is unkind, then there may be nothing deserving the title of 'laws.'

To extend this account to probabilistic laws (or even deterministic laws in a chancy world) the systems must be limited to those that never had any chance of being false (Lewis 1986b). Probabilistic laws are those generalizations about chance that are theorems in the best system, where the best system balances simplicity, strength, and fit, and fit is understood in terms of the degree to which the probabilistic laws conform to the actual course of history. In other words, the chance of the actual history occurring will be higher according to some systems than others; the higher the chance, the better the fit. Since fit is not the only criterion the best system must satisfy, probabilities will not in general be equivalent to actual frequencies. Lewis was initially doubtful that this account of laws would work, since it seemed to lead to a contradiction when combined with his *principal principle* (Lewis 1986b) and ultimately led him to revise it (Lewis 1994). Others have argued that the contradiction need not have arisen in the first place (e.g., Roberts 2001).

This account of laws appears to avoid the epistemic difficulties faced by other accounts. Scientists can justify law claims by determining which universal generalizations are axioms or theorems in the best deductive systems. It is even reasonable to think that scientists already use strength and simplicity as a guide to theory choice. Nevertheless, epistemic problems do arise. A central difficulty is accounting for how scientists could distinguish natural from nonnatural kinds. One common answer to how this is done—that natural kinds are those picked out by the best theories—is not open to Lewis, assuming simplicity plays a role in theory selection. If simplicity guides theory choice (and it must if science is to discover laws, according to Lewis), then science's best theories might fail to pick out natural kinds, since it will be possible to use nonnatural predicates to achieve gains in simplicity. This is precisely why Lewis added this requirement in the first place. (For a discussion of other potential epistemic difficulties, see van Fraassen 1989, 55–59.) Nevertheless, the epistemic problems faced by this account seem, at least on the surface, to be less daunting than those of other accounts. Lewis's account also has the advantage of making sense of the connection between laws and modality, without relying on modal notions to define laws. Nomic necessity is defined in terms of laws, rather than the other way around. A proposition is nomically necessary if and only if it is entailed by the laws of nature. In terms of possible worlds, a world W' is nomically possible relative to another world W if and only if the laws of W are true in W', though they may not be laws in W'. A proposition is nomically necessary if and only if it is true in all nomically possible worlds. Lewis is also able to distinguish between laws and accidental generalizations, since not all generalizations will be axioms or theorems in the best system. This will presumably apply to such claims as "All pieces of gold have a mass of less than 10,000 kg."

This leads to the fundamental problem faced by such Humean positions: They fail to adequately account for the necessity involved in laws and therefore incorrectly distinguish between laws and accidental generalizations. While Lewis's account does provide for a distinction between laws and accidental generalizations, the question is whether it draws the distinction correctly. Criticisms have generally taken the form of counterexamples that attempt to show that Lewis's account yields the wrong answer about which generalizations are laws (Armstrong 1983; Tooley 1987; Carroll 1994; van Fraassen 1989). While these counterexamples are decisive against Lewis only if one shares the intuitions about which generalizations ought to count as laws or about how laws and modal notions are connected (Loewer 1996), they nevertheless make clear why many find Lewis's account inadequate. One counterintuitive consequence of Lewis's account is that there will be some nomic possibilities compatible with the laws but not compatible with their being the laws. This leads to a revision of the distinction between initial conditions and laws. As a result, Lewis's account carves the distinction between laws and accidental generalizations in a way many find counterintuitive.

Other Approaches and Issues

In response to these difficulties, Carroll (1994) and Lange (2000) argue that it is impossible to give a reductive analysis of laws that does not rely on nomic notions, while van Fraassen (1989) and Giere (1999) argue that there are no laws. Others have argued that there are no strict laws, even in fundamental physics (Cartwright 1983 and 1989). Instead, law statements require *ceteris paribus* clauses or perhaps might be better understood as claims about capacities. Earman and Roberts (1999) disagree with Cartwright about fundamental physics, but, building on an insight of Hempel's (1988), argue that the special sciences can have no strict laws, at least not formulated purely in the language of that science. (For related discussion about psychological laws, see Davidson 1970; for biological laws, see Beatty 1995; for economic laws, see Hausman 1992; and for social science laws, see Kincaid 1990.) The use and nature of laws has also played an integral part in debates about numerous other philosophical issues, such as causation, explanation, reductionism, determinism, confirmation, and induction.

Jessica Pfeifer

References

Armstrong, David (1983), *What Is a Law of Nature?* Cambridge: Cambridge University Press.

——— (1997), *A World of States of Affairs*. Cambridge: Cambridge University Press.

Beatty, John (1995), "The Evolutionary Contingency Thesis, " in Gereon Wolters and James Lennox (eds.), *Concepts, Theories and Rationality in the Biological Sciences*. Pittsburgh: Pittsburgh University Press, 45–81.

Bigelow, John, and Robert Pargetter (1990), *Science and Necessity*. Cambridge: Cambridge University Press.

Braithwaite, R. (1953), *Scientific Explanation*. Cambridge: Cambridge University Press.

Carroll, John (1994), *Laws of Nature*. Cambridge: Cambridge University Press.

Cartwright, Nancy (1983), *How the Laws of Physics Lie*. Oxford: Clarendon Press.

——— (1989), *Nature's Capacities and Their Measurement*. Oxford: Oxford University Press.

Davidson, Donald (1970), "Mental Events," in *Essays on Actions and Events*. Oxford: Clarendon Press, 207–225.

Dretske, Fred (1977), "Laws of Nature," *Philosophy of Science* 44: 248–268.

Earman, John, and John Roberts (1999), "Ceteris Paribus, There Is No Problem of Provisos," *Synthese* 118: 439–478.

Giere, Ronald (1999), *Science without Laws*. Chicago: University of Chicago Press.

Goodman, Nelson (1954), *Fact, Fiction, and Forecast*. London: Althone Press.

Hausman, Daniel (1992), *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.

Hempel, Carl (1988), "Provisos: A Problem Concerning the Inferential Function of Scientific Laws," in Adolf Grunbaum and Wesley Salmon (eds.), *The Limits of Deductivism*. Berkeley and Los Angeles: University of California Press, 19–36.

Hempel, Carl, and Paul Oppenheim (1948), "Studies in the Logic of Explanation," *Philosophy of Science* 15: 135–175.

Kincaid, Harold (1990), "Defending Laws in the Social Sciences," *Philosophy of the Social Sciences* 20: 56–83.

Kitcher, Philip (1986), "Projecting the Order of Nature," in Robert Butts (ed.), *Kant's Philosophy of Physical Science*. Dordrecht, Netherlands: D. Reidel, 201–235.

——— (1989), "Explanatory Unification and the Causal Structure of the World," in Philip Kitcher and Wesley Salmon (eds.), *Scientific Explanation*. Minneapolis: University of Minnesota Press, 410–505.

Lange, Marc (2000), *Natural Laws in Scientific Practice*. Oxford: Oxford University Press.

Lewis, David (1973), *Counterfactuals*. Cambridge, MA: Harvard University Press.

——— (1986a), *On the Plurality of Worlds*. Cambridge: Basil Blackwell.

——— (1986b), *Philosophical Papers: Volume II*. New York: Oxford University Press.

——— (1994), "Humean Supervenience Debugged," *Mind* 103: 473–490.

Loewer, Barry (1996), "Humean Supervenience," *Philosophical Topics* 24: 101–127.

Mormann, Thomas (1994), "Accessibility, Kinds, and Laws: A Structural Explication," *Philosophy of Science* 61: 389–406.

Pargetter, Robert (1984), "Laws and Modal Realism," *Philosophical Studies* 46: 335–347.

Ramsey, Frank (1928), "Universals of Law and of Fact," in David Mellor (ed.), *Philosophical Papers*. Cambridge: Cambridge University Press.

Roberts, John (2001), "Undermining Undermined: Why Humean Supervenience Never Needed to Be Debugged (Even if It's a Necessary Truth)," *Philosophy of Science* 68 (Proceedings): S98–S108.

Shoemaker, S. (1980), "Causality and Properties," in van Inwagen, P. (ed.), *Time and Cause*. Dordrecht: Reidel, 109–136.

Skyrms, Brian (1980), *Causal Neccessity*. New Haven, CT: Yale University Press.

Swoyer, C. (1982), "The Nature of Natural Laws," *Australian Journal of Philosophy* 60: 203–223.

Tooley, Michael (1987), *Causation*. Oxford: Clarendon Press.

Vallentyne, Peter (1988), "Explicating Lawhood," *Philosophy of Science* 55: 598–613.

van Fraassen, Bas (1989), *Laws and Symmetry*. Oxford: Clarendon Press.

Woodward, James (1992), "Realism about Laws," *Erkenntnis* 36: 181–218.

See also **Biology, Philosophy of; Causality; Chemistry, Philosophy of; Confirmation Theory; Determinism; Economics, Philosophy of; Explanation; Hempel, Carl Gustav; Induction, Problem of; Mechanism; Physicalism; Reductionism; Scientific Models; Theories**

PHILOSOPHY OF LINGUISTICS

Linguists study various topics, ranging from politeness register to the etymology of the word "dude"; but what has caught the attention of philosophers (and what has given rise to interesting questions in the philosophy of science) has been work in generative linguistics as developed by Noam Chomsky (see Chomsky, Noam) and several generations of his students. Generative linguistics per se has a tradition that dates back further than Chomsky (e.g., to work in phonology by Roman Jakobson), but Chomsky was the first to bring serious formal methods to the field and the first to pursue ways of embedding linguistic theory in other more basic sciences, such as biology. The field that has emerged in the wake of this effort has seen its share of empirical successes and continues to show promise, but it has also given rise to a series of interesting debates about the nature of language and scientific practice.

Among the issues that have emerged with the development of generative linguistics are questions about the nature of language and the object of inquiry in linguistics, about the role of reference and language/world relations, and about the plausibility and nature of rules and representations in linguistic theorizing, as well as a cluster of issues surrounding methodology in linguistics (among other things, for example, generative linguistics incorporates an unabashed appeal to intuitions as part of its data). The most radical issue, however, concerns how to understand the object of study itself.

The Object of Study

Most current work in generative linguistics holds that the object of study is not an external abstract language that one comes to acquire, but rather the study of the faculty that underwrites one's linguistic competence. The motivation for this is clear enough: The commonsense notion of a language has been under pressure for some time. Consider Weinreich's contention that a language is a dialect with an army and a navy: Why is Italian a language and Veneto a dialect? It seems to be a purely political decision, not driven by any facts about the linguistic forms themselves. Individuating dialects (or identifying a natural group of "same-language speakers") is no more feasible than separating dialects from languages. Exactly when do two people speak the same dialect? There are differences between the way any two people speak—does that make them speakers of different dialects? In the end, people say that they "speak the same dialect" when they identify with each other enough. Once again, linguistic identity recapitulates political identity. Even trying to say that a particular agent speaks a single idiolect (his/her own individual language), which might be identified by a series of rules, does not stand up to serious scrutiny. Which set of rules is appropriate for describing one's idiolect? The way one speaks shifts radically depending upon one's age, discourse partners, and context. Is the conversation in a classroom? in a bar? with an interlocutor from a foreign country? What sentences one produces will vary markedly from situation to situation. Then, too, there is the question of errors. One may stutter and stammer or hiccup during a conversation. Are those noises to be counted as part of one's idiolect? If not, then, why not?

The considerations just reviewed provide good reason for being suspicious of language as an abstract external object, whether construed nationally, locally, or individually. But what is to replace this conception? Generative linguists hold that one should focus on the faculty that underwrites linguistic competence. That is, they are not interested in languages so much as the tacit theory that the agent deploys in the production and comprehension of linguistic behavior (or at least in judging the acceptability of certain linguistic forms).

This tacit theory is also enlisted to account for the fact that the data the language learner is exposed to are not sufficient (by themselves) to explain the linguistic competence that the agent comes to have. The language learner faces the problem, familiar in the philosophy of science, that the available evidence radically underdetermines the theory. The sentences that a child hears—and even the explicit corrections and affirmations it receives—when first acquiring a language are compatible with infinitely many, wildly divergent grammars. Nonetheless, the learner apparently manages to select a grammar. And, by and large, learners in similar environments select similar grammars. Infinitely many grammars, perfectly compatible with the evidence the child has access to, are simply ignored. Understanding the nature of the tacit theory that language users employ is one way of illuminating just how it is that agents are able to acquire the linguistic competence they do. Current linguistic and psychological theories suggest that the tacit theory employed by a linguistic agent may be part of human biological endowment and, as a corollary, that the mechanisms under study in linguistic theory will be embedded in human cognitive psychology and ultimately in human biology (see Pinker 1994 for a readable account and defense of this view).

This view about the object of study is not shared by all linguists. For example, Katz (1985) endorsed a Platonist view that takes the object of study in linguistics to be an abstract mathematical object outside of space and time—this would contrast with a position like Chomsky's, in which the object of study is a mental object of some form. The Platonist view has been advanced most visibly by Katz (1981), and it has been at least endorsed by Gazdar et al. (1985). It may be, however, that the position rests on a confusion. Higginbotham (1983) has observed that even if grammars are abstract objects, there is still the empirical question of which grammar a particular agent is employing. George (1989) has further clarified the issue, holding that one needs to distinguish between (i) a grammar, which is the abstract object that an agent knows, (ii) a psycho-grammar, which is the cognitive state that constitutes the agent's knowledge of the grammar, and (iii) a physio-grammar, which is the physical manifestation of the psycho-grammar in the brain. If this picture is right, then the Platonist position in linguistics may be trading on a failure to distinguish between grammars and psycho-grammars.

Perhaps more pressing is the dispute between what Chomsky (1986) has characterized as conceptions of language of *E*-language and *I*-language. From the *E*-language perspective, a natural language is a kind of social object the structure of which is purported to be established by convention (Lewis 1975) (see Conventionalism), and persons may acquire varying degrees of competence in their knowledge and use of that social object. In Chomsky's view, such objects would be of little scientific interest if they did exist (since they would not be "natural" objects), but in any case such objects *do not* exist. Alternatively, an *I*-language is not an external object, but is rather a state of an internal system that is part of the agent's biological endowment. An agent might have *I*-language representations of English sentences, but those internal representations are not to be confused with spoken or written English sentences. They are rather data structures in a kind of internal computational system.

Chomsky understands the *I*-language computational system to be individualistic (see Methodological Individualism). That means that the properties of the system can be specified independently of the environment that the agent is embedded in. Thus, it involves properties like the agent's rest mass and genetic makeup, but not relational properties like the agent's weight and IQ. The difference can be described by analogy to the difference between primate physiology and primate ecology. The former study is "narrow" in that it is concerned with the properties and structure of the primate in isolation (e.g., with bone and muscle architecture). The latter study is "wide" in that it is concerned with primate/environment relations (e.g., with the role the primate's musculature might play in its interactions with its environment—allowing it to swing from tree to tree, say). In Chomsky's view, linguistic theory is much more analogous to physiology, and is narrow in precisely that sense. Both the claim that *I*-language is individualistic and the claim that it is computational have led to a number of philosophical skirmishes.

The Language Faculty and the External World

One of the immediate questions raised by the idea that *I*-language is individualistic has to do with how semantics could be possible—in particular, how *referential* semantics could be possible, where referential semantics is a theory of the relation between linguistic forms and aspects of the external world. The worry is this: If generative linguistics is a chapter of narrow (individualistic) psychology, and semantics (and indeed meaning) is concerned with relations of language/world (or at least mind/world), then how are the two enterprises to be squared? There are two parts to the question. Is it really the case that meaning involves language/world relations? And, if so, then how can semantics (and the theory of meaning) be reconciled with generative linguistics?

The case for meaning involving language/world relations was put most vividly by Putnam (1975), who offered a number of thought experiments intended to show that the meanings of linguistic utterances (and, indeed, of tokens of mentalese, should there be any) are not determined solely by the internal psychological states of the speaker, but by those states along with the speaker's local environment and social milieu. Consider the example of an agent, Hilary, who has in his lexicon the words *elm* and *beech.* However, all Hilary knows about elms is that they are trees called 'elm,' and all he knows of beeches is that they are trees called 'beech.' He has no knowledge, linguistic or otherwise, that would allow him to identify elms or beeches, or to tell them apart; in particular, there is nothing in his concept *elm* (*beech*) that makes it true of all and only elms (beeches). (The fact that elms are called 'elm' while beeches are called 'beech' will not do: For this to work, Hillary would need to know what '*elm*' and '*beech*' pick out—but this, of course, is exactly what is in question. Kripke [1980] first made this point with respect to the meanings of proper names.) What, then, allows Hilary's uses of '*elm*' to pick out exactly those things that are, in fact, elms? (And likewise for his uses of '*beech.*') Putnam's answer is that the tokens get their reference, in part, from Hilary being a member of a linguistic community that contains experts who *can* identify elms and beeches. And it is these experts who determine the reference of '*elm*' and '*beech*'; Hilary's use of the words is normatively constrained by the expert's use—Hilary means by '*elm*' whatever the experts mean by '*elm.*' So the reference of Hilary's tokens of 'elm' and 'beech' is partly determined by knowledge about how to identify elms and beeches. But this knowledge is not in Hilary's head, it is distributed across Hilary's linguistic community.

Next, consider Oscar. Oscar is a normal English speaker, with the word '*water*' in his lexicon. Oscar's uses of '*water*' refer to the substance with the chemical formula H_2O (this being the substance that the relevant experts identify as water). Now somewhere in the universe, there is a planet, Twin-Earth, that is qualitatively very similar to Earth. In particular, it has "lakes," "rivers," and "oceans" filled with a colorless, odorless liquid that often falls from the sky and is necessary for the life on Twin-Earth. The intelligent inhabitants of Twin-Earth—twin-earthlings—drink large quantities of this liquid, wash their dishes in it, have bubbling fountains of it; in fact, they use it in all the ways Oscar and other earthlings use water. Despite its superficial resemblance to water, however, this substance does not have the chemical formula H_2O, but the formula XYZ. Among the twin-earthlings is Twin-Oscar, a normal Twin–English speaker, with the word '*water*' in his lexicon. Twin-Oscar and the other twin-earthlings use this word to refer to the liquid that plays the same role on Twin-Earth as water does on Earth. But this substance, on Twin-Earth, is not H_2O, but XYZ. So despite having qualitatively identical mental states—and perhaps even qualitatively very similar linguistic communities—Oscar's uses of '*water*' and Twin-Oscar's uses of '*water*' refer to different substances. What Oscar's and Twin-Oscar's uses of '*water*' are about depends crucially on their local environments.

So Hilary and an arborist, despite having different concepts of elms, refer to the same things by tokens of '*elm*'. And Oscar and Twin-Oscar, despite having the same concept associated with '*water*', refer to different things by tokens of that word. Thus, Putnam's examples apparently show that the meanings of words cannot depend entirely on the mental states of the speaker: Meaning depends crucially on facts external to the speaker. Call such an approach to meaning *referential semantics.*

Chomsky (e.g., 1995a and 2000) has launched numerous attacks on referential semantics. The initial worry is that referential semantics conflicts with the internalist, individualist nature of *I*-language. *I*-language is to be characterized in terms of nonrelational properties of the linguistic agent: It is a "narrow" property (or system) of the agent's mind/brain. Referential semantics, however, deals with relational properties: It is in the business of detailing relations between linguistic items and parts of the world external to the agent.

Chomsky admits that the naturalistic study of *I*-language is not restricted a priori to internalist investigations. The relations that bits of *I*-language (lexemes, phrases, and sentences, conceived of as elements of a mental computation system) bear to the extra-mental world could, *in principle*, be examined naturalistically. Chomsky believes, however, that these relations are, in fact, naturalistically intractable—the relations that linguistic items bear to the external world are asystematic enough that no serious, naturalistically acceptable explanatory theory can be given for them. Chomsky (1995) illustrates this point with the word *water*, which Putnam took to pick out all and only those things with the property of being H_2O (give or take impurities). Suppose Peter fills a cup out of the tap; it is a cup of water. But after a tea bag is dipped in it, it is no longer water—it is tea. Suppose further that, across town, Noam fills another cup from the tap; Noam's tap draws water from a reservoir into which a large quantity of tea leaves has been dumped as part of a purification process. Noam's cup contains water (albeit contaminated), even if what it contains is chemically indistinguishable from what is in Peter's cup. Thus, whether a substance counts as tea or water (containing tea only as an impurity) depends crucially on the particular interests and intentions of the speaker in the context. There simply is no single substance that infallibly serves as the reference of '*water*'.

Any object that could serve as a referent in referential semantics, argues Chomsky, is bound to be so gerrymandered and ill-behaved that it could not plausibly be a real constituent of the world. Is it to be believed that there is some single substance, an "object" in the world, that is what is in Noam's cup—as well as the River Thames, the Pacific Ocean, and bottles of Evian—but not what is in Peter's cup—nor bottles of Windex or cans of Coca Cola? And the complexity that attends '*water*' is entirely typical of natural language. Consider that, during World War II, Dresden was burned to the ground; but it was rebuilt, and it deserves to be proud of its many new buildings. Or that a bank, which raised its interest rates because it hoped to bring in new business, might be destroyed in an earthquake and have to move across the street. What are these things, picked out by '*Dresden*' and '*bank*', that are at once concrete and abstract, apparently intentional, and can survive destruction? Consider, too, that there is likely a flaw somewhere in this paper, or that the average family has 2.3 children. In the simplest referential picture, '*flaw*' refers to flaws and '*the average family*' refers to the average family. But, surely, flaws and average families are naturalistically dubious; it is difficult to imagine a naturalistic inquiry into the nature of flaws. Of course, the surface form of sentences containing '*flaw*' and '*average family*' might be misleading; at the level of logical form, these linguistic items might not be referential—they could instead be adjectival modifiers or adverbials. But natural language is replete with apparent reference to naturalistically suspicious entities; to pass naturalistic muster, referential semantics must contend with each and every case.

Chomsky concludes that there is no relation of reference holding between linguistic items and objects in the world, at least not one about which anything interesting and general can be said. Any account of reference must ultimately advert to intentionality, a subject forever out of reach of naturalistic inquiry (see Intentionality).

Ludlow (2003) mounts a defense of referential semantics for *I*-languages. He suggests that semanticists bite the metaphysical bullet and accept that—perhaps in addition to the substances, objects, and properties catalogued by physical science—there are things exactly like those needed for referential semantics. There are cities and banks that survive destruction and act intentionally; there is a substance, water, the nature of which depends sensitively on its origin and the uses to which it is put; there are, perhaps, even such things as flaws and average families. In this view, metaphysical intuitions—about, say, whether or not a particular substance in a particular context is water—are underwritten by the structure of *I*-language. And if metaphysical intuitions—at least of the sort probed in Putnam-like examples—are by and large correct, semanticists and philosophers are in a position to reason from the structure of *I*-language to the structure of the world and vice versa.

This Kantian view of the nature of referential semantics presents a dilemma, however. If the semanticist insists that the entities and substances invoked as referents really exist, but agrees with Chomsky that such things are not fit for naturalistic study, then referential semantics is nonnaturalistic. It then lies outside the *scientific* study of language and is instead a philosophical epicycle on naturalistic linguistics. If, on the other hand, the entities and substances of referential semantics are natural objects, it is a serious question as to what place they hold in the vast array of objects posited by the other sciences. How, for example, is the referent of '*water*', water, related to atoms and molecules (objects of chemistry and physics) and to the representational/computational systems of the brain (objects of linguistics, psychology, and neurobiology)? (If Chomsky is correct, any adequate answer will be more complicated than "The referent of '*water*' is identical with H_2O.") This seems tantamount to asking questions about the relation between *I*-language and the world that Chomsky thinks is naturalistically legitimate but also naturalistically intractable.

Questions About Rules and Representations

The idea that linguistic theory involves the investigation of rules and representations (or principles and parameters) of an internal computational system has also led to philosophical questions about the nature of these rules and representations. For example, Quine (1970) has argued that, since many possible grammars may successfully describe an agent's linguistic behavior, there is no way in principle to determine which grammar an agent is using (see Quine, Willard Van). For his part, Chomsky (1980) has argued that, if one considers the *explanatory adequacy* of a grammar in addition to its *descriptive adequacy*, then the question of which grammar is correct is answerable in principle. That is, because the theory of grammar must be consistent with the theory of language acquisition, acquired language deficits, and, more generally, cognitive psychology, then there are many constraints available to rule out competing grammatical theories. To illustrate, two descriptively adequate theories of tense may differ in their assumptions about whether tenses like past and future are more basic, or whether a grasp of terms of temporal order (like 'before' and 'after') are more basic. Acquisition data might shed light on such a standoff if it could be shown that children acquire the use of one set of linguistic items markedly before the other.

Another set of worries about rule following have stemmed from Kripke's (1982) reconstruction of arguments in Wittgenstein (1953 and 1956). The idea is that there can be no brute fact about what rules and representations a system is using apart from the intentions of the designer of the system. Since, when studying humans, there is no access to the intentions of the designer, there can be no fact of the matter about what rules and representations underlie linguistic abilities. The conclusion drawn by Kripke (1982) is that "it would seem that the use of the idea of rules and of competence in linguistics needs serious reconsideration, even if these notions are not rendered meaningless" (1983, 31n, 22).

Chomsky (1986) appears to argue that one can know certain facts about computers in isolation, but Chomsky's current position (1995) is that computers, unlike the human language faculty, are artifacts and hence the product of human intentions. The language faculty is a natural object and embedded within human biology, so the facts about its structure are no more grounded in human intentions than are facts about the structure of human biology.

Kripke's argument is often stated in the form of a problem about justification: Speakers believe they know what they mean by their words; what justifies that belief? Chomsky (1986) responds to this worry by observing that, in the syntactic realm, there is no reason to suppose that speakers have first-person authority with respect to the rules they follow. Speakers produce grammatical sentences effortlessly, but the procedures they use in producing them are highly abstract and can remain obscure even under careful scrutiny. Justification is not available, or necessary, for successful linguistic communication. Scientific inquiry into the nature of those rules, by contrast, is no more or less justified than inquiry into the guiding principles of any system whose operation cannot be directly observed.

Such considerations address the epistemic side of Kripke's argument, but they do not address the metaphysical problem: What is it about someone that makes him or her a follower of rule *R*? One possible response may come from Chomsky's (1965) suggestion that the object of inquiry for syntactic theorizing is "an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge in actual performance" (3). Such an idealized speaker could in principle produce evidence capable of distinguishing between any two substantively different rule systems. The apparent reliance on the relevant problematic notions ("grammatical," "ungrammatical") in linguistic theory might thus be seen as an artifact of considerations that are appropriately idealized away. This line of response may, however, be susceptible to Kripke's critique of solutions that depend on *ceteris paribus* clauses. In that critique, Kripke argues that there is no non-question-begging way to decide which features of the system should be idealized away from and which should not.

Perhaps a more promising solution can be found in an argument, due to Soames (1998), that an inability to know what fact about someone makes that person a follower of rule *R* in no way undermines the metaphysical possibility that there is such a brute fact. Soames suggests that Kripke, of all people, should have seen the error here—a conflation of metaphysical and epistemic possibility.

Methodological Issues

If the language faculty is an internal computational/representational system, a number of questions arise about how to best go about investigating and describing it. For example, there has been considerable attention paid to the role of formal rigor in linguistic theory. On this score, a number of theorists (e.g., Gazdar et al. 1985; Bresnan and Kaplan 1982; Pullum 1989) have argued that the formal rigor of their approaches—in particular their use of well-defined recursive procedures—count in their favor. However, Ludlow (1992) has argued that this sort of approach to rigorization would be out of synch with the development of other sciences (and indeed, of branches of mathematics), where formalization follows in the wake of the advancing theory.

Another methodological issue concerns the nature of evidence available to investigations of the language faculty. For generative linguists, evidence from a written or spoken corpus is at best twice removed from the actual object of investigation, and given the possibility of performance errors, is notoriously unreliable at that. Much of the evidence adduced in linguistic theory has therefore been from speakers' intuitions of acceptability, as well as intuitions about possible interpretations. This raises a number of interesting questions about the reliability of introspective data and the kind of training required to have reliable judgments. There is also the question of why one should have introspective access to the language faculty at all. It is fair to say that these questions have not been adequately explored to date (except in a critical vein; see Devitt 1995; Devitt and Sterelny 1987).

A third methodological issue relates to the use of parsimony and simplicity in the choice between linguistic theories (see Parsimony). While tight definitions of simplicity within a linguistic theory seem to be possible (see Halle 1961; Chomsky and Halle 1968; Chomsky 1975), finding a notion of simplicity that allows one to choose between two competing theoretical frameworks is another matter. Some writers (e.g., Postal 1972; Hornstein 1995) have argued that generative semantics and the most recent version of generative linguistics, minimalism (discussed in the next section), are simpler than their immediate competitors because they admit fewer levels of representation. In response, Ludlow (1998) has maintained that there is no objective criterion for evaluating the relative amount of theoretical machinery across linguistic theories. Ludlow offers that the only plausible definition of simplicity would be one that appealed to "simplicity of use," suggesting that simplicity in linguistics may not be a feature of the object of study itself, but rather an ability to easily grasp and utilize certain kinds of theories.

An alternative approach would be to take a leaf from Sober (1975) and argue for a view of simplicity according to which a theory is simpler than another if it is more easily embedded within more basic sciences. In such a view the simpler linguistic theory would be the one that could more naturally be embedded into cognitive psychology or even, following recent work in the minimalist program, into low-level biophysical and mathematical principles.

Issues Raised by the Minimalist Program

Although generative linguistics has gone through a number of permutations over the last 50 years, perhaps the most interesting has been the recent development of the minimalist program, outlined in Chomsky (1995a). Setting aside the technical details of the project, the headline idea is that the core language faculty did not evolve slowly over an extended period of time but was, rather, the result of a sudden mutation that, in effect, wired together two discrete cognitive systems—the conceptual/intentional (C/I), involving meaning and thought; and the perceptual/articulatory (P/A), responsible for speech production and perception. The working hypothesis is that the wiring solution was "optimal" and governed by basic low-level biological and mathematical constraints (such as those that account for the prevalence of recursive and fractal patterns in nature). This speculative hypothesis, if correct, would suggest that the linguistic theory should be looking for very specific kinds of properties and principles linking the C/I and P/A systems—properties that might naturally emerge from low-level biophysical principles (much as the Fibonacci pattern in a sunflower does). While speculative in the extreme, this new research program has shown some surprising successes and clearly calls for further investigation of its basic guiding assumptions. Although the project has yet to be explored in a formal way by philosophers of linguistics, it should prove a fascinating domain for future investigations.

Conclusion

Although the philosophy of linguistics is not as well explored as the philosophy of physics or of biology, it is certainly no less rich a domain of inquiry. Not only does it involve the usual concerns of scientific methodology (simplicity, the nature and trustworthiness of the data, etc.), but it also deals with kinds of entities (rules and representations) that are not routinely found in the basic sciences and are not well understood. Furthermore, the philosophy of linguistics is concerned with questions of the embeddability of linguistics into more basic sciences (possibly even into low-level biophysical and mathematical systems), as well as which parts of linguistics (if any) involve agent/environment relations and which parts are purely individualistic. For all these reasons, this subdiscipline of the philosophy of science promises to be a fertile area of investigation, plausibly able to illuminate some of the deeper cognate questions being explored in the philosophy of other sciences, as well as the philosophy of science generally.

Joshua Brown
Peter Ludlow
Tim Sundell

References

Bresnan, J., and R. Kaplan (1982), "Introduction: Grammars as Mental Representations of Language," in Bresnan (ed.), *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press, xvii–lii.

Chomsky, N. (1965), *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

——— (1975), *The Logical Structure of Linguistic Theory*. New York: Plenum.

——— (1980), *Rules and Representations*. New York: Columbia University Press.

——— (1986), *Knowledge of Language*. New York: Praeger.

——— (1993), "Explaining Language Use," in J. Tomberlin (ed.), *Philosophical Topics 20*, 205–231.

——— (1995a), "Language and Nature," *Mind* 104: 1–61.

——— (1995b), *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N., and M. Halle (1968), *The Sound Pattern of English*. New York: Harper and Row.

Devitt, M. (1995), *Coming to Our Senses: A Naturalistic Program for Semantic Localism*. Cambridge: Cambridge University Press.

Devitt, M., and K. Sterelny (1987), *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge, MA: MIT Press.

Gazdar, G., E. Klein, G. Pullum, and I. Sag (1985), *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard UP.

George, A. (1989), "How Not to Become Confused about Linguistics," in George (ed.), *Reflections on Chomsky*. Oxford: Basil Blackwell, 90–110.

Halle, M. (1961), "On the Role of Simplicity in Linguistic Description," in *Proceedings of Symposia in Applied Mathematics* 12 (Structure of Language and Its Mathematical Aspects). Providence, RI: American Mathematical Society, 89–94.

Higginbotham, J. (1983), "Is Grammar Psychological?" in L. Cauman, I. Levi, C. Parsons, and R. Schwartz (eds.), *How Many Questions: Essays in Honor of Sydney Morgenbesser*. Indianapolis: Hackett.

Hornstein, N. (1984), *Logic as Grammar*. Cambridge, MA: MIT Press.

——— (1995), *Logical Form: From GB to Minimalism*. Oxford: Blackwell.

Katz, J. (ed.) (1985), *The Philosophy of Linguistics*. Oxford: Oxford University Press.

——— (1981), *Language and Other Abstract Objects*. Totowa, NJ: Rowman and Littlefield.

Kripke, S. (1980), *Naming and Necessity*. Cambridge, MA: Harvard University Press.

——— (1982), *Wittgenstein on Rules and Private Language*. Cambridge: Harvard University Press.

Lewis, D. (1975), "Language and Languages," in K. Gunderson (ed.), *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press, 3–35.

Ludlow, P. (forthcoming), "Referential Semantics for *I*-Languages?" in N. Hornstein and L. Antony (eds.), *Chomsky and His Critics*. Oxford: Blackwell.

——— (1998), "Simplicity and Generative Grammar," in R. Stainton and K. Murasugi (eds.), *Philosophy and Linguistics*. Boulder, CO: Westview Press.

——— (1992), "Formal Rigor and Linguistic Theory," *Natural Language and Linguistic Theory* 10: 335–344.

Pinker, S. (1994), *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Company.

Postal, P. (1972), "The Best Theory," in S. Peters (ed.), *Goals of Linguistic Theory*. Englewood Cliffs, NJ: Prentice-Hall, 131–179.

Pullum, Geoffrey (1989), "Formal Linguistics Meets the Boojum," *Natural Language and Linguistic Theory* 7: 137–143.

Putnam, H. (1975), "The Meaning of Meaning," in Gunderson (ed.), *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science* (Vol. 7). Minneapolis: University of Minnesota Press, 131–193.

Quine, Willard Van (1970), "Methodological Reflections on Current Linguistic Theory," *Synthese* 21: 368–398.

Soames, S. (1998), "Skepticism about Meaning: Indeterminacy, Normativity, and the Rule-Following Paradox," in A. Kazmi (ed.), *Meaning and Reference. Canadian Journal of Philosophy* 23(Suppl).

Sober, E. (1975), *Simplicity*. Oxford: Oxford University Press.

Wittgenstein, L. (1953), *Philosophical Investigations*. Translated by G. E. M. Anscombe. New York: Macmillan.

——— (1956), *Remarks on the Foundations of Mathematics*. Translated by G. E. M. Anscombe. Cambridge, MA: MIT Press.

See also **Chomsky, Noam; Innate-Acquired Distinction; Intentionality; Quine, Willard Van; Social Sciences, Philosophy of**

LEVELS OF SELECTION

See **Biology, Philosophy of; Natural Selection**

LOCALITY

The principle of local action has played an important role in the development of modern physics, and has been taken by many philosophers to be a necessary condition for intelligible causal explanations. However, recent evidence from quantum mechanics and quantum field theory seems to point toward fundamental limitations on the ability to provide locally causal explanations of physical phenomena (see Quantum Mechanics; Quantum Field Theory).

Locality in the History of Philosophy

According to a popular view, the most primitive cause/effect relation is that which holds between two physical objects that make contact in space and time, known as *contact action.* Furthermore, it is supposed that between any cause and effect, there must be a continuous chain of primitive causes by contact action; and if there is no continuous chain in space and time between two events, neither can be a cause of the other. This view has been advocated in one form or another by a number of philosophers of diverse persuasions. For example, Aristotle claims that "it is evident, therefore, that in all locomotion there is nothing intermediate between mover and moved" (Aristotle 1941). Similarly, in establishing the foundations for his new physics, Descartes takes it as an a priori principle that causation occurs only by local contact (see Suppes 1954). Moreover, Hume (1978)—the arch-critic of a priori knowledge of causal relations—claims that the concept of causation includes the concept of contiguity:

> [W]hatever objects are consider'd as causes or effects, are *contiguous*; and ... nothing can operate in a time or a place, which is the ever so little remov'd from those of its existence. (§I.3.2)

Einstein (1948) claims that the principle of local action is a necessary presupposition for the existence of empirically testable natural laws:

> For the relative independence of spatially distinct things (A and B), this idea is characteristic: an external influence on A has no immediate effect on B; this is known as the "principle of local action" ... . The complete suspension of this basic principle would make impossible ... the establishment of empirically testable laws in the sense familiar to us. (321)

Finally, a number of influential contemporary accounts of causation tie the notion of causal connectedness to a space-time picture (see, e.g., Salmon 1984).

Locality in Modern Physics

The principle of local action was a cornerstone of the mechanical philosophy of Cartesian and neo-Cartesian physics. However, in Newton's theory of gravitation, the inverse square law seems to entail that some causes are spatially separated from their effects. Although numerous attempts, both physical and philosophical, were made to explain nonlocal gravitational forces (see McMullin 1989), a satisfactory resolution was not reached until Einstein supplied a field-theoretic formulation of gravity. Einstein's general theory of relativity was the culmination of a line of development that had begun in the early nineteenth century with Michael Faraday's introduction of the concept of "lines of force" emanating from a magnet. Faraday's idea was incorporated into Maxwell's dynamical theory of the electromagnetic field, in which electromagnetic field quantities are associated with each point of space, and disturbances in the field propagate through space via wave motion. Problems arising from Maxwell's theory ultimately led to Einstein's special theory of relativity, which grounds the principle of local action in the assumption of the constancy of the speed of light in all reference frames.

According to textbook presentations, special relativity is based on the limit principle: No physical process can propagate faster than the speed of light. If two events cannot be connected by a light signal, then they are said to be *spacelike separated* (i.e., they are simultaneous in some inertial reference frame). Thus, there can be no cause/effect relation between two spacelike separated events. However, the status of the principle of locality in special relativity continues to be a subject of dispute among philosophers. For example, it has been claimed that special relativity is not premised on the limit principle and probably does not entail it (Nerlich 1982). It has also been claimed that the limit principle is a statistical generalization that need not hold for individual processes (Cushing 1996; Maudlin 1994 provides an extended discussion of the role of locality in special relativity) (see Space-Time).

Entanglement and the Einstein–Podolsky–Rosen Result

In quantum mechanics, a pair of spatially separated systems can occupy an "entangled" state in which the values of their dynamical variables are perfectly correlated. In particular, if S_1 and S_2 are spatially separated systems with state spaces H_1 and H_2, then the state space for the composite system $S_1 + S_2$ is the tensor product $H_1 \otimes H_2$. For any $u_2 \in H_1$ and $v_2 \in H_2$, there is a vector $u \otimes v \in H_1 \otimes H_2$ called the "product" of u and v. Product states can be thought of as describing conjunctive states of affairs: $S_1 + S_2$ is in state $u \otimes v$ just in case S_1 is in state u and S_2 is in state v. However, since $S_1 + S_2$ is a quantum system, it also has states that are superpositions of product states. In particular, if u_1, u_2 are distinct states of S_1 and v_1, v_2 are distinct states of S_2, then

$$\psi = \frac{1}{\sqrt{2}}(u_1 \otimes v_1) + \frac{1}{\sqrt{2}}(u_2 \otimes v_2),$$

is a state of $S_1 + S_2$ that cannot be decomposed as a simple product. Such states are said to be *entangled.*

Composite systems in classical physics also have correlated states. But unlike the classical case, an entangled quantum state cannot be taken to represent an ensemble of composite systems each of which is in some definite product state. Indeed, unlike any correlated state in classical physics, entangled states are "pure" (i.e., statistically irreducible) states of the composite system. However, entangled states look mixed to local observers at S_1 or S_2. In fact, there is no pure (vector) state v of S_1 that agrees with ψ on the probabilities assigned to the various propositions about S_1, and similarly for S_2. In general, when $S_1 + S_2$ is in an entangled state, it is impossible to think of S_1 and S_2 as having their own (pure) quantum states.

The entangled state ψ predicts perfect correlations between measurements on the component systems. In particular, there is a measurement M_1 on S_1 that discriminates between the states u_1 and u_2, and there is a measurement M_2 on S_2 that discriminates between the states v_1 and v_2. If $S_1 + S_2$ is in the state ψ, then the two outcomes of M_1 are equally likely, and the two outcomes of M_2 are equally likely. However, outcomes of M_1 and M_2 are perfectly correlated: If M_1 yields an outcome corresponding to u_1, then M_2 will yield an outcome corresponding to v_1; and if M_1 yields an outcome corresponding to u_2, then M_2 will yield an outcome corresponding to v_2.

In their argument against the completeness of quantum mechanics, Einstein, Podolsky, and Rosen [EPR] (1935) made use of an entangled state that predicts perfect correlations between both the positions and the momenta of a pair of particles. They note that if the position of the first particle is ascertained, then the position of the second particle can be predicted with certainty. Similarly, if the momentum of the first particle is ascertained, then the momentum of the second particle can be predicted with certainty. Now, if one assumes (as EPR did) that the principle of local action holds, then a measurement on the first particle can neither alter nor bring into being properties of the second particle. Thus, a *position* measurement on the first particle should be thought of as a means of discovering the preexisting position of the second particle; and a *momentum* measurement on the first particle should be thought of as a means of discovering the preexisting momentum of the second particle. But then the second particle must have had a definite position and momentum before any measurement was performed. Since, however, a quantum-mechanical state never assigns a definite position and momentum to any object, EPR concluded that the quantum-mechanical state does not provide a complete description of the properties of the second particle.

EPR claimed to have shown that each particle must have a "hidden" state that determines the values of all of its dynamical variables—in other words, there are hidden variables. EPR also assumed that the hidden state of one system cannot be instantaneously influenced by events in distant locations. Nonetheless, neither EPR nor any other similarly inclined physicists were able to find a hidden variable theory that obeys the principle of local action. This failure, it is now known, was inevitable: Bell's theorem (Bell 1964) shows that no local hidden variable theory can explain the correlations described in the EPR experiment.

Bell's Theorem

According to hidden variable theories, quantum states merely provide statistical information about ensembles of systems, each of which has its own definite state (which includes a specification of the values of all relevant dynamical variables). A local hidden variable theory attributes a definite state to each local system and requires that changes in the state of one system cannot instantaneously bring about changes in the state of a distant system. Bell's theorem shows that no such local hidden variable theory can reproduce the predictions of quantum mechanics.

In the thirty years prior to the proof of Bell's theorem, the question of hidden variables had been largely pushed aside, in particular since von Neumann (1932) had supposedly shown that hidden variables are inconsistent with the empirical predictions of quantum mechanics. Few physicists took notice at the time when, in 1952, David Bohm constructed a hidden variable theory that is empirically equivalent to quantum mechanics. As Bell (1982) points out, Bohm's theory is not ruled out by the (mathematically valid) no-go theorems of von Neumann and Kochen-Specker because Bohm's hidden variables are contextual; that is, the hidden state of a system cannot be specified without taking into account its context, including the setting of measurement devices in distant locations. In fact, Bohm's hidden variables are patently nonlocal. Bell's theorem shows that this feature of Bohm's theory holds for any hidden variable theory that reproduces the predictions of quantum mechanics.

Consider the most simple correlation experiment, in which there is a pair of measurement devices situated in distant wings of a laboratory and each measurement device has (at least) two distinct settings. Let L_a denote the event that the left device is in setting a, and let R_b denote the event that the right device is in setting b. Suppose that each experiment has two possible outcomes, denoted by $-$ and $+$. Let $L_a^{\pm}$ denote the event that the device on the left registers a $\pm$ outcome when in setting a, and let $R_b^{\pm}$ denote the event that the device on the right registers a $\pm$ outcome when in setting b. A quantum-mechanical realization of such an experiment is given by a pair of spin-$\frac{1}{2}$ particles in the singlet state:

$$\psi = \frac{1}{\sqrt{2}}(x_1 \otimes y_1 - x_2 \otimes y_2). \tag{1}$$

The measuring devices can be taken to be a pair of Stern-Gerlach magnets, each of which can be oriented at various angles in a plane from a common (arbitrarily chosen) axis. For each possible measurement on each particle, there are two possible outcomes, spin up and spin down. In this case, quantum mechanics supplies the following probabilities:

$$P_{\mathrm{QM}}(L_a^+) = P_{\mathrm{QM}}(R_b^+) = \frac{1}{2}, \tag{2}$$

and

$$P_{\mathrm{QM}}(L_a^+ \cap R_b^+) = \frac{1}{2}\cos^2\theta_{ab}, \tag{3}$$

where θ_{ab} is the difference between the angles of orientation of the magnets on the left and right. If θ_{ab} is not an integer multiple of $\pi/4$, then the left and right measurement outcomes are statistically correlated:

$$P_{\mathrm{QM}}(L_a^+ \cap R_b^+) \neq P_{\mathrm{QM}}(L_a^+) \times P_{\mathrm{QM}}(R+b). \tag{4}$$

Of course, the existence of a correlation between spatially separated events does not necessarily indicate a nonlocal connection, because the two events might have a common cause in the intersection of their past light cones. Suppose then that the quantum state corresponds to a probability distribution P_{HV} over a space Λ of hidden variables. Suppose for simplicity that Λ is finite. Suppose also that the domain of the probability function P_{HV} includes the events L_a, R_b^+ etc.). Let

$$P_{ab} = \sum_{\lambda \in \Lambda} [P_{\mathrm{HV}}(L_a^+ \cap R_b^+ \mid L_a \cap R_b \cap \lambda) \times P_{\mathrm{HV}}(\lambda)], \tag{5}$$

for all a, b, and let

$$P_1 = \sum_{\lambda \in \Lambda} [P_{\mathrm{HV}}(L_1^+ \mid L_1 \cap \lambda) \times P_{\mathrm{HV}}(\lambda)], \tag{6}$$

$$P_3 = \sum_{\lambda \in \Lambda} [P_{\mathrm{HV}}(R_3^+ \mid R_3 \cap \lambda) \times P_{\mathrm{HV}}(\lambda)]. \quad (7)$$

Thus, this hidden variable model reproduces the quantum mechanical probabilities in state ψ just in case $P_a = P_{\mathrm{QM}}(L_a^+)$, $P_b = P_{\mathrm{QM}}(R_b^+)$, and $P_{ab} = P_{\mathrm{QM}}(L_a^+ \cap R_b^+)$.

The hidden variable λ is local just in case it determines the outcomes of measurements on S_1 independently of what is occurring at S_2, and vice versa. That is, once the value of the hidden variable λ and the setting of the measurement apparatus at S_1 is fixed, then the outcomes of measurements on S_1 are determined; and similarly for S_2. (In the more general case of stochastic hidden variables, λ and the setting at S_1 will fix the probabilities for various outcomes at S_1.) This assumption is captured succinctly by Bell's locality condition:

$$\begin{aligned} &P_{\mathrm{HV}}(L_a^+ \cap R_b^+ \mid L_a \cap R_b^+ \cap \lambda) \\ &= P_{\mathrm{HV}}(L_a^+ \mid L_a \cap \lambda) \times P_{\mathrm{HV}}(R_b^+ \mid R_b \cap \lambda)]. \quad (8) \end{aligned}$$

If Bell's locality condition is conjoined with a "no conspiracy" condition (*viz.,* the event that a certain measurement occurs is probabilistically independent of λ), then Bell's inequality follows:

$$0 \le P_1 + P_3 + P_{24} - P_{14} - P_{23} - P_{13} \le 1. \quad (9)$$

Thus, Bell's inequality is satisfied by the statistical predictions of any "reasonable" local hidden variable model of this experiment.

For specific choices of angles for the measurement devices, the quantum mechanical predictions violate Bell's inequality. For example, for the settings $\theta_{24} = \pi/2$, $\theta_{13} = \theta_{14} = \theta_{23} = \pi/6$, the sum of the quantum-mechanical probabilities equals $-\frac{1}{8}$. Thus, the predictions of quantum mechanics cannot be reproduced by any local hidden variable model. Moreover, these predictions have now been verified in a number of different experiments (for a review, see Redhead 1994, 107ff). Thus, the phenomena cannot be explained by a local hidden variable model.

Interpretations of Bell's Theorem

Many philosophers and physicists think that the violation of Bell's inequality points toward some form of nonlocality, whether or not quantum mechanics is a complete theory. However, a small group of dissenters claim that the violation of Bell's inequality has nothing to do with locality but should be seen as a consequence of the use by quantum mechanics of a nonclassical probability theory. Moreover, even among those who think that the violation of Bell's inequality entails nonlocality, there is still widespread disagreement about what exactly this means. Some claim that the violation of Bell's inequality shows that the world is thoroughly interconnected and holistic (or "nonseparable"), but not that the principle of local action is false. Others claim that the violation of Bell's inequality shows that causes can be spatially separated from their effects. The following section examines arguments for these three positions.

Quantum Mechanics Is Local

In a minimalist interpretation of Bell's theorem, the violation of Bell's inequality is due to the fact that local systems have incompatible observables, that is, observables for which there are no joint probabilities. The primary support for this interpretation comes from a theorem by Arthur Fine (1982a and 1982b) that shows that Bell's inequality is satisfied if and only if all joint probabilities are well defined. More precisely, suppose that there are joint probabilities $P(L_a^+, L_b^\pm)$ and $P(R_c^+, R_d^\pm)$ that return the already-given marginal probabilities:

$$P_{\mathrm{QM}}(L_a^+) = P(L_a^+, L_b^+) + P(L_a^+, L_b^-), \quad (10)$$

$$P_{\mathrm{QM}}(R_c^+) = P(L_c^+, L_d^+) + P(R_c^+, R_d^-). \quad (11)$$

Note that these joint probabilities are not supplied by quantum mechanics.

If such joint probabilities exist, then the marginal probabilities must satisfy Bell's inequality (de Muynck 1986). The minimalist will then point out that this derivation of Bell's inequality does not use any locality condition, and so the violation of Bell's inequality does not entail nonlocality. Contrapositively, the minimalist claims that since the existence of joint probabilities entails Bell's inequality, the violation of Bell's inequality shows that joint probabilities do not exist. The minimalist interpretation of Bell's theorem has also been defended within the context of particular interpretations of quantum mechanics. For example, advocates of the *consistent histories* interpretation have argued that "locality is not the only assumption that goes into the proof of the Bell inequalities, and thus their violation by quantum theory is not a proof of nonlocality" (Brun and Griffiths 2000). Furthermore, it has recently been claimed that if quantum mechanics is approached from an information-theoretic perspective, then the theory is "essentially local" (Fuchs and Peres 2000).

However, the minimalist interpretation of Bell's theorem has been criticized on the grounds that it ignores the issue of contextuality (van Fraassen 1991, 102; Shimony 1993, II.9). In particular, in a contextual hidden variable theory, unconditional probabilities such as $P_{\mathrm{HV}}(L_a^+ \mid \lambda)$ are not physically significant. Rather, the physically significant probabilities are those conditionalized on all relevant measurement settings, for example, $P_{\mathrm{HV}}(L_a^+ \mid L_a \cap R_b \cap \lambda)$.

However, Bell's inequality cannot be derived for the latter conditional probabilities. In other words, the existence of joint conditional probabilities does not entail Bell's inequality.

Holism and Nonseparability

Some philosophers have argued that the violation of Bell's inequality may entail nonlocality but it does not entail that there is superluminal causation. The main supporting argument for this position draws on Jarrett's (1984) analysis of Bell's locality condition. Jarrett shows that Bell's locality condition is equivalent to the conjunction of two conditions (the labels here are due to Shimony):

1. Outcome Independence:

$$\begin{aligned} P_{\mathrm{HV}}(L_a^+ \cap R_b^+ \mid L_a \cap R_b \cap \lambda) \\ = P_{\mathrm{HV}}(L_a^+ \mid L_a \cap R_b \cap \lambda) \\ \times P_{\mathrm{HV}}(R_b^+ \mid L_a \cap R_b \cap \lambda). \end{aligned} \tag{12}$$

2. Parameter Independence:

$$P_{\mathrm{HV}}(L_a^+ \mid L_a \cap R_b \cap \lambda) = P_{\mathrm{HV}}(L_a^+ \mid L_a \cap \lambda), \tag{13}$$

$$P_{\mathrm{HV}}(R_b^+ \mid L_a \cap R_b \cap \lambda) = P_{\mathrm{HV}}(R_b^+ \mid R_b \cap \lambda). \tag{14}$$

According to the orthodox interpretation of quantum mechanics, the quantum state ψ gives maximal information about the system. Thus, the orthodox interpretation can be thought of as the trivial hidden variable theory in which λ supplies no information beyond that supplied by the quantum state. Since the probabilities assigned by the quantum state are insensitive to which measurements are being performed on distant systems, the orthodox interpretation satisfies parameter independence. On the other hand, since distant measurement outcomes are correlated (see equation 4), outcome independence is violated. Moreover, it has been claimed that since outcomes are not determined by the quantum state—and hence cannot be controlled—this nonlocality could not be exploited to send a signal faster than the speed of light. Thus, orthodox quantum mechanics is consistent with special relativity (Shimony 1993, II.10). Proponents of the orthodox interpretation have also claimed that hidden variable theories violate parameter independence and that such a violation allows for superluminal signaling, and therefore hidden variable theories are inconsistent with special relativity. However, advocates of hidden variables have replied by pointing out that superluminal signaling is possible only if the hidden variables could be controlled, and this is not generally possible (e.g., in Bohm's theory).

Action at a Distance

Bell's inequality follows from the conjunction of outcome independence and parameter independence. So, either outcome independence or parameter independence (or both) is false. Bell's theorem by itself says no more. Nonetheless, there have been many attempts over the years to narrow down the interpretive options by deriving Bell's inequality from one of Jarrett's conditions. On the one hand, some argue that Bell's inequality can be derived from the assumption of hidden variables (*viz.,* the existence of joint probabilities), and therefore its violation supplies good evidence against "realism." Others, however, argue that Bell's inequality follows from locality alone, and so its violation entails nonlocality.

Maudlin (1994) argues that quantum-mechanical correlations can be explained only on the supposition of nonlocal causes. For example, he parodies Jarrett's analysis of Bell's locality condition by showing that it is equivalent to the conjunction of two conditions:

$$P_{\mathrm{HV}}(L_a^+ \mid L_a \cap R_b^+ \cap \lambda) = P_{\mathrm{HV}}(L_a^+ \mid L_a \cap \lambda), \tag{15}$$

$$P_{\mathrm{HV}}(R_b^+ \mid R_b \cap L_a^+ \cap \lambda) = P_{\mathrm{HV}}(R_b^+ \mid R_b \cap \lambda), \tag{16}$$

and,

$$P_{\mathrm{HV}}(L_a^+ \mid L_a \cap R_b \cap R_b^+ \cap \lambda) = P_{\mathrm{HV}}(L_a^+ \mid L_a \cap R_b^+ \cap \lambda), \tag{17}$$

$$P_{\mathrm{HV}}(R_b^+ \mid L_a \cap R_b \cap L_a^+ \cap \lambda) = P_{\mathrm{HV}}(R_b^+ \mid R_b \cap L_a^+ \cap \lambda). \tag{18}$$

Maudlin (1994) then points out that that it would be appropriate to call the first condition "outcome independence" and the second condition "parameter independence" (95). But now orthodox quantum mechanics violates "parameter independence" but not "outcome independence." The conclusion that should be drawn, claims Maudlin, is that Jarrett's analysis does nothing to show that the nonlocality found in orthodox quantum mechanics is more benign than the nonlocality found in hidden variable theories. (For another argument that quantum mechanics by itself—i.e., without additional interpretive assumptions—entails nonlocality, see Stapp 1997.)

The Subtleties of Nonlocality

While philosophers have been mainly concerned with investigating the consequences of the violation of Bell's inequality, physicists have also been trying to find ways to use nonlocality as a physical resource (e.g., to speed up computation). In the course of these investigations, it has been discovered that the violation of Bell's inequality is just one of many manifestations of nonlocality in quantum mechanics. Each vector state for a composite system is either a product state, or it is entangled. If a vector state is entangled, then it violates Bell's inequality (Gisin and Peres 1992), and therefore its correlations cannot be reproduced by a local hidden variable model. More generally, an arbitrary (possibly mixed) state ρ of a composite system is said to be separable just in case it is a mixture of product vector states; otherwise it is said to be nonseparable.

It is not difficult to see that separable states satisfy Bell's inequality. (Indeed, the "hidden variables" can be taken to be the quantum product states that are mixed together to form the separable state.) Werner (1989), however, shows that not all nonseparable states violate Bell's inequality. In particular, consider the mixture

$$W_n = \frac{1}{2}M_n + \frac{1}{2}P_s, \tag{19}$$

where $M_n = (1/n^2)(I \otimes I)$ is the maximally mixed state of $\mathbf{C}^n \otimes \mathbf{C}^n$, and P_s is any maximally entangled, symmetric, pure state of $\mathbf{C}^n \otimes \mathbf{C}^n$. (For example, when $n = 2$, P_s could be the singlet state.) Werner uses an ingenious argument to show that W_n is nonseparable. However, he then goes on to construct a local hidden variable model for the correlations of W_n in Bell-type experiments. So, is W_n local or nonlocal?

There are at least two good reasons for thinking that the Werner state is nonlocal. First, a local hidden variable model, in Bell's sense, need account for the outcomes of only a certain special class of measurements; there might be other measurements whose statistics in W_n cannot be reproduced by such a model. In fact, Popescu (1995) shows that (for $n \geq 5$) a local observer can select a subensemble from W_n that violates Bell's inequality. That is, after an initial preparatory measurement on W_n is performed, then a Bell-type measurement can be performed that yields manifestly nonlocal results. Since the initial preparatory measurement is purely local, it cannot create entanglement where none already existed. Therefore, it seems plausible to say that the original state W_n was already nonlocal (although its nonlocality was "hidden").

Second, the Werner state permits a teleportation scheme with higher fidelity than any classical communication channel (see Popescu 1994). Suppose that an observer O_1 has a particle P_1 in some unknown quantum state ψ, and O_1 wants to supply enough information to a second observer O_2 so that O_2 can prepare a particle in an identical state. On the one hand, if O_1 has access only to classical means of communication, the best O_1 can do is to make a measurement on P_1 (which can supply only partial information about its state), and then report the outcome to O_2. On the other hand, suppose that O_1 and O_2 share a pair (P_2, P_3) of particles in the Werner state. Suppose also that O_1 makes a measurement on the pair (P_1, P_2) and reports the outcome of this measurement to O_2. It can then be shown that O_2 is more likely to infer correctly the initial state ψ of P_1 than would be possible if O_2 had access to only classical communication from O_1. Thus, the Werner state allows for the transmission of more information than any classical procedure. Whether this improved communication ability amounts to superluminal information transfer is a matter of dispute.

Nonlocality in Relativistic Quantum Field Theory

The special theory of relativity (at least according to its most popular interpretation) prohibits action at a distance, while quantum mechanics seems to require it. Surely this poses a serious problem of consistency: How can two theories be true (or at least approximately true) when their most basic principles contradict each other? Philosophers have often confronted this apparent contradiction by looking for creative ways to reinterpret relativity and quantum mechanics; some have even concluded that special relativity must be false. And yet, there already is a theory, *viz.,* relativistic quantum field theory, that is both relativistic and quantum mechanical. Although relativistic quantum field theory has been immensely successful in applications (in fact, it forms the basis for all of contemporary particle physics), philosophers have hardly begun to investigate how it manages to combine relativistic causality and quantum nonlocality (see Quantum Field Theory).

There are two structural features of relativistic quantum field models that mark them as distinctively relativistic. First, if *A* and *B* are spacelike separated regions, then any observable that can be measured in *A* is compatible with any observable that can be measured in *B*. This compatibility relation ensures that (nonselective) measurements performed in *A* cannot influence the statistics of measurements performed in *B*, and vice versa. Second, in relativistic quantum field models the spectrum of the four-momentum observable (i.e., the set of its possible measurement outcomes) is contained in the forward light cone. Thus, any measurement of the four-momentum will yield a result consistent with the predictions of special relativity; in particular, there can be no detectable energy-momentum transfer faster than light.

However, recent investigations have shown that these relativistic features of quantum field models do not preclude them from having nonlocal states. In fact, it has been shown (roughly speaking) that the percentage of nonlocal states grows in proportion to the dimension of the state space of the system. Thus, while systems with low-dimensional state spaces (e.g., spin-$\frac{1}{2}$ particles) have relatively few nonlocal states, systems with infinite-dimensional state spaces (e.g., field theories) have a very high percentage of nonlocal states. More specifically, for any two spacelike separated regions *A, B,* the set of field states that are Bell correlated across *A* and *B* is everywhere dense in the state space (Halvorson and Clifton 2000). Furthermore, every field state is maximally Bell correlated across unbounded tangent space-time regions, that is, "Rindler wedges" (Summers and Werner 1988). Finally, the vacuum state is non-separable across any pair of space-like separated regions, no matter how distant (Halvorson and Clifton 2000).

HANS HALVORSON

References

Aristotle (1941), "Physics," in Richard McKeon (ed.), *The Basic Works of Aristotle*. New York: Random House, VII, 2; 244b 16.

Bell, John S. (1964), "On the Einstein-Podolsky-Rosen Paradox," *Physics* 1: 195–200.

——— (1982), "On the Impossible Pilot Wave," *Foundations of Physics* 12: 989–999.

Brun, Todd, and Robert Griffiths (2000), Letter to the Editor, *Physics Today* 53.

Cushing, James (1996), "What Measurement Problem?" in Rob Clifton (ed.), *Perspectives on Quantum Reality.* Dordrecht, Holland: Kluwer, 167–181.

de Muynck, Willem (1986), "The Bell Inequalities and their Irrelevance to the Problem of Locality in Quantum Mechanics," *Physics Letters A* 114: 65–67.

Einstein, Albert (1948), "Quantenmechanik und Wirklichkeit," *Dialectica* 2: 320–324.

Einstein, Albert, Boris Podolsky, and Nathan Rosen (1935), "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review* 17: 777–780.

Fine, Arthur (1982a), "Hidden Variables, Joint Probability, and the Bell Inequalities," *Physical Review Letters* 48: 291–294.

——— (1982b), "Joint Distributions, Quantum Correlations, and Commuting Observables," *Journal of Mathematical Physics* 23: 1306–1310.

Fuchs, Christopher, and Asher Peres (2000), Letter to the Editor, *Physics Today* 53.

Gisin, Nicolas, and Asher Peres (1992), "Maximal Violation of Bell's Inequality for Arbitrarily Large Spin," *Physics Letters A* 162: 15–17.

Halvorson, Hans, and Rob Clifton (2000), "Generic Bell Correlation Between Arbitrary Local Algebras in Quantum Field Theory," *Journal of Mathematical Physics* 41: 1711–1717.

Hume, David (1978), *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge. New York: Oxford University Press.

Jarrett, Jon (1984), "On the Physical Significance of the Locality Conditions in the Bell Arguments," *Noûs* 18: 569–589.

Maudlin, Tim (1994), *Quantum Non-Locality and Relativity.* New York: Blackwell.

McMullin, Ernan (1989), "The Explanation of Distant Action: Historical Notes," in McMullin and James Cushing (eds.), *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem.* South Bend, IN: University of Notre Dame Press, 272–302.

Nerlich, Graham (1982), "Special Relativity Is Not Based on Causality," *British Journal for the Philosophy of Science* 33: 361–382.

Popescu, Sandu (1994), "Bell's Inequalities Versus Teleportation: What is Non-Locality?" *Physical Review Letters* 72 (1994): 797–799.

——— (1995), "Bell's Inequalities and Density Matrices: Revealing 'Hidden' Nonlocality," *Physical Review Letters* 74: 2619–2622.

Redhead, M. L. G. (1994), *Incompleteness, Nonlocality, and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics*. New York: Oxford University Press.

Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World.* Princeton, NJ: Princeton University Press.

Shimony, Abner (1993), *Search for a Naturalistic World View*. New York: Cambridge University Press.

Stapp, Henry (1997), "Nonlocal Character of Quantum Theory," *American Journal of Physics* 65: 300–304.

Summers, Stephen, and Reinhard Werner (1988), "Maximal Violation of Bell's Inequalities for Algebras of Observables in Tangent Spacetime Regions," *Annales de l'Institut Henri Poincaré* 49: 215–243.

Suppes, Patrick (1954), "Descartes and the Problem of Action at a Distance," *Journal of the History of Ideas* 15: 146–152.

van Fraassen, Bas (1991), *Quantum Mechanics: An Empiricist View*. New York: Oxford University Press.

von Neumann, John (1932), *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer.

Werner, R. F. (1989), "Quantum States with Einstein-Rosen-Podolsky Correlations Admitting a Hidden-Variable Model." *Physical Review A* 40: 4277–4281.

See also **Emergence; Quantum Field Theory; Quantum Mechanics; Reductionism**

LOGICAL EMPIRICISM

Logical empiricism dominated philosophical thinking about science from the late 1930s through the 1970s, so much so that nearly all philosophical writing about science in this period located itself, quite consciously, either within the fold of or in opposition to logical empiricism. Middle ground was, for a time, not easily found, and to ignore logical empiricism was to betray a profound ignorance of professional philosophy of science. Indeed, for many, the distinctly professional philosophy of science that emerged in the 1950s and 1960s had been made possible by—and was, perhaps, even identical to—logical empiricism. It is the task of this article to convey the main ideas and development of this profoundly influential philosophical movement.

Logical Positivism Versus Logical Empiricism

'Logical empiricism' is often used to refer to a philosophical school thought to be developed from *logical positivism*, whose more stringent *verifiability* criterion of cognitive meaning was replaced by a looser *confirmability* criterion; and an endorsement of scientific realism replaced logical positivism's rejection of the very question of whether the terms of a scientific theory refer to or of whether a successful theory's main claims should be taken to be, approximately true (see Cognitive Significance; Scientific Realism; Verifiability). (For the distinction between logical positivism and logical empiricism in this form, see Salmon 1999, 334.) But the notion that 'logical empiricism' suggests an identifiable, discrete, and conscious departure from something called logical positivism is tantamount to historical falsehood. In fact, logical empiricism, logical positivism, and *wissenschaftliche Philosophie* (scientific philosophy) had no fixed referents even in their heyday, and for good reason: What was usually behind these terms was an *attitude* or *approach,* rather than a theory or doctrine (a study of the terms used by participants themselves might suggest *wissenschaftliche Philosophie* as the most apt general term for the movement). And for this reason, current historical work should resist any urge to fix these terms' referents retrospectively (Hardcastle and Richardson 2003); logical positivism thus is not best described as having turned into, or given rise to, any particular successor movement, let alone one that embraced relaxed criteria of cognitive significance or scientific realism. It is worth noting that in contrast to 'logical positivism,' a term introduced in Blumberg and Feigl (1931), the provenance of 'logical empiricism' is murky.

On the other hand, insofar as logical empiricism denotes something like an intellectual development from and reaction to the main texts of logical positivism, the term is both useful and historically appropriate. Accordingly, this article on logical empiricism will present the tenets of logical positivism and the subsequent intellectual efforts that related themselves, directly and, on occasion, in opposition to these tenets; logical empiricism is really the story of the *development* of themes articulated within logical positivism. Following a discussion of five of logical positivism's main themes, then, logical empiricism will be described as it was reflected in ensuing philosophical work on analyticity, cognitive significance, holism, explanation, and the proper attitude toward scientific theories (see Analyticity; Cognitive Significance; Explanation; Theories).

Logical empiricism developed alongside intellectual projects with which it conflicted (socially and politically, as well as intellectually). The account of logical empiricism given here describes these conflicts without the typical "mortality" metaphor, in which philosophical views are born, mature, age, and die, becoming then objects of historical study. Instead, logical positivism is presented as the *distillation* of a particular scientific/philosophical ethos, and logical empiricism as the gradual, but detectable, *dilution* of that ethos over several decades and across several realms. This alternative metaphor presents logical empiricism as not a dead philosophical idea of mere historical interest, but as a set of identifiable traces left by the problems, approaches, and thinking associated with logical empiricism—problems, approaches, and thinking that animate and explain philosophy of science in the twenty-first century.

Distillation: Central Themes of Logical Positivism

The central themes of logical positivism are:

1. Antimetaphysics
2. A close relation to the natural sciences
3. Logic

At logical positivism's heart is antimetaphysics, or, to put it positively, a "spirit of enlightenment and anti-metaphysical factual research" (*geist der Aufklärung und der antimetaphysischen Tatsachenforschung*) (Hahn, Carnap, and Neurath [1929] 1973, 301). Indeed, it is important to express this central theme positively, since the image of logical positivism (and logical empiricism) as a negative, even destructive, movement is an egregious misrepresentation. The clearest expression of logical positivism's spirit of enlightenment is the 1929 pamphlet *The Scientific Conception of the World: The Vienna* Circle [*Wissenschaftliche Weltauffassung: Der Wiener Kreis*] (hereafter SCW), the 64-page "manifesto" of the Vienna Circle, the group of scientists, mathematicians, and like-minded thinkers that gathered around Moritz Schlick in Vienna in the 1920s, including Rudolf Carnap, Herbert Feigl, Philipp Frank, Hans Hahn, Otto Neurath, and Friedrich Waismann (see the listings for many of these figures in this volume, along with Vienna Circle). The pamphlet's author was listed simply as the "Wiener Kreis," but Neurath in fact wrote the bulk of the text, with contributions from Hahn and Carnap (Neurath 1973). The document announced to the European intellectual world the Circle's "scientific world-conception," which, it made clear, was characterized not so much by theses of its own, but rather by its basic attitude, its points of view, and direction of research:

> Neatness and clarity are striven for, and dark distances and unfathomable depths rejected. In science there are no 'depths'; there is surface everywhere: all experience forms a complex network, which cannot always be surveyed and can often be grasped only in parts. Everything is accessible to man; and man is the measure of all things... . The scientific world-conception knows no *unsolvable riddle.* Clarification of the traditional philosophical problems leads us partly to unmask them as pseudo-problems, and partly to transform them into empirical problems and thereby subject them to the judgment of empirical science. The task of philosophical work lies in this clarification of problems and assertions, not in the propounding of special 'philosophical' pronouncements. (Hahn et al. [1929] 1973, 305–306; cf. Carnap 1963, 20–34)

To what was this "spirit of enlightenment" directed? In practice, logical positivism directed itself toward questions concerning the logical structure of the sciences—the "clarification of problems and assertions" in them. The SCW summarized, for example, various "fields of problems" still to be addressed at "the foundations" of arithmetic, physics, geometry, biology, psychology, and the social sciences. Significantly, though, the scientific world-conception also addressed "questions of life" (304–305) in an enlightened manner, and thereby linked the Vienna Circle and logical positivism with political and social sensitivity, if not activism. As the SCW declared:

> [E]ndeavors toward a new organization of economic and social relations, toward the unification of mankind, toward a reform of school and education all show an inner link with the scientific world-conception; it appears that these endeavors are welcomed and regarded with sympathy by the members of the Circle, some of whom indeed actively further them. (Hahn et al. [1929] 1973, 305; cf. Carnap 1963, 20–26)

The later impression that logical positivism, and thus logical empiricism, consisted mainly in the systematic *rejection* of various philosophical traditions and methods as (worthless) "metaphysics" is owed largely to A. J. Ayer's *Language, Truth, and Logic*, an early, popular, and polarizing report on the Vienna Circle (see Ayer, Alfred Jules). The young Ayer visited the Circle from late 1932 through March of 1933 and managed to grasp its opposition to metaphysics, but little of its positive program. Nevertheless, *Language, Truth and Logic*'s infamous first line—"The traditional disputes of philosophers are, for the most part, as unwarranted as they are unfruitful" (Ayer 1936, 33)—fixed for the next three decades an image of logical positivism and empiricism. (In fairness, Blumberg and Feigl's [1931] earlier but less influential English presentation of logical positivism also failed to convey logical positivism's enlightenment theme.) Ayer's employment of logical positivism's verifiability criterion as a test for cognitive meaningfulness, and thus as a means to attack metaphysical claims, will be taken up below.

The second theme of logical positivism, one subsequently prominent in logical empiricism, is its close relationship to the natural sciences. Indeed, the distance between Ayer's understanding of logical positivism and logical positivism itself is clearest at the very end of *Language, Truth, and Logic*, where Ayer calls for the "philosopher to become a scientist... if he is to make any substantial contribution towards the growth of human knowledge" (Ayer 1936, 153). Ayer took himself to be calling for an unrealized philosophical future, but in fact logical positivism was already steeped in science. Its adherents aspired in their philosophical work to the sobriety and clarity of science, as represented particularly in relativistic physics (see Conventionalism; Space-Time). And its practitioners in Vienna, Berlin, and Prague were themselves adept at relativistic and quantum physics; the Circle's Philipp Frank, in fact, succeeded Einstein in Prague in 1912 when Einstein went to Berlin. Yet Ayer's insulation from actual science, combined with his authorship of *Language, Truth, and Logic*, led to the unfortunate, mistaken, and ironic impression that logical positivism was actually *detached* from contemporary science.

Steeped in science, logical positivism nevertheless distinguished itself from science, although what sort of project it *was* exactly would be subject to continued debate. The distinction between logical positivism and science itself was reflected most clearly in the fact that the former attended not to particular domains of experience (that was the focus of the various separate sciences) but to experience, and its *logical* structure, *as a whole.* Logic—the study of the *form* of scientific statements and the experience they describe—emerges as a third theme of logical positivism. And it is in this context that logical positivism's verifiability criterion is best appreciated. In SCW the criterion can be glimpsed as the claim that "for us, *something is 'real' through being incorporated into the total structure of experience*" (Hahn et al. [1929] 1973, 308, emphasis in original; see also Schlick [1930–1931] 1959). A claim or entity that could *not* be incorporated into the "total structure of experience" was therefore not real, in the strongest sense, and thus not even describable. Attempts to express such a claim or describe such an entity were at best confused and at worst disingenuous: metaphysics in the most damning sense (see, e.g., Blumberg and Feigl 1931; Carnap [1932] 1959, esp. §7).

Logic, particularly the quantificational logic articulated by Gottlob Frege and developed in 1910 by Whitehead and Russell (1963) and in 1921 by Wittgenstein (1961), was thus of enormous use to logical positivism. Indeed, the Circle read Wittgenstein's *Tractatus* painstakingly (Feigl 1969, 634). Moreover, the logical positivists' view of logic allowed for a statement to be true in a language purely by virtue of the meaning of its terms, that is, to be *analytically* true; such sentences provided an essential component of their epistemological account of mathematics and the formal sciences generally (see Analyticity). If mathematical statements were ultimately analytic statements of logic, and if logical statements were truths that could be produced at will, as it were, by the articulation of a language (perhaps an artificial language) in which they always came out true, then the truth of a mathematical sentence could be explained by reference to nothing more puzzling than a simple convention—a decision, freely taken, to render the sentence in question true (see Conventionalism). Knowledge of the truths of mathematics, *qua* sentences of logic, would then be as transparent or obvious as the language itself. Such a result embodied the enlightenment ethos of the scientific world-conception, and, correspondingly, challenges to it (as well as to, more generally, the view of logic at the heart of logical positivism) carried particular weight.

It remains to recognize two further important themes of logical positivism. One, implicit in the *generality* of logic, is the unity of science (*Einheitswissenshaft*) (see Unity and Disunity of Science; Unity of Science Movement). The SCW duly sounded the unity-of-science theme, which would become central to logical empiricism in the United States: "The goal ahead," the SCW stated, "is *unified science*. . . .The endeavor is to link and harmonize the achievements of individual investigators in their various fields of science" (Hahn et al. [1929] 1973, 306). Sciences, that is, were not to be distinguished on the basis of different methods, subject matters, or attitudes; their attitudes and methods did not differ, and differences in subject matter reflected merely pragmatic divisions of labor. Just beneath the surface of the unity-of-science theme was the denial of a division between the natural and social sciences, that is, between *Natur*- and *Geisteswissenschaften*; this denial had substantial political import for logical empiricism (Reisch 2004).

Finally, emphasis on the unity of science served to underscore a final theme implicit in the other three, *viz.,* the question of logical positivism's own place in the realm of inquiry or knowledge. In the 1920s and 1930s, the logical positivists struggled to describe their work in its own terms, that is, to describe their perspective as a philosophy informed by but not identical to science and its world-conception. Some, such as Schlick, subscribed to Wittgenstein's apparent view that the propositions of philosophy, properly understood, were themselves meaningless (Wittgenstein 1961; Schlick [1930–1931] 1959). Against Schlick's view, Carnap (1934) argued that a sufficiently rich metalanguage allowed for the expression of truths about the logical structure of science, including truths about that metalanguage itself. This was a matter within logical positivism that was not resolved and, indeed, became a central question for logical empiricism.

Dilution of Logical Empiricism: Analyticity and Cognitive Significance

World War II's significance to the development of logical empiricism is enormous, and at present only partly understood and appreciated (Reisch 2004). The war interrupted work within scientific philosophy in the obvious, material, way, but significantly also by virtue of logical positivism's perceived peripheral significance compared with the war effort. After World War II, scientific philosophy—now a distinctively North American endeavor, with Carnap, Hempel, Feigl, Frank, and Hans Reichenbach having emigrated to the United States—resumed work put aside in the late 1930s and early 1940s, although it did so in a very different cultural and political climate (see Carnap, Rudolf; Hempel, Carl Gustav; Reichenbach, Hans).

The result, by the early 1950s, consisted of three papers that would each have, for decades to come, a profound effect on philosophical thinking about science, challenging many (but, importantly, not all) of the themes associated with logical positivism. These papers were Quine's (1951) *Two Dogmas of Empiricism* and two essays published by Hempel in the early 1950s (Hempel 1965) (see Hempel, Carl Gustav; Quine, Willard Van). The former argued against the widely accepted view that certain truths were to be accounted for as true by virtue of the meanings of their terms (and, indeed, the article rejected as dogma the doctrine that there *were* analytic truths in this sense), while Hempel's two papers taken together reviewed and eventually rejected the notion that statements on their own had cognitive significance, by which was meant empirical meaning (see Analyticity; Cognitive Significance). Together the line of argument in these works pointed scientific philosophers sympathetic to logical positivism toward a holism about the meaning of statements and a pragmatism about any separation between science and metaphysics. The result, ultimately, was a nearly complete dilution, by the 1960s, of logical empiricism. An examination of these papers (each of which, incidentally, summarized discussions and work of several previous years) will illustrate this development.

Two Dogmas of Empiricism ostensibly takes up the following question: In virtue of what, precisely, are certain "analytic" sentences, the truth of which seems to be unavoidable, true? (Quine's example is "All bachelors are unmarried men.") Recognizing that the logical positivists' answer appealed to the meaning of the nonlogical terms in such sentences ('bachelor' and 'unmarried men'), Quine presses the question of how, precisely, the meaning of these terms manages to accomplish such a feat (see Analyticity). A good portion of *Two Dogmas of Empiricism* is given over, then, to showing how successive versions of the putative semantic relations between terms, and between sentences and a language, fail to provide the needed explanation of the analytic truths in question, typically because the purported explanation rests on concepts as much in need of clarification as meaning itself (Quine considers, specifically, definition, interchangeability *salva veritate*, semantic rules, and verificationism). Having presented a comprehensive, if not exhaustive, list of potential accounts and found them all wanting, Quine proceeds to doubt the presumption of the question, doubting, that is, that so-called analytic sentences are true by virtue of meaning. He then offers an alternative view. The single exception Quine recognizes to his list of failed accounts of analyticity, though, is telling. Quine allows that truths arising from the "explicitly conventional introduction of novel notations for sheer abbreviation" *does* suffice to account for some analytic truths, for here the definiendum becomes synonymous with the definiens simply because it has been created expressly for this purpose: "Here we have a really transparent case of synonymy created by definition; would that all species of synonymy were as intelligible" (26). Such "explicitly conventional" definition is rejected simply because it in fact is at the root of very few, if any, of the *actual existing* instances of analytic statements *in the present language.* Radically revising, or even discarding, present language, or more generally present theories, is not a live option.

Quine's rejection of explicitly conventional definition in *Two Dogmas of Empiricism* reflects not just his own conservatism but, significantly, that of the philosophy of science from the early 1950s on (see Quine, Willard Van). The enlightenment optimism of the SCW was no longer seriously considered; a departure from the past by means of the adoption of a new, modern, scientific attitude and the creation of a new, transparent language suitable to modern needs was simply dismissed.

Quine's critical treatment/rejection of various explanations of analytic truth is followed by his own account of those truths. His alternative is at once novel and conservative:

> The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the

> profoundest laws of atomic physics or even pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges .... A conflict with experience at the periphery occasions readjustments in the interior of the field ... . But the total field is so underdetermined by ... experience that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience .... If this view is right ... it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold come what may. Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system .... Conversely ... no statement is immune to revision. (42–43)

Analytic statements, then, are simply those abandoned last, if ever, in the face of experience. Contained in this influential metaphor, which would come to be known as the web of belief, is a powerful challenge to several of logical positivism's central themes (see Duhem Thesis). Quine's holism—reflected particularly in his insistence that present and future language and theory be continuous with past theory and language—is in considerable tension with the progressiveness of the Vienna Circle (a tension embodied, indeed, in the work of Neurath (1973) (see Neurath, Otto). Further, Quine's (1951) pragmatism, displayed above in his recognition of several different but acceptable reactions to "contrary experience," "blurs," as he puts it, the "supposed boundary between speculative metaphysics and natural science" (10).

The effect of *Two Dogmas of Empiricism* was dramatic, and was only heightened when an analysis from a somewhat different starting point led to a very similar set of recommendations. Hempel's *Empiricist Criteria of Cognitive Significance: Problems and Changes,* which in many ways continued work begun by Carnap (1936–1937), reviewed and ultimately abandoned the thesis that there existed a "sharp dividing line ... between those sentences which do have cognitive significance and those which do not," or, for that matter, between significant and insignificant, or metaphysical, *systems* or *theories* (see Cognitive Significance; Hempel, Carl Gustav). Posing the problem in terms of a search for a formal relation between putatively cognitively significant statements and "observation sentences" (which contain only observational terms and are thus empirically unobjectionable), Hempel examined and rejected both verificationism (construed as the thesis that a cognitively significant sentence must be entailed by a consistent finite set of observation sentences) and falsificationism (that its negation be so entailed), as well as closely related proposals, on the grounds that each criterion either admitted clearly insignificant claims or, alternatively, barred clearly significant ones.

Pursuing the different tack of isolating the cognitive significance of *terms* (on the basis of "observation" terms), Hempel (1965) established first the inadequacy of the reductionist strategy Carnap (1936–1937) had outlined for theoretical terms, and then endorsed the semantic holism Quine also forwarded:

> It is not correct to speak ... of the "experiential meaning" of a term or a sentence in isolation. [A] single statement usually has no experiential implications .... [T]he occurrence of certain observable phenomena can be derived from it only by conjoining it with a set of other, subsidiary, hypotheses ... [that] will usually be observation sentences [and] accepted theoretical statements. (112)

More significantly, Hempel was led from this holism to philosophical morals that, again, echo Quine's. After a failed search for criteria to separate cognitively significant from cognitively insignificant systems or theories (by way of barring those containing "isolated" sentences, that is, those without "experiential bearing"), Hempel (1965) cautiously suggests that "it is not possible to formulate ... criteria which would separate those ... systems whose isolated sentences ... have a significant function from those in which the isolated sentences are ... mere useless appendages. (117)

Rather,

> [C]ognitive significance . . . is a matter of degree: Significant systems range from those whose entire extralogical vocabulary consists of observation terms, through theories whose formulation relies heavily on theoretical constructs, on to systems with hardly any bearing on potential empirical findings. (117)

As with Quine's conclusion in *Two Dogmas of Empiricism,* such a claim constitutes (and was understood at the time to constitute) an abandonment of the *spirit* of the verifiability criterion, and of logical positivism. Moreover, here, as again with Quine (1951), logical empiricism took direction from Hempel's conclusion, which (like Quine's) revealed a certain conservatism in opposition to logical positivism's "spirit of enlightenment and anti-metaphysical factual research." For example, Hempel's critique of cognitive significance relied upon the appeal to the value of theory, even (indeed, especially) theory disconnected from experience or practice. As Hempel (1965) put it:

> The history of scientific endeavor shows that if we wish to arrive at precise, comprehensive, and well-confirmed

> general laws, we have to rise above the level of direct observation .... In following ... a narrowly phenomenalistic or positivistic course, we ... deprive ourselves of the tremendous fertility of theoretical constructs, and we ... often render the formal structure of the expurgated theory clumsy and inefficient. (116)

Hempel's brief for empirically isolated theory contrasts deeply with the antimetaphysical spirit of logical positivism, and would shape logical empiricism in many ways, for decades.

Dissolution of Logical Empiricism: Unity, Realism, and the Philosophy of Science

It is important to recognize that the challenges to logical empiricism posed by Quine and Hempel, among others, as deep as they were, remained in a broadly scientific philosophical context. In other respects, logical empiricist themes were endorsed and elaborated. This is the case, for example, regarding the unity of science, which figured in logical empiricism in two significant ways.

Hempel's studies of cognitive significance were undertaken at nearly the same time as studies of what Hempel called the "logic of explanation," the attempt to characterize, as with cognitive significance, the formal relation obtaining between an event or a regularity to be explained, an *explanandum,* and that which explains the explanandum, the *explanans* (see Explanation). Hempel was motivated to pursue a logic of explanation in order to counter the view that methodology within the social sciences, notably history, differed in kind from what was found in the natural sciences. Both, Hempel urged, accomplish explanation by showing how the explanandum was to be expected given the laws inevitably cited in the explanans. In his first discussion of explanation, *The Function of General Laws in History,* Hempel (1942) argued that historical explanation "aims at showing that the event in question was not 'a matter of chance,' but was to be expected in view of certain . . . conditions. The expectation referred to is not prophecy or divination, but rational scientific anticipation which rests on the assumption of general laws" (39).

Hempel later attempted to extend the fundamental idea contained in his account of explanation—that explanation consisted in subsuming the explanandum under general laws—to both the use of statistical laws to explain particular facts and the explanation of the laws themselves. The explication of scientific explanation in all its guises emerged as a research project of great fecundity in the 1950s and 1960s, as philosophers such as Braithwaite (1953) and Nagel (1961) pursued the essentially Hempelian project of providing a single explication of scientific explanation in all its guises, including, notably, explanations that made use of functions, particularly within a biological context. In this respect, Nagel's (1961) account of functional explanation came to exercise particular influence (see Function). This research project remained both popular and fruitful through the 1980s, as exemplified by Kitcher and Salmon (1989). Its roots, however, lay in the unity-of-science thesis. (For an authoritative overview of research on explanation, see Salmon 1990 and 1998.)

Hempel also pursued a parallel, and equally fecund, project with respect to confirmation; the effort here was to express the single relationship between evidence and hypothesis as it was to be found across all the sciences (Hempel 1965) (see Confirmation Theory). Hempel's efforts to capture the confirmation relation syntactically were dealt a fatal blow by Goodman (1955) (see Induction, Problem of), although valuable and influential work in this direction continued under the heading of inductive logic in the hands of Carnap, who made extensive use of the logical (as opposed particularly to a frequentist) interpretation of probability (Carnap 1950 and 1952) (see Inductive Logic; Probability).

In another, different guise, the unity of science proved far less successful. The Unity of Science movement, an official collaboration of Neurath, Carnap, and Charles Morris, had begun in Europe in 1934 with enormous ambition and promise (see Unity of Science Movement). It included international congresses, the *Journal of Unified Science* (a reincarnated version of the earlier logical positivists' organ *Erkenntnis*), and a separate Library of Unified Science, containing, in one vision, two hundred separate monographs (Neurath, Carnap, and Morris 1971). But World War II, a series of personal disputes between the collaborators, and, possibly, increasing (and justified) association of the Unity of Science movement with socialism and communism during the onset of the Cold War in the United States resulted in the nearly complete disintegration of the enterprise by the early 1950s. Thus the idea of scientific unity survived, even thrived, in the search for universal formal accounts of explanation and confirmation, while it languished in the cultural forum that had previously identified it with political aims—specifically, progressive socialism (Reisch 2004).

Hempel's welcoming of theoretical terms and entities disconnected from observation, as well as Quine's appeal to purely pragmatic criteria for preferring certain ontologies to others in an account of the world, contributed significantly to the consideration of the question of under what conditions, if any, the success of a scientific theory warranted belief that its central claims were in fact true or that the entities it mentioned in fact existed. Scientific realism—the position that success did warrant actual belief in, rather than mere acceptance of, a theory—was subsequently the focus of much debate from the late 1950s on, leading in turn to careful attention to the distinction between observable and nonobservable entities and the nature of success for scientific theories (see Scientific Realism). The debate itself, quite apart from any resolution it reached, demonstrated the intellectual distance logical empiricism had come from the attitude heralded in the SCW.

Thus by the late 1950s, logical empiricism embodied a tension between the historically oriented holism suggested by Quine and (less so) Hempel and the pursuit of general, unified accounts of science by way of specifying formal relations definitive of explanation or confirmation. This tension provides one (but hardly the only) means to understand the reaction to Thomas Kuhn's (1962/1970) profoundly influential *The Structure of Scientific Revolutions,* an essay that pitted an image of science that Kuhn understood to be logical empiricism against another image garnered from a close reading of scientific changes and the texts surrounding them (see Kuhn, Thomas; Scientific Revolutions). The latter image, Kuhn argued, engaged on its own terms, would transform the former and possibly lead to a rejection of both the unity-of-science thesis and the notion that science itself was a social institution with epistemic authority and privilege that it had earned. Yet, while *The Structure of Scientific Revolutions* was offered (and is now perceived) as an attack on logical empiricism, in fact several logical empiricists (most notably, Carnap) endorsed it. Such were the tensions within logical empiricism by the early 1960s.

Logical empiricism, understood as the reexamination, modification, and (alternatively) rejection and endorsement of the themes of logical positivism, is perhaps no more detectable within philosophy of science in the early twenty-first century than in current discussions of the nature and place of the philosophical examination of science, a topic, as mentioned above, that exercised the Vienna Circle. A good number of logical empiricists or their heirs subscribe to some version of Carnap's understanding of philosophy as the analysis of the logical structure of the concepts of science; work on confirmation, explanation, and other general concepts proceeds on several fronts. Others take up Quine's development of his own holism in his call for continuity between science and philosophical thought *about* science; Quine's "naturalized" account of epistemology describes philosophy of science as science itself. So motivated, philosopher-scientists have contributed to a number of scientific fields since the late twentieth century, most notably biology and physics. Finally, a felt need to connect philosophy of science in either guise to social and political matters, as well as study of the traditional aims and history of logical positivism and empiricism, has reopened discussion of the social and political dimensions of the philosophy of science and informed, for example, the *Institut Wiener Kreis,* an especially active institute of the University of Vienna dedicated to promoting historical study and understanding of the Vienna Circle and its aims. The perspective of logical empiricism thus informs the best philosophy of science done today.

Gary Hardcastle

References

Ayer, Alfred Jules (1936), *Language, Truth, and Logic.* London: Victor Gollancz.

——— (ed.) (1959), *Logical Positivism.* New York: Free Press.

Blumberg, Albert E., and Herbert Feigl (1931), "Logical Positivism: A New Movement in European Philosophy," *Journal of Philosophy* 28: 281–296.

Braithwaite, R. B. (1953), *Scientific Explanation.* Cambridge: Cambridge University Press.

Carnap, Rudolf ([1932] 1959), "The Elimination of Metaphysics through Logical Analysis of Language" in A. J. Ayer (ed.), *Logical Positivism.* New York: Free Press, 60–81. Originally published as "Überwindung der Metaphysik durch der Logische Analyse der Sprache," *Erkenntnis* 2: 219–241.

——— (1934), *Logische Syntax der Sprache.* Wien: Springer. Translated as *Logical Syntax of Language.* London: Kegan Paul, 1937.

——— (1936–1937), "Testability and Meaning," *Philosophy of Science* 3 (1936): 419–471, and 4 (1937): 1–40.

——— (1950), *Logical Foundations of Probability.* Chicago: University of Chicago Press.

——— (1952), *The Continuum of Inductive Methods.* Chicago: University of Chicago Press.

——— (1963), "Intellectual Autobiography," In Paul A. Schilpp (ed.), *The Philosophy of Rudolf Carnap.* LaSalle, IL: Open Court, 3–84.

Feigl, Herbert (1969), "The Wiener Kreis in America," in D. Fleming and B. Bailyn (eds.), *The Intellectual Migration: Europe and America, 1930–1960.* Cambridge, MA: Belknap, 630–673.

Goodman, Nelson (1955), *Fact, Fiction, and Forecast.* Cambridge, MA: Harvard University Press.

Hahn, Hans, Rudolf Carnap, and Otto Neurath ([1929] 1973), "The Scientific Conception of the World: The Vienna Circle," in Marie Neurath and Robert S. Cohen (eds.), *Otto Neurath: Empiricism and Sociology.* Dordrecht, Holland: Reidel, 299–318. Originally published as "Wissenshaftliche Weltauffassung: Der Wiener Kreis," *Veröffentlichungen des Vereines Ernest Mach*. Wien: Artur Wolf Verlag.

Hardcastle, Gary L., and Alan Richardson (eds.) (2003), *Logical Empiricism in North America*. Minneapolis: University of Minnesota Press, xiv–xvi.

Hempel, Carl G. (1942), "The Function of General Laws In History," *Journal of Philosophy* 39: 35–48.

——— (1965), "Empiricist Criteria of Cognitive Significance: Problems and Changes," in Hempel, *Aspects of Scientific Explanation*. New York: Free Press, 101–122.

Kitcher, Philip S., and Wesley Salmon (1989), *Scientific Explanation*. Minneapolis: University of Minnesota Press.

Kuhn, Thomas (1962), *Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Nagel, Ernest (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World.

Neurath, Otto (1973). *Empiricism and Sociology*. Edited by Marie Neurath and Robert S. Cohen. Dordrecht, Holland: Reidel.

Neurath, Otto, Rudolf Carnap, and Charles C. Morris (1971), *Foundations of the Unity of Science: Toward an International Encyclopedia of Unified Science*, 2 vols. Chicago: University of Chicago Press.

Quine, Willard Van (1951), "Two Dogmas of Empiricism," *Philosophical Review* 60: 20–43.

——— (1953), *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

Reisch, George (2004), *How the Cold War Transformed Philosophy of Science*. Cambridge: Cambridge University Press.

Salmon, Wesley C. (1990), *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

——— (1998), "Scientific Explanation: How We Got from There to Here," in *Causality and Explanation*. Oxford: Oxford University Press.

——— (1999), "The Spirit of Logical Empiricism: Carl G. Hempel's Role in Twentieth-Century Philosophy of Science," *Philosophy of Science* 66: 333–350.

Schlick, Moritz ([1930–31] 1959), "The Turning Point in Philosophy," in A. J. Ayer (ed.), *Logical Positivism*. New York: Free Press, 53–59.

Whitehead, Alfred North, and Bertrand Russell (1963), *Principia Mathematica*, 2nd ed. Cambridge: Cambridge University Press.

Wittgenstein, Ludwig (1961), *Tractatus Logico-Philosophicus*. Translated by D. F. Pears and B. F. McGuinness. London: Routledge and Kegan Paul.

See also **Analyticity; Ayer, Alfred Jules; Carnap, Rudolf; Cognitive Significance; Confirmation Theory; Explanation; Explication; Feigl, Herbert; Hahn, Hans; Hempel, Carl Gustav; Induction, Problem of; Inductive Logic; Kuhn, Thomas; Nagel, Ernest; Neurath, Otto; Phenomenalism; Physicalism; Popper, Karl Raimund; Probability; Protocol Sentences; Quine, Willard Van; Rational Reconstruction; Reichenbach, Hans; Scientific Realism; Schlick, Moritz; Theories; Unity and Disunity of Science; Unity of Science Movement; Verifiability; Vienna Circle**

M

ERNEST MACH

(18 February 1838–19 February 1916)

Mach was a physicist, psychologist, philosopher, and historian of science, as well as a political figure (for biographical detail, see Blackmore 1972). He studied physics at the University of Vienna from 1855 to 1861, continuing there as a lecturer until 1864. After spending three years as professor of mathematics at Graz, he received a chair at Prague, where he stayed until 1895. For the next six years, Mach occupied a chair in the History and Philosophy of the Inductive Sciences at Vienna. He suffered a stroke in 1898 and retired in 1901.

Mach's Influence

As a political figure, Mach served in the Austrian parliament and was so influential amongst the Austrian and Russian left that Lenin wrote *Materialism and Empirico-Criticism* as a criticism of Mach's anti-materialism. In physics, he was the first to understand supersonic shock waves and was a major influence upon a generation of physicists, including Einstein (who credited Mach for being a philosophical forerunner of relativity theory), Schrödinger, Planck, and Heisenberg. In psychology, Mach was a founder of Gestalt theory and made numerous contributions to sense physiology (see Psychology, Philosophy of). His research on Mach bands anticipated the modern understanding that the senses have neural nets that preprocess information before sending it to the brain. In philosophy, he was a major influence on the Vienna Circle (especially Frank, Hahn, and Carnap) and remains an inspiration to empiricist conceptions of science (see Vienna Circle). He was one of the most influential intellectuals in Vienna at a time when Vienna was the center of Western intellectual activity.

Mach's Psychology and Biology

Although he received his degree in physics, Mach attended many classes in physiology, and from the beginning of his career the majority of his research was not in physics but in the physiology of the senses. His central project, developed most extensively in his *Analysis of Sensations*, was to understand the relationship between sensations (the actual phenomenal experience) and the physical stimuli that trigger them (Mach [1914] 1984).

How, for instance, do eyes convert light particles into three-dimensional visual fields, when those particles themselves contain no information about their point of origin?

Although an empiricist in many regards, Mach held a strong notion of the a priori, maintaining that three-dimensional space is biologically innate (Banks 2003). Mach differed from the Kantian tradition by holding that the a priori is formed through evolution and development. Thus, there is nothing necessary about the current human spatial intuition; had humans evolved differently, they would perceive space differently. A similar account is given for other aspects of cognition, such as the understanding of matter, time, color, and even mathematics. The current human condition is historically contingent upon the particular evolutionary pathway humans accidentally took, and thus the world known through human senses should not be confused with the actual world.

Mach was strongly influenced by Darwin and by a variety of evolutionary ideas that are rejected today but were prominent within nineteenth-century German culture (see Evolution). Borrowing from the latter, Mach saw humans as nature's way of understanding itself, almost a waking up of nature. Science was a continuation of this process in that it was the "waking up" of humans. Evolution in prescientific times was an unconscious adaptation of organisms to their physical environments. The selection pressure was on survival and not on truth or higher social ideals. Humans thus adapted to convenient local minima and have come to misinterpret the human view of the world for the world itself. Thus there is a confusion between biological programming and reality. It is the biological purpose of science to lift humans out of this condition by understanding the nature of our psychological programming, and to provide us with a stable environment for future cognitive growth. It was an optimistic outlook; Mach's brand of positivism held out the hope that psychology embedded within an evolutionary framework would change the way humans saw the world, and thus lead not only to knowledge but, more importantly, to social-political harmony.

Mach's Philosophy of Physics

Two central areas dominate Mach's philosophy of physics: his opposition to atomism and his opposition to Newton's conception of absolute space and time. Both arise from his antimetaphysical attitude, which in turn derives from his placing physics within a biopsychological framework.

His opposition to atomism arose from his biological conception of the purpose of science. For Mach, the social-political value of a scientific worldview is that it is nonmetaphysical and stable, and thus can provide a basis for positive scientific and social progress. Science should be nonmetaphysical so that all humans can agree to it. Furthermore, a nonmetaphysical science is stable in that it is not grounded in speculation but in description. This led him to favor a view of physics that emphasized description of the phenomena over the positing of ontologies and theories. Descriptions are simply more stable and less likely to be overthrown by future science, and thus provide a stable environment for cognitive change.

Turning to Mach's theories of space and time, it is important to note that Mach writes far more on the *psychology* of space than on the physics of space. In particular, his research on Mach bands develops the idea that even at a sensory level, one experiences only relations between things, and not the things themselves (Ratliff 1965). Mach's critique of Newton's theories of absolute space and time (which in turn influenced Einstein) was thus derivative of this psychological outlook. The history of physics thus owes a small but important debt to psychology. This is still controversial. What is agreed upon is that Mach rejected the mechanical view of physics and developed an influential alternative in which space, time, and matter were redefined as relations. For instance, he defined matter in terms of its relational interactions with other matter, that is, how much acceleration two objects impart to each other. With space and time, Mach similarly argued for a relational view. Humans have no psychological access to what space and time "are," so physics should simply mathematically describe the relations between objects.

In his influential *Science of Mechanics*, he argues that space and time cannot be used to measure the absolute changes of objects, as the very concepts of space and time are arrived at by observing the changes in objects (Mach [1893] 1960). That is, an intuition of space and time is prerequisite to measuring motion. Thus one cannot claim after measuring a motion that one has measured 'space' and 'time.' All one can do is give mathematical accounts of the spatial–temporal relationships of things. Mach, then, was a radical naturalistic epistemologist who turned to biology and psychology to give a naturalistic account of the entire human condition, including physics.

Paul Pojman

References

Banks, Erik (2003), *Ernest Mach's World Elements*. Dordrecht and Boston: Kluwer Academic.

Blackmore, John T. (1972), *Ernest Mach: His Work, Life, and Influence*. Berkeley and LA: University of California Press.

Mach, Ernest ([1914] 1984), *The Analysis of Sensations and the Relation of the Physical to the Psychical*. Translated by C. M. Williams. LaSalle, IL: Open Court.

——— ([1893] 1960). *The Science of Mechanics: A Critical and Historical Exposition of its Principles*. Chicago: Open Court Publishing.

Ratliff, Floyd (1965), *Mach Bands: Quantitative Studies on Neural Networks in the Retina*. San Francisco: Holden-Day.

See also **Logical Empiricism; Phenomenalism; Space–Time; Verifiability; Vienna Circle**

MATERIALISM

See **Physicalism**

MECHANISM

Interest in mechanisms has experienced a recent upsurge in the philosophy of science generally (e.g., Salmon 1984; Glennan 1996) and in the philosophy of biology and neuroscience in particular (see e.g., Bechtel and Richardson 1993; Craver and Darden 2001; Machamer, Darden, and Craver 2000). Scientific explanation often involves identifying the mechanism responsible for a phenomenon of interest. This entry provides a generic account of what mechanisms are and how they are appealed to in explanations and then turns to the question of how scientists discover them. What are mechanisms?

Four Aspects of Mechanisms

The notion of mechanism has four aspects: (i) a phenomenal aspect, (ii) a componential aspect, (iii) a causal aspect, and (iv) an organizational aspect. Mechanisms can differ from one another in each of these aspects. Consider them in turn, first using a common mousetrap as an example and then considering the more complicated mechanism of action potential generation in neurons:

The Phenomenal Aspect

Mechanisms do things; they are the mechanisms *of* the things that they *do*. A mousetrap traps mice, and the mechanism for generating action potentials generates action potentials. These tasks performed by the mechanism as a whole are the *phenomena* explained by the working of the mechanism. There are no mechanisms *simpliciter*—only mechanisms *for* phenomena. A mechanism's phenomenon partially determines the mechanism's boundaries (i.e., what is "in" the mechanism and what is not). As Kauffman (1971) clearly emphasized, an item is considered "part of" the mechanism only if it is relevant to a mechanism's phenomenon.

The Componential Aspect

Mechanisms have components, or working parts. Mechanisms all have at least two components. The old-fashioned mousetrap has six: a platform, a trigger, a latch, a catch, a spring, and an impact bar (see Figure 1). Trivially, the components are proper parts of the mechanism as a

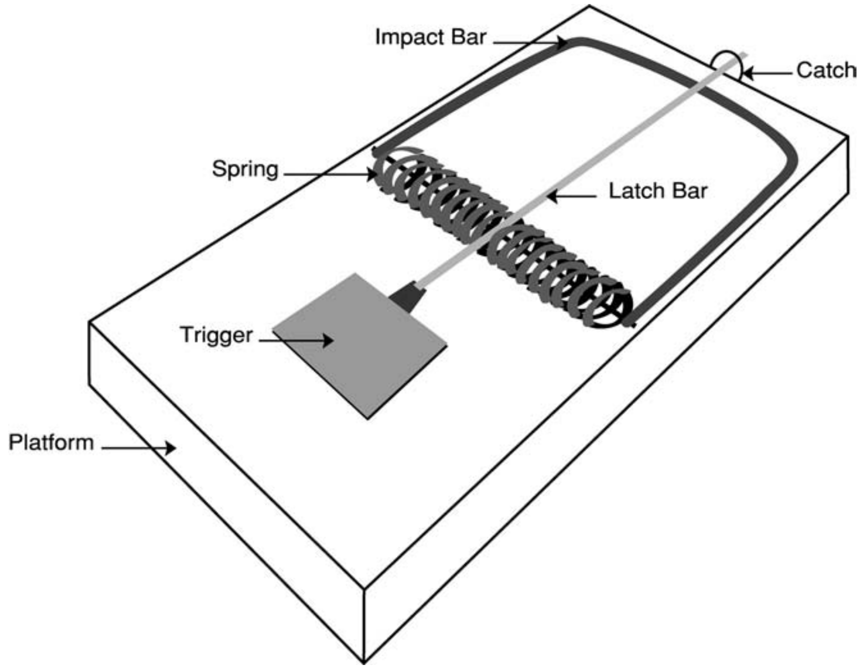


Fig. 1. The Mousetrap.

whole. More restrictively, as just noted, the parts of a mechanism are those that are relevant to the phenomenon explained by the mechanism. The parts are relevant to the phenomenon by virtue of certain of their properties (and not others). But for the rigidity of the bar and the tension on the spring, a mousetrap would catch no mice. The buoyancy of the platform, in contrast, is not properly included in the mechanism for catching mice.

The Causal Aspect

The components of mechanisms act and interact with one another. If they did not, they would not do anything. *Pressing* the trigger *releases* the catch, *allowing* the spring to *launch* the impact bar. The verbs in this description of the mousetrap refer to the relevant causal relations among the component parts. Talk of causal relations is a schematic placeholder to be filled in with one or more appropriate accounts of the kinds of causing exhibited in a given case. Philosophical attempts to develop univocal analyses of such causal relationships have yet to garner widespread acceptance. Yet the intransigence of the causal relation to a single uniform philosophical analysis should not distract attention from the central role that causal relations play in mechanistic explanations.

The Organizational Aspect

The components of mechanisms and their causal relations are organized spatially and temporally in the production of the phenomenon. The *spatial organization* of a mechanism includes the relative locations, shapes, sizes, orientations, connections, and boundaries of the mechanism's components. In the mousetrap, the trigger and the catch have to be so located with respect to one another that a small amount of pressure on the trigger moves the trigger bar enough to dislodge from the catch. The catch is circular and accommodates the size of the trigger bar. When the mechanism is loaded, the parts are connected to one another: The trigger bar restrains the blunt bar because it is stuck in the catch. As the mousetrap ‘‘fires,’’ temporal organization takes center stage. The temporal organization of a mechanism includes the order, rates, durations, and frequencies of the activities in the mechanism. If a mousetrap is to work, it should work quickly, it should not discharge until there is pressure on the trigger, and there should not be significant delays between the steps of its working. Spatial and temporal organization are two important varieties of mechanistic organization. There are familiar patterns of mechanistic organization that can be found in different mechanisms for different phenomena. Some mechanisms are feed-forward, with each step following upon its predecessor without forks, joins, or cycles (like the common mousetrap); others may work in parallel or have significant feedback connections.

A Neurobiological Example: The Mechanism of the Action Potential

Mousetraps fire, and so do neurons. The firing of a neuron is known as an action potential. Action potentials are changes in the electrical potential difference across the cell membrane that propagate along the length of the neuron. This difference, known as the membrane potential (V_m), consists of the separation of charged ions on either side of the membrane. In the neuron's resting state, positive ions line up against the membrane's extracellular surface, and negative ions line up on the intracellular side, producing a polarized resting potential (V_{rest}) of roughly −60 mV. The action potential (as indicated in Figure 2) consists of (I) a rapid rise in V_m (reaching a maximum value of roughly +20 mV), followed by (II) an equally rapid decline in V_m to values below V_{rest}, and then (III) an extended hyperpolarized afterpotential during which the neuron is less excitable. These three features characterize the phenomenon to be explained by the action potential mechanism.

The components of this mechanism include the cell membrane, positively charged sodium (Na^+) ions, positively charged potassium (K^+) ions, and two types of voltage-sensitive ion channels that selectively allow, respectively, Na^+ or K^+ ions to diffuse through the membrane. It is the temporally organized activities of these channels that produce the action potential phenomenon.

The mechanism of the action potential starts with a cumulative depolarization of the cell body (i.e., V_m becomes greater than V_{rest}), typically through the effect of neurotransmitters on ion channels in the cell's dendrites (the "receiving" ends of the neuron). Action potentials are generated in the axon hillock, an ion-channel-dense region of membrane at the interface of the cell body and the axon (the "sending" end of the neuron). Depolarization of the cell body opens voltage-sensitive Na^+ channels (increasing membrane conductance to Na^+), allowing Na^+ to diffuse down its concentration gradient from the Na^+-rich extracellular fluid into the relatively Na^+-poor intracellular fluid (illustrated by the membrane conductance curve for Na^+ in Figure 2). The resulting flood of Na^+ drives the voltage of the cell toward the Na^+ equilibrium potential (E_{Na}; roughly +55 mV), accounting for the rapid rising phase of the action potential (I).

This rapid depolarization of the membrane has two consequences that account for the declining phase of the action potential (II). The first is the inactivation of the Na^+ channel, which slows and eventually stops the ascent of V_m toward E_{Na}. The second is the delayed activation of voltage-sensitive K^+ channels, increasing the K^+ conductance of the membrane and allowing K^+ to diffuse down its concentration gradient from the K^+-rich intracellular fluid into the K^+-poor extracellular fluid. This diffusion of K^+ drives the membrane potential back down toward the K^+ equilibrium potential (E_K; roughly −75 mV) and even below the resting potential of the membrane.

Thus begins the final, afterpotential phase of the action potential (III), which is characterized by both the hyperpolarization of the membrane (i.e., V_m is

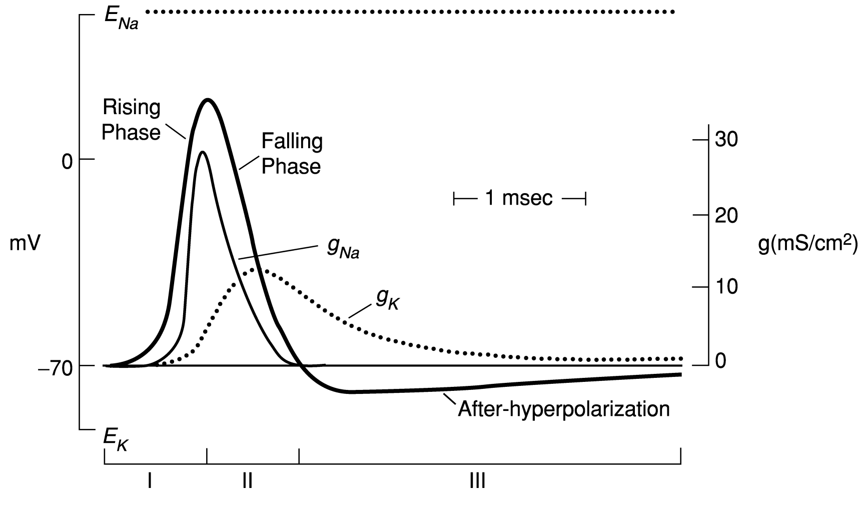


Fig. 2. The Action Potential.

lower than V_{rest}) and a period of reduced excitability. The membrane hyperpolarizes after the action potential because K^+ channels are slow to return to their resting closed state. The residual K^+ conductance tugs V_m away from V_{rest} and toward E_K.

The parts in the mechanism for generating action potentials are the membrane, the ions, and the ion channels. These parts are causally connected; they act and interact in regular ways to produce the action potential. These activities depend crucially upon the spatial organization of the components; ion channels *span* the membrane, allowing ion *movement* between the intracellular and extracellular fluids. Spatial organization is also fundamental to understanding the molecular mechanisms of channel activation and inactivation and for understanding the propagation of action potentials along axons. Yet, it is temporal organization that is most evident in the mechanism of the action potential; it is the relative orders and durations of the activation and inactivation of Na^+ and K^+ channels that explain the characteristic waveform (I, II, and III) of the action potential.

Levels of Mechanisms

Often mechanisms are nested within mechanisms. In such cases, some phenomenon (ψ) of a mechanism (M) is explained by the organized activities (ϕ) of lower-level components (X) that can themselves be taken as phenomena to be explained by the activities (ρ) of still lower level components (Z). Thinking about mechanisms provides a straightforward way to think about levels (see Craver 2001b). In this case, the relationship between lower and higher mechanistic levels is a compositional relationship with the additional restriction that the lower-level parts are components of (and hence organized within) the mechanism at the higher level. The requirement that lower-level parts be organized (at least spatially and temporally) within the higher-level mechanism distinguishes mechanistic levels from mere aggregates, such as piles of sand (Wimsatt 1986); from mere collections of improper parts, such as the cubes into which a television might be arbitrarily sliced (Haugeland 1998); and from mere inclusive sets, such as the collected songs of the Ramones. Lower mechanistic levels are entities and activities organized to exhibit the behavior of the mechanism as a whole.

Mechanistic levels should not be confused with intuitive ontic levels (e.g., Oppenheim and Putnam 1958), which map out a monolithic stratigraphy of levels across theories, sciences, and types of entities. Just as there are no mechanisms *simpliciter*, there are no mechanistic levels *simpliciter*. Mechanistic levels, instead, are defined only with respect to some highest-level mechanism M and its phenomenon ψ (pronounced "psi"). This, however, does not mean that the investigator cannot move upward, treating M as part of a yet higher level mechanism that generates its own phenomenon. Different levels of a mechanism involve different entities and activities. Accordingly, different vocabularies are typically used to describe mechanisms at different levels (Bechtel 1995). Exactly how many levels there are and how they are to be individuated are empirical questions that are answered differently for different phenomena.

Representing Mechanisms

There are many conventions for describing and representing mechanisms. Verbal accounts are generally insufficient to convey an understanding of a mechanism, especially if there are any nonlinearities in its behavior. Accordingly, verbal descriptions are often accompanied by diagrams representing the components, their activities (often depicted with arrows), and the relevant features of their organization (see Figure 1 above). Temporal relations are often represented spatially, either with labeled events conjoined by arrows or in separate frames. Diagrams afford the viewer the opportunity to follow through the parallel sequences of activities within the mechanism in one glance. With increasing frequency, the working of a mechanism may be represented in animated shorts. Extremely complicated mechanisms, however, frequently require the viewing time afforded by static two-dimensional representations so that aspects of the mechanism can be taken in piecemeal.

Descriptions of mechanisms, whether verbal or pictorial, may be more or less gappy, with holes or question marks to be filled in as details of the mechanism are discovered. Sometimes these are appreciated by the person portraying the mechanism, but many times the gaps are not even recognized until, for example, another component is discovered and researchers try to figure out what it contributes. Descriptions of mechanisms may also be more or less abstracted from the details of the operation of any particular mechanism, highlighting broad patterns of organization (e.g., with equations) or exhibiting precisely the spatial, temporal, and hierarchical organization of the components and activities of the mechanism.

Often the activities within a mechanism are characterized mathematically. For example, in describing the action potential mechanism, equations are advanced describing the changes in magnitude of Na^+ concentrations over time. Once such equations are developed, mathematical models of the overall operation of the mechanism can be advanced.

Mechanistic Explanations

Since mechanisms are often responsible for generating phenomena for which explanations are sought, it is not surprising that scientists frequently advance accounts of mechanisms as explanations. That is, to explain an action potential, they proceed much as in the example above—identifying the components of the responsible mechanism, describing the activities performed by the components, and showing how these components and activities are organized. They frequently present this information in diagrams, and often the account offered is gappy. Presenting a mechanism as an explanation, however, does not fit the standard deductive-nomological account of explanation, according to which explanation involves deriving a statement of the phenomenon to be explained from laws and relevant initial conditions. It is not laws that do the explanatory work but the account of the operation of the mechanism.

One might try to reconcile the two accounts of explanations by insisting that there is a law characterizing each mechanism. Typically, however, there is too much variability in a given mechanism (e.g., in the generation of action potentials in different neurons) for this to be plausible. It is better to recognize mechanistic explanation as an alternative model of explanation. Its prevalence in a variety of sciences such as physiology and neuroscience may account for the fact that these sciences have not been the primary source of examples of deductive-nomological explanation and have been relatively neglected by philosophers of science. Once mechanisms are recognized for their explanatory role in these sciences, though, we can also identify a number of philosophical issues to be pursued. One of these concerns their discovery.

How Are Mechanisms Discovered?

Characterizing the Phenomenon

One of the first tasks in discovering mechanisms is identifying the phenomenon—determining what it is that the mechanism does. The world does not come obviously prepackaged in terms of phenomena. How one characterizes the phenomenon critically affects how one goes about trying to discover the responsible mechanism and whether that quest will prove successful. Accordingly, characterizations of phenomena prove controversial and frequently are revised in the course of inquiry as one discovers that the mechanism does something different than one thought.

Phenomena are often subdivided, consolidated, or reconceptualized entirely as the discovery process proceeds. Researchers may recognize the need to subdivide a phenomenon into many distinct phenomena, as when learning and memory researchers were forced to recognize that there were many different kinds of memory requiring more or less distinct mechanisms to explain them. Alternatively, researchers may be forced to consolidate many different phenomena into a single phenomenon, as when it became understood that burning, respiring, and rusting were all due to a common mechanism and thus are examples of one phenomenon, oxidation. Finally, investigators may need to reconceptualize the phenomenon to be explained entirely. For example, early physiologists focused on the fact that animals burn foodstuffs and release heat. But after further investigation, researchers recharacterized this phenomenon as transforming energy into usable forms (e.g., ATP bonds).

Identifying Components

The discovery of mechanisms also involves identifying the components of the mechanism and their activities. Bechtel and Richardson (1993) used the term *decomposition* to describe analysis of a phenomenon into activities that, when properly organized, exhibit the phenomenon. In one of their main examples, they describe how the biological process of fermentation, over three decades of research, was decomposed into a set of more basic chemical reactions (oxidations, reductions, phosphorylations, etc.). This is *functional decomposition*. But frequently the process of decomposition begins by breaking the mechanism apart into component entities and only then investigating what the components do. This is *structural decomposition*. Ultimately, one measure of the adequacy of either form of decomposition is that it maps onto the other so that specific components are related to particular activities. Bechtel and Richardson call this identification of activities with components *localization*.

Often the search for the components of a mechanism is guided by an accepted store of components and activities that are reasonably well understood by a science at a particular time and that are available for use in thinking about how a mechanism works (Craver and Darden 2001). In the early stages of mechanism discovery, there may be no such store; there is either no idea of, or considerable controversy over, what the components and activities might be. The brain provides a useful example (Mundale 1998). There has been considerable controversy, for example, about what counts as a brain region, with different investigators using different criteria to divide the brain into parts at different times. Early attempts to map brain areas focused on the sulci and gyri resulting from the folding of cortex. Although prominent features of the brain, these tend not to be closely linked to component activities. With the identification of different types of neurons and the existence of cortical layers of varying thicknesses, numerous early-twentieth-century scientists, including Korbinian Brodmann, used these cytoarchitectural features to demarcate brain areas. Brodmann explicitly thought that different areas were likely to perform different operations, but he lacked any means for linking the regions he differentiated with function. More recent brain mappers have invoked yet additional criteria such as connectivity to other regions to identify brain areas. A major reason for controversy over these is that researchers are interested in components that perform the activities that generate the relevant phenomena. As the sulci and gyri of the brain illustrate, it is possible to differentiate structures within a mechanism that are not working components, that is, parts that carry out the relevant activities. In the relevant sense, these are not components of the mechanism. Similar challenges arise in functional decomposition—one may propose a decomposition into activities but not ones performed by any of the mechanism's components. Moreover, as hinted above, the search for components and for activities is interdeterminate—conceptions of the activities thought to be performed guide the identification of components, and vice versa.

How do scientists arrive at satisfactory decompositions that describe mechanisms in terms of their organized parts and activities? Often scientists begin the discovery process by proposing that there is a single component in the mechanism that alone is responsible for the phenomenon (e.g., attributing pleasure to activation of the brain's pleasure center). Sometimes this claim is correct, but even when it is, the task of identifying the mechanism that generates the phenomenon awaits decomposition of that component itself.

True decomposition is frequently guided either by the available store of components or by the available tools for investigating these components. Often scientists, functioning much like engineers, attempt to organize known components and activities in such a way that they might possibly produce the phenomenon. This process may involve reasoning analogically from other mechanisms (discovered in nature or human artifacts) and the activities performed in them. Such "how possibly" reasoning is, of course, fallible, since even two phenomena that are very similar may be generated by two very different mechanisms. In fact, sometimes the discovery process is slowed dramatically by pursuit of false leads generated by this engineering heuristic. On the other hand, even an erroneous proposal often advances the inquiry, since now experimental evidence can be generated that points to a more adequate decomposition. Experimental strategies for decomposing a mechanism are discussed further below.

Discovering the Organization of a Mechanism

Beyond delineating the phenomenon and revealing the components, a third major goal in the discovery of a mechanism is to determine how these components and activities are organized in the mechanism. Typically there are both spatial and temporal aspects to the organization of a mechanism. For example, the rate and duration of the phenomenon places time constraints on the activities of the components, and uncovering the order, rate, and duration of the steps in a mechanism often provides important clues into how the mechanism works. Likewise, discovering aspects of the spatial organization of a mechanism (the size, shape, position, orientation, etc., of the components) is often crucial for suggesting possible mechanisms and for ruling out others (see Craver and Darden 2001).

The relative importance of spatial and temporal organization varies from mechanism to mechanism. Spatial organization is of fundamental importance in, for example, the mechanisms of enzyme degradation because enzymes that can break down cellular substances need to be kept separate from other cellular substances that are not to be broken down. Spatial organization also helps provide efficiency in production mechanisms in which intermediate products are literally passed from one activity to the next (as in the Kreb's cycle).

If a phenomenon involves a change from one state or set of conditions to another (e.g., from glucose to alcohol, from sensory stimulus to recognition), it is common to think of that change as being executed by a linear sequence of steps. In part this is common because human conscious cognitive activities are serial—humans proceed from thinking of one thing to thinking of another. But, for very good reasons, such as ensuring proper regulation of a process, many natural mechanisms are not organized linearly. As a result, they are difficult for humans to conceptualize, at least without the aid of external representations such as diagrams in which one can represent backward as well as forward linkages.

The naturalness of linear organization means that in trying to fit multiple parts and activities together into a coherent description of a mechanism, researchers often begin by trying to organize them linearly. Often researchers begin to appreciate more complex modes of organization only when these attempts fail to account for the phenomenon. In modeling a chemical process, for example, one may find that there is no way to link together known basic reactions to get from the initial input to the product. This often leads to the exploration of more complex modes of organization such as a cycle. Thus, one common pattern in the process of discovering mechanisms is to begin with linear organization and then add complexity as required.

Experiments in Mechanism Discovery

Typically, the components, activities, and organization of a mechanism cannot be understood without the aid of well-designed experiments. Experimentation figures not just in the testing of models of mechanisms that have been hypothesized independently, but in the very process of discovering the mechanism.

Experimentation requires some means of intervening in the operation of the mechanism as well as a means of recording the effects of those interventions. Sometimes interventions into a mechanism are performed "by nature," through accidental damage, disease, or genetic mutation or variation. Other times the interventions are intentional and designed by the researcher to perturb some isolated aspect of the phenomenon or some component or activity in the mechanism.

A taxonomy of experimental approaches to developing and testing descriptions of mechanisms can be developed by focusing on where the intervention and recording techniques are applied (Bechtel and Richardson 1993; Craver 2001a). In the sense discussed above, a phenomenon and the mechanism that produces it are at two mechanistic levels, the phenomenal level (L_P) and the level of the mechanism (L_M) (see Figure 3). As illustrated in Figure 3, experiments may intervene and record entirely at the phenomenal level, bridge phenomenal and mechanistic levels, or intervene and record entirely within the mechanistic level.

First, both the intervention and the recording may be conducted at L_P without going down to L_M (see Figure 4). For example, one can intervene to vary the inputs to a mechanism or the conditions under which it operates (e.g., temperature) and record variations in the phenomenon. Much experimentation in cognitive psychology (e.g., requiring subjects to perform a task under varying conditions, such as cognitive load, and using reaction time as the measure of the effect) is of this sort and, when done well, can provide abundant information about the internal design of the mechanism. For example, evidence that two tasks interfere with each other provides further evidence that some component or components may be involved in both tasks. A great deal can also be learned about a mechanism by determining the range of input conditions under which it works properly and under which it fails or malfunctions.

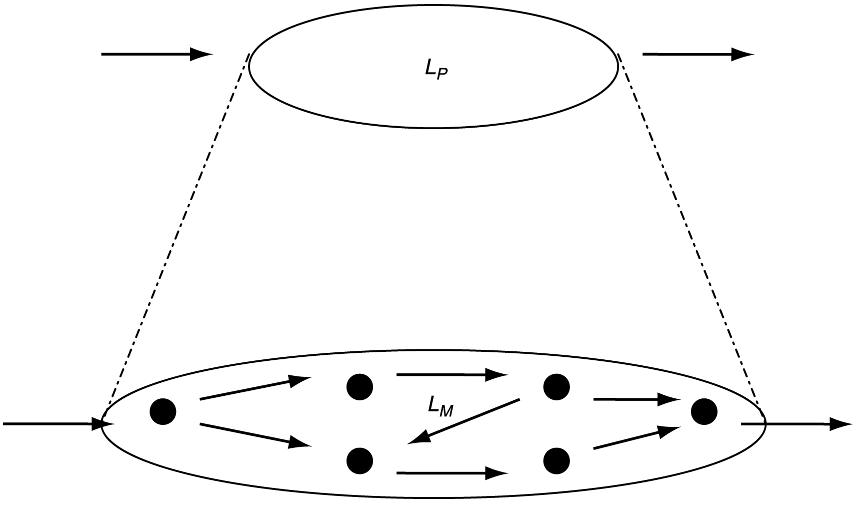


Fig. 3. Phenomenal Level (top) and Mechanism Level (bottom).

Second, experiments may bridge L_P and L_M. (Many experiments bridge several such levels at once.) Such experiments may be top-down (intervening at L_P and recording at L_M) or bottom-up (intervening at L_M and recording at L_P), and the experimental intervention may be either excitatory (somehow stimulating the target of the intervention) or inhibitory (somehow removing or impairing the target of the intervention). Top-down excitatory experiments are prevalent in cognitive neuroscience, where researchers intervene to engage an organism in some cognitive task while recording the activities of component brain regions, neurons, or molecules. Bottom-up excitatory experiments are also common. Neural stimulation studies, for example, use electrodes to excite individual neurons, and the effects are recorded for the cognitive phenomenon in which those neurons are involved. Additionally, bottom-up inhibitory experiments are a staple of most sciences that search for mechanisms. In neuroscience, for example, one may intervene to remove a brain region, a receptor molecule, or a neurotransmitter and record the effects on the phenomena in which those components are putatively involved. It is not uncommon for researchers to find a way to impair the activity before they figure out what the relevant components are or which are being affected. For example, one can discover a chemical poison that impairs a metabolic process but not know what component of the mechanism the poison is acting upon.

Third, inhibitory and excitatory techniques can also be applied within L_M. In this case, one intervenes to excite or inhibit some component or activity in the mechanism and then records the results of that intervention elsewhere in the mechanism. This form of experiment is especially important for determining how the components of the mechanisms are organized together in the production of the phenomenon.

There are significant epistemological challenges in interpreting the results of excitatory and inhibitory interventions into the working of the mechanism. Bottom-up inhibitory experiments may be foiled by redundancy, reorganization, and failures of specificity in the intervention. Intervention to remove or inhibit a component or activity may result in little or no change to the phenomenon if the removed or inhibited component is redundant (like the human kidney). Likewise, the mechanism may reorganize in the face of a loss of its component, leaving the phenomenon intact or only mildly transformed. In general, in removing a part of a mechanism and observing the behavior of the mechanism as a whole, researchers learn not what the removed part does but rather what the rest of the mechanism can do in its absence. Finally, the intervention may have nonspecific effects on other components in the mechanism, thereby indirectly altering the phenomenon and foiling the inference from the recorded changes to the function of the inhibited part. This problem is often exacerbated in ''natural

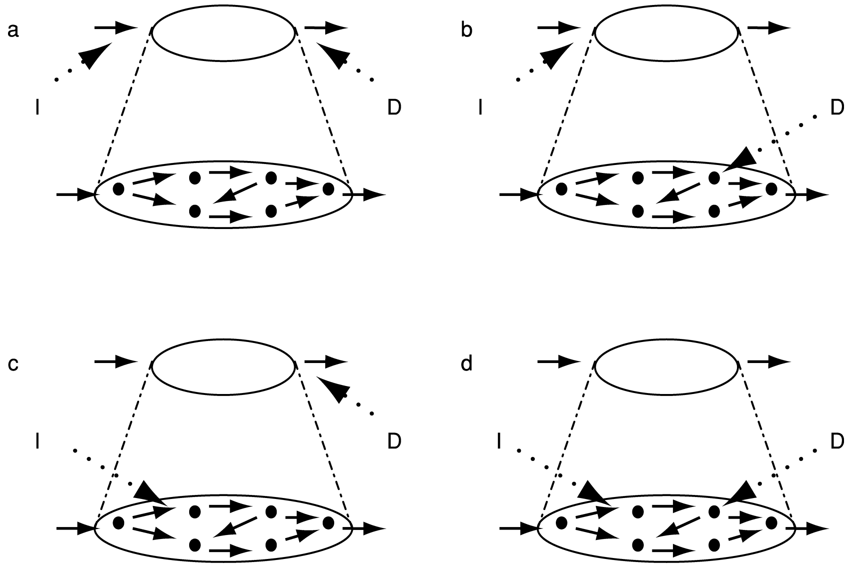


Fig. 4. Points of intervention and recording in experiments. Experiments may (a) both intervene and record at the phenomenal level, (b) intervene at the phenomenal level and record at the mechanistic level, (c) intervene at the mechanistic level and record at the phenomenal level, or (d) both intervene and record at the mechanistic level.

experiments" in which the intervention has not been tightly controlled by an investigator and so may have had a rather nonlocal impact on the components of the mechanism.

Similar epistemological difficulties attend the use of top-down excitatory experimental strategies. One example of such an experiment is to provide a stimulus to an organism and record from individual neurons in its brain or to use neuroimaging to record where there is increased blood flow in the brain. The epistemic challenges here are no less than when the intervention is within the mechanism. Activity in a part of a mechanism when the whole mechanism has been stimulated shows only that the component in question does respond to the stimulation. It does not yet show what activity it performs. Many neurons in the brain, for example, will respond when the organism is presented with a visual stimulus. One can gain more of a clue as to what a component is contributing by varying the intervention and determining the range of interventions to which the component is responsive (e.g., that it is only responsive to visual stimuli moving to the left). Even so, a given active neuron may perform an activity that is largely incidental to the phenomenon one is investigating (e.g., how objects are identified).

One way investigators begin to acquire confidence in their physical and functional decompositions is by drawing upon multiple modes of investigation, especially by invoking both inhibition and excitatory interventions. If lesioning a component eliminates the phenomenon of interest and exciting it produces the phenomenon, compelling evidence is provided that the component figures in generating that phenomenon. But just what does it contribute? Often answering that question depends on formulating a hypothesis about what many different components are contributing, and developing an account of how the components together produce the phenomenon. For example, researchers working on how the brain recognizes objects identified different brain regions in which individual cells would respond to different aspects of a stimulus—some responded whenever a given color was present, another when a given shape was present, and yet others when a particular object was present. By also knowing how these various brain regions were connected to each other, researchers began to piece together an account of the overall mechanism (Bechtel 2001).

As researchers reach the stage of reasonably worked out hypotheses about what different components contribute, additional tools can be invoked to help figure out the mechanism. For example, researchers often begin to build models, including computational ones, that characterize what each component is thought to contribute and to simulate their interaction. To the degree that the model predicts the phenomenon, one acquires confidence that one's account is at least close to correct. (The fit between a model and the phenomenon is often a matter of degree, and the degree of fit deemed sufficient often changes as research on the mechanism proceeds.) But failures are equally informative, since they often lead researchers to posit yet unidentified components and activities and begin to seek evidence for them.

Not surprisingly, there is no foolproof procedure for discovering mechanisms. But there are a range of strategies that can be identified by careful examination of actual science.

Conclusion

Four aspects of mechanisms have been identified: (i) the phenomenal, (ii) the componential, (iii) the causal, and (iv) the organizational. The generation of a phenomenon is often the product of a mechanism, and describing the mechanism provides an explanation of the phenomenon. The sciences concerned with identifying mechanisms have developed a variety of conceptual and experimental tools for this purpose. The philosophical analysis of mechanisms and their discovery is still in a relatively early stage but has advanced far enough that it is safe to predict that careful attention to mechanisms and mechanistic explanation is likely to yield significant advance in the philosophical understanding of science.

CARL CRAVER
WILLIAM BECHTEL

References

Bechtel, William (1995), "Biological and Social Constraints on Cognitive Processes: The Need for Dynamical Interactions Between Levels of Organization," *Canadian Journal of Philosophy* 20 (supplement): 133–164.

——— (2001), "Decomposing and Localizing Vision: An Exemplar for Cognitive Neuroscience," in Bechtel, Pete Mandik, Jennifer Mundale, and Robert S. Stufflebeam (eds.), *Philosophy and the Neurosciences: A Reader*. Oxford: Basil Blackwell, 225–249.

Bechtel, William, and Robert C. Richardson (1993), Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research. Princeton, NJ: Princeton University Press.

Craver, Carl (2001a), "Interlevel Experiments and Multilevel Mechanisms in the Neuroscience of Memory," *Philosophy of Science* 68 (supplement): S83–S97.

——— (2001b), "Role Functions, Mechanisms, and Hierarchy," *Philosophy of Science* 68: 53–74.

Craver, Carl, and Lindley Darden (2001), "Discovering Mechanisms in Neuroscience: The Case of Spatial Memory," in Peter K. Machamer, Rick Gush, and Peter

McLaughlin (eds.), *Theory and Method in Neuroscience*. Pittsburgh: University of Pittsburgh Press.

Glennan, Stuart (1996), "Mechanisms and the Nature of Causation," *Erkenntnis* 44: 49–71.

Haugeland, John (1998), *Having Thought*. Cambridge, MA: Harvard University Press, chap. 10.

Kauffman, Stuart A. (1971), "Articulation of Parts Explanation in Biology and the Rational Search for Them," in R. C. Bluck and R. S. Cohen (eds.), *PSA 1970*. Dordrecht, Netherlands: Reidel.

Machamer, Peter, Lindley Darden, and Carl Craver (2000), "Thinking about Mechanisms," *Philosophy of Science* 67: 1–25.

Mundale, Jennifer (1998), "Brain Mapping," in William Bechtel and George Graham (eds.), *A Companion to Cognitive Science*. Oxford: Basil Blackwell, 129–139.

Oppenheim, Paul, and Hilary Putnam (1958), "Unity of Science as a Working Hypothesis," in Herbert Feigl, Grover Maxwell, and Michael Scriven (eds.), *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 3–36.

Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

Wimsatt, William (1986), "Forms of Aggregativity," in A. Donagan, A. Perovich, and M. Wedin (eds.), *Human Nature and Natural Knowledge*. Dordrecht, Netherlands: Reidel, 259–291.

See also **Explanation; Cognitive Science; Neurobiology; Reductionism**

METHODOLOGICAL INDIVIDUALISM

Methodological individualism (MI) is a set of related but distinct theses about how the social sciences should proceed. Drawing inspiration from the atomist program in natural science, MI holds that all social explanation should be in terms of individuals. Classic figures in the history of the social sciences such as Weber advocated some version of the doctrine, which is espoused by major schools of thought in economics and elsewhere in the social sciences (Blaug 1992; Gordon 1991). It is opposed by the holist tradition that began with the founders of sociology (Durkheim 1965) and continues to this day throughout the social sciences. While it has been usually treated as a conceptual or philosophical thesis, MI is probably best thought of as a series of more or less empirical theses.

Most individualist claims can be classified under one of the following four types (Kincaid 1996 and 1997):

1. Ontological claims:
   - (i) Societies are composed of individuals.
   - (ii) Societies do not act independently of individuals.
   - (iii) Social entities do not exist.
2. Claims about theory reduction:
   - (iv) Any social theory is in principle reducible to a theory referring entirely to individuals. (see Reductionism)
3. Claims about explanation:
   - (v) Theories referring only to individuals can fully explain all social phenomena.
   - (vi) Individualist mechanisms are a necessary condition for social explanation. (see Explanation)
4. Claims about confirmation:
   - (vii) No social theory without individualist mechanisms can be well confirmed; and
   - (viii) Searching for individualist theories is the best route to successful social science (see Confirmation Theory)

There are, of course, various real and alleged interconnections among these claims.

Perhaps the most common and logically central claim is the reductionist one. To reduce one theory to another is to show that the reducing theory can do all the explanatory work of the reduced theory, which is only a special case of the reducing theory. Since different theories have different vocabularies, theory reduction requires systematic linkages between the categories of the two theories. Individualism in its reductionist guise thus claims that the concepts of the social sciences can be equated with descriptions of individual behavior in such a way that all social science explanations can be put in individualist terms.

Many have claimed that this reductionist thesis follows from the ontological truism that society is made of and does not act independently of individual people (Watkins 1973). That conclusion does not follow, however. Chairs are made of molecules, but there are so many ways of making a chair that it is unlikely that a systematic connection will be found that would allow the replacement of the category 'chair' with some molecular description, however complex. Social entities such as corporations or states may have a similarly loose relation to the individual behaviors constituting them. Whether that is the case is an empirical issue that is unlikely to have any general answer across all domains of social research.

While the ontological claim that society does not exist and act independently of individuals is quite plausible, the presumed failure of theory reduction mentioned above argues against the further ontological claim that societies do not exist. If there are successful theories that make essential reference to social entities, that is one good reason to countenance their existence.

Of course the methodological individualist might deny the assumption that social explanations—explanations in terms of social entities—ever succeed in the first place. Two versions of this claim were identified above: No explanation is adequate unless it is entirely in terms of individuals (claim [v]) and some reference to individuals is necessary (claim [vi]). The latter, logically weaker claim is usually put as a claim about mechanisms: explanations in terms of social entities are only acceptable if the individualist mechanisms bringing them about are provided (see Mechanism).

To be interesting, these theses should be independent of assertions about theory reduction. It is unclear, however, that the stronger of the two is independent. If an individualist theory can fully explain everything that a nonindividualist theory can, then it can state those explanations only if it can relate the categories of the nonindividualist theory in every case to its own. But this goes back to the requirements for reduction.

The claim that individualist mechanisms are needed for explanation is widespread (Elster 1985; Little 1989). It is often defended as an instance of the general truth that all explanations require citing the mechanism involved. The general principle is implausible and of dubious value. It is implausible because one seemingly explains one macroscopic physical event by a previous one without any idea of the mechanism—"the crash of the plane caused the collapse of the building" is explanatory even without describing how it did so. The general principle is of dubious value because 'mechanism' is ill-defined. There can be mechanisms at many different levels of detail, so a demand for mechanisms is ambiguous. Individualists need to show that the mechanism cannot be social—as when competition between firms is cited to explain prices—and that it must be about individuals instead of neurons or genes.

Even if individualist mechanisms are needed to explain, this may not be much of a victory for methodological individualism. The mechanisms in question are likely to invoke the role individuals play and the constraints they face in institutions. Many game-theoretic and rational choice accounts provide individualist explanations, but they take norms, rules of the game, institutional constraints, and other such nonindividualist explainers as givens (see Game Theory).

Another motivation for individualist mechanisms is confirmation. The basic idea is that mechanisms are needed to rule out confounding cases. Aside from the question of levels raised above, this rationale suffers from the obvious fact that scientific claims can be confirmed without mechanisms. Newton's laws of motion were predictively very successful for a long period of time without any account of the underlying cause of gravity; Darwin had a similar success in describing evolution with an incorrect mechanism of genetic transmission. No doubt mechanisms play a useful role in confirming and explaining. But when, where, and at what level of detail they are useful is an empirical question that depends on the context.

HAROLD KINCAID

References

Blaug, M. (1992), *The Methodology of Economics*. Cambridge: Cambridge University Press.

Durkheim, Emile (1965), *The Rules of the Sociological Method*. New York: Free Press.

Elster, Jon (1985), *Making Sense of Marx*. Cambridge: Cambridge University Press.

Gordon, S. (1991), *The History and Philosophy of the Social Sciences*. London: Routledge.

Kincaid, Harold (1996), *Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research*. Cambridge: Cambridge University Press.

——— (1997), *Individualism and the Unity of Science: Essays on Reduction, Explanation, and the Special Sciences*. Lanham, MD: Rowman and Littlefield.

Little, Daniel (1989), *Understanding Peasant China*. New Haven, CT: Yale University Press.

Watkins, John (1973), "Methodological Individualism: A Reply," in John O'Neill (ed.), *Modes of Individualism and Collectivism*. London: Heineman, 179–185.

See also **Confirmation Theory; Decision Theory; Economics, Philosophy of; Explanation; Game Theory; Mechanism; Reductionism; Social Sciences, Philosophy of**

MIND-BODY PROBLEM

See **Consciousness; Intentionality; Physicalism; Supervenience**

MOLECULAR BIOLOGY

The term 'molecular biology' was introduced by Warren Weaver in 1938 in an internal report of the Rockefeller Foundation: "And gradually there is coming into being a new branch of science—molecular biology . . . in which delicate modern techniques are being used to investigate ever more minute details of certain life processes" (as quoted in Olby 1974, 442). What Weaver may have only dimly foreseen is that these new techniques would ultimately transform the practice of biology in a way comparable only to the emergence of the theory of evolution in the nineteenth century. At the beginning of the twenty-first century, molecular biology has become most of biology, either *constitutively,* insofar as biological structures are characterized at the molecular level as a prelude for further study, or at least *methodologically,* as molecular techniques have become a preferred mode of experimental investigation of a domain. Recent biological work at the organismic and lower levels of organization (cytology, development, neurobiology, physiology, etc.) increasingly fall under the former rubric. Work in demography, epidemiology, and ecology falls under the latter, with ecology perhaps being the subdiscipline within biology that has most resisted molecularization. Work in evolution falls under both: constitutively, when the evolution of molecules and molecular structures forming organisms is studied for its own sake, and methodologically, when molecular techniques (most notably, DNA sequencing) are used to reconstruct evolutionary history. This article will be largely restricted to the constitutive aspect of molecular biology, since that is what has so far (perhaps deservedly) commanded most philosophical attention.

The decade following Weaver's introduction of molecular biology saw the steady increase in the use of "delicate" molecular techniques, in particular, x-ray crystallography, to study biological macromolecules "minutely," increasingly with an emphasis on proteins. The central problem was the elucidation of the three-dimensional structures (the relative positions of the atoms) of biological macromolecules. The structure of proteins was supposed to explain their behavior. Proteins were singled out because they were believed to be the most important of these macromolecules. In particular, since the establishment of biochemistry as a discipline in the 1920s, enzymes and their interactions had been held to be the key to understanding *metabolism* (the catchall term for the complex chemical reaction systems that characterize life). All enzymes are proteins. Until the early 1940s, it was believed that the hereditary material (the genes) was also likely to be proteins. The nucleic acids, constructed out of only four nucleotide base types (adenosine [A], cytosine [C], guanine [G], and thymine [T]), were believed to be insufficiently complex to be able to specify the immense variety of known genes.

However, experimental work starting in the early 1940s showed that the hereditary substance—specifying 'genes' (see Genetics)—was deoxyribonucleic acid (DNA). Attention then shifted to deciphering the physical structure of DNA, a problem that was solved by Watson and Crick (1953) with their double-helix model. The construction of this model and its subsequent confirmation marks a development of signal importance for modern biology (Sarkar 2005, Ch. 1). It ushered in the "classical" age of molecular biology (see the next two sections) with an intriguing informational interpretation of biology (see Biological Information). Important conceptual innovation also came from Monod and Jacob in the early 1960s, who constructed the allostery model to explain cooperative behavior in proteins and the operon model of gene regulation (Monod 1971; Jacob 1973; see below). Genes were interpreted as DNA sequences either specifying proteins (the *structural* genes) or controlling the action of other genes (the *regulatory* genes). Perhaps the most important development in classical molecular biology was the establishment of a genetic "code" delineating the relation of DNA sequences to amino acid residue sequences in proteins. (Both DNA and protein are linear molecules in the sense that they consist of units connected in a chain through strong [covalent] chemical bonds.) Gene *expression* took place by the *transcription* of DNA to ribonucleic acid, RNA, at the chromosomes (in the nucleus), and the *translation* of these transcripts into protein at the ribosomes (in the cytoplasm). The one gene–one enzyme credo of classical genetics was transformed into the one DNA segment–one protein chain credo of molecular biology (see Genetics).

Crucial to the program of molecularizing biology was the expectation—first explicitly stated by Waddington (1962)—that gene regulation explained tissue differentiation and, ultimately, morphogenesis in complex organisms. Genetic reductionism, the thesis that genes alone can explain organismic features, long predates molecular biology (Sarkar 1998). However, the molecular interpretation of the gene allowed the general explanatory success of molecular biology to be co-opted as a success of molecular genetics. In such a context, Waddington's thesis was positively received and helped usher in an era dominated by *developmental genetics,* according to which organismic development was to be understood through the action of genes. Mayr (1961) and others introduced the metaphor of the genetic program to characterize the putative relation between genomic DNA and organismic features. As molecular genetics began to dominate the research agenda of molecular biology in the 1970s, the emergence of organismic features came to be viewed as determined by "master control genes" (Gehring 1998). This view was initially supported by the demonstration that some DNA sequences (such as the homeobox) were conserved across a wide variety of species. DNA came to be viewed as the molecule "defining" life, a view that helped initiate the massive genome sequencing projects of the 1990s, which were supposed to produce a gene-based complete biology that delivered on all the promises of molecular developmental genetics. In general, because of the presumed primacy of DNA in influencing organismic features, starting in the early 1960s, molecular genetics began to dominate research in molecular biology.

Genetics and development were the earliest biological subdisciplines to be redefined by molecular biology. In the case of evolutionary biology, as early as the 1950s, Crick (1958) pointed out that the genotype/phenotype relation could be reinterpreted as the relation between DNA and protein, with proteins constituting the subtlest form of the expression of a phenotype of an organism. Consequently, the evolution of proteins (and, later, DNA sequences), especially the question of what maintained their diversity within a population, became a topic of investigation—in the 1960s, these studies led to the neutralist challenge to the received view of evolution (see Evolution). More importantly, changes at the level of DNA sequences, provided that these were selectively neutral, permitted the construction of a "molecular clock" that could be used to reconstruct evolutionary history more accurately than could be achieved by traditional morphological methods (see Population Genetics).

Meanwhile, biochemistry and immunology were reconstituted by the new molecular biology in ways that were not unexpected. That enzyme interactions and specificity would be explained in molecular terms was no surprise (see "Classical Molecular Biology" below). However, immunological specificity was also believed to be explainable by the same mechanism. This model of immune action was coupled to a selectionist theory of cell proliferation to generate the clonal theory of antibody formation, which combined molecular and cellular mechanisms in a novel fashion (see Immunology). In both biochemistry and immunology, what was largely at stake was the development of models that could explain the observed specificity of interactions: Enzymes reacted with only very few substrates; antibodies were highly specific to their antigens.

By the late 1970s, it became clear that the simplicity of the picture of genetics inherited from the 1960s was being lost. The initial picture was generated from an exploration of the genomes of prokaryotes (single-celled organisms without a nucleus), especially the bacterium *Escherichia coli.* In prokaryotes, every piece of DNA has a structural or regulatory function. In the 1970s, it was discovered that the genetics of eukaryotes (organisms with cells with nuclei) turned out to have an unexpected complexity. In particular, large parts of the genomic DNA sequences apparently had no function: These segments of "junk" DNA were interspersed between genes on chromosomes and also within genes. After RNA transcription, noncoding segments within genes were *spliced out* before translation. Gene regulation in eukaryotes was qualitatively different and more complicated than in prokaryotes. Some organisms used nonstandard genetic codes and other alternatives (see Sarkar 1996 for a detailed account.)

Subsequent work in molecular biology has only added to this picture of complexity, so much so that it is reasonable to suspect that the classical picture is breaking down. RNA transcripts are subject to *alternative splicing,* with the same DNA gene corresponding to several proteins. RNA is edited, with bases added and removed, before translation at the ribosome, to such an extent that it is sometimes difficult to maintain that some gene actually does code for a given protein. There is no obvious relation between the number of genes in an organism and its morphological or behavioral complexity. Most importantly, it now appears that a fair amount of the DNA thought to be junk is transcribed into RNA though not translated. Thus, presumably, much of the so-called junk DNA is functional, though the nature of these functions remains controversial.

This article will concern both classical molecular biology and the postgenomic molecular biology of the modern era. It will not only discuss issues in the philosophical interpretation of the classical era, which are fairly well characterized, but also include more speculative discussions of issues raised by recent developments.

Classical Molecular Biology

Classical molecular biology can be viewed in continuity with both the genetics and the biochemistry of the era that preceded it. From biochemistry—in particular, the study of enzymes in the 1920s and 1930s—early molecular biology inherited the mechanistic proposal that the function or behavior of biological molecules was "determined" by its structure, an idea that went back to Ehrlich's "side-chain" theory in the late nineteenth century. In the 1950s, structural modeling of biological macromolecules, especially proteins, was pioneered by Pauling and his collaborators using data from x-ray crystallography (see, e.g., Pauling and Corey 1950). By the early 1960s, a handful of such structures were fully solved. These structures, along with the structure of DNA, seemed to confirm the hypothesis that structure explains behavior. Perhaps more surprisingly, it was found that structural interactions seemed to be mediated entirely by the shape of active sites on molecules and that the sensitive details of structure and shape were maintained by very weak interactions.

These experimental observations led to four seemingly innocuous rules about the behavior of biological macromolecules, which in the 1960s and 1970s formed the theoretical core of molecular biology (Sarkar 1998, 149–150):

1. The weak interactions rule: The interactions that are critical in molecular processes are very weak.
2. The structure-function rule: The behavior of biological macromolecules can be explained from their structures, as determined by techniques such as crystallography.
3. The molecular shape rule: These structures, in turn, can be characterized entirely by molecular size, external shape (especially), and some general properties (such as hydrophobicity) of the different regions of the surfaces;
4. The lock-and-key fit rule: In molecular interactions, molecules interact only when there is a lock-and-key fit between the two molecular surfaces. There is no interaction when these fits are destroyed.

A lock-and-key-fit thus based on shape is an obvious way of achieving stereospecific capacity, thus resolving the critical problem for classical molecular biology. Because they are most intimately involved in the explanation of specificity, the molecular shape and lock-and-key-fit rules are the most important in this respect. In what follows, these will be called the rules of classical molecular biology.

In the 1960s and 1970s, these rules were deployed with remarkable success. As noted earlier, enzymatic and immunological interactions were among those that were immediately brought under the aegis of the new molecular biology. Two other cases are even more philosophically interesting:

- The allostery model explains why some molecules such as hemoglobin show *cooperative* behavior. In the case of hemoglobin, there is a nonlinear increase in the binding of oxygen after binding is first initiated. This is explained by conformational—shape—changes in the molecular subunits of hemoglobin; and

- The operon model explains *feedback*-mediated gene regulation in prokaryotes: The presence of a substrate activates the production of a protein that interacts with it, and its absence inhibits that production (see Monod 1971 for an accessible accurate account of these two examples and a conceptual summary of theoretical reasoning in early molecular biology).

Both cooperativity and feedback phenomena formed part of the traditional repertoire of holists in biology (see the next section, which will discuss the philosophical significance of the success of such structural explanation in molecular biology).

However, the 1950s also saw the elaboration of a radically different model of biological specificity, based on the concept of *information*, which was introduced into genetics only in 1953 (Sarkar 1996). This concept soon came to play a foundational role in molecular genetics. DNA was supposed to be the repository of biological information, a genetic "program" was supposed to convert this information into the adult organism, and new information was supposed to result from random mutation (and be maintained by selection) and never incorporated into the genome from the environment. Crick (1958) enshrined these assumptions in what he called the central dogma of molecular biology:

> This states that once "information" has passed into protein *it cannot get out again*. In more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible. (153) (emphasis in original)

Information, according to Crick, was the sequence of nucleotide bases in DNA or the sequence of amino acid residue in protein molecules. Note the contrast here with the stereospecific physical model of specificity. The dogma has continued to be an important regulative principle of molecular biology in the sense that it is presumed for further theoretical reasoning: Whether it survives recent developments will be discussed later in this essay.

However, the complexities of eukaryotic genetics, as discovered in the 1970s and 1980s, already began to challenge the central dogma (but see Thiéffry and Sarkar 1998). Much of this work was made possible by the development of technologies based on the polymerase chain reaction in the 1980s. There were five salient discoveries that challenged the simple picture inherited from prokaryotic genetics (Sarkar 2005, Ch. 8) (see Genetics):

1. The genetic code is not fully universal, the most extensive variation being found in mitochondrial DNA in eukaryotes. However, there is also some variation across taxa (see Fox 1987 for a review).
2. DNA sequences are not always read sequentially in blocks. There are overlapping genes, genes within genes, and so on (Barrell, Air, and Hutchison 1976). Thus, two or more different proteins could be specified by the same gene.
3. As noted earlier, not all DNA in the genome is functional. Intervening sequences—within and between structural genes—must be spliced out from transcripts (Berget, Moore, and Sharp 1977; Chow et al. 1977). This discovery helped resolve the so-called C-value paradox (Cavalier-Smith 1978), that is, the absence of any obvious correlation between the size of the genome and the morphological and behavioral complexity of an organism.
4. The same transcript may be spliced in different ways (Berk and Sharp 1978). One consequence of such alternative splicing is that, as with overlapping genes, two or more different proteins could be specified by the same gene.
5. Besides splicing, RNA is sometimes subject to extensive editing before translation at the genome (Cattaneo 1991).

These developments have led to skepticism of the relevance of the coding model of the DNA/protein relationship and of the informational model of specificity (see the next section). Though philosophers (and some biologists) have been slow to recognize this, the credo of one DNA segment–one protein chain has long become irrelevant in molecular biology. The modern era presents even more significant challenges, as later sections of this essay will underscore.

Philosophical Interpretations

Philosophy of biology only emerged as a recognizable part of philosophy of science only in the late 1960s. In the early years, considerable attention was paid to molecular biology, especially with respect to the issue of reductionism, but starting in the late 1970s, attention within philosophy of biology began to be concentrated solely on evolutionary theory, much to the detriment of the field. Attention shifted back to molecular biology in the 1990s, with some work now being done on the question of biological information besides reductionism. Since then, classical molecular biology has been increasingly scrutinized by philosophers, though not as much as it deserves. This section will focus on reduction and information. However, important philosophical work has also been done on other forms of conceptual change in molecular biology and, lately, experimentation in the field (Culp 1995; Rheinberger 1997).

Reduction

The first question about molecular biology that interested philosophers was whether it could be interpreted as a reductionist enterprise in the same way as the kinetic theory of matter was reductionist within classical physics (see Reductionism). The model of reduction then in vogue was due to Nagel (1961) with some modification by Schaffner (1967): It viewed reduction as a deductive-nomological explanation but with the reduced laws as the explananda (see Explanation; Nagel, Ernest). The debate soon centered on the question of whether molecular genetics was reducing or replacing Mendelian genetics. While Schaffner (1967) made the case for successful reduction, this position was attacked by Hull (1972) on the grounds that molecular biology did not have laws and theories (as logical empiricists envisioned those entities). Subsequently, an antireductionist consensus developed (also influenced by Kitcher [1984]).

This consensus was subsequently challenged by Sarkar (1989 and 1998), Waters (1990), and others, but only by rejecting the Nagel-Schaffner formal model as being relevant to substantive questions about reduction. (Even earlier, Wimsatt [1976] had argued against the relevance of the Nagel-Schaffner model.) In these analyses, what is at stake is that properties of wholes are being explained by properties of parts interacting locally. The allostery and operon models are philosophically critical exemplars of this approach because the former explains cooperativity and the latter feedback, both of which formed part of the conceptual repertoire of traditional holists (see Emergence). Enzymatic and immunological specificity provide more mundane examples. However, most of these cases are much simpler than that of providing fully reductionist explanations of quintessentially Mendelian genetic phenomena such as the segregation or assortment of alleles. In these cases—central to the question of reducing Mendelian genetics to molecular genetics—reductionist explanation remains piecemeal and, in many ways, incomplete. However, there is every reason to believe that the relevant lacunae will be filled without requiring new conceptual or theoretical resources.

Nevertheless, even during the classical era, a few anomalies remained, though none serious enough to call into question the viability of the reductionist project. In particular, there has never been a successful parts-whole account of dominance (that is, the dominance of one trait or allele over another) (see Genetics). There is some reason to believe that explaining dominance at the molecular level will require appeal to topological properties of networks, but such a move would take explanation beyond the reductionist realm (see "Philosophical Speculations" below).

Information

Though it is commonplace to talk of biological information, no successful formal definition of the concept in the context of molecular biology has ever been given. Because of difficulties that the concept of information encountered in the late 1980s and 1990s, this failure led Sarkar (1996) to suggest that information in molecular biology was a metaphor masquerading as a theoretical concept (Griffiths 2001) (see Scientific Metaphors). For Crick (1958), information consisted of sequences, of DNA or protein. Informally, this is what 'information' is probably taken to mean in most contexts. The first point to note is that any such definition would require that the concept of information being used *not* be Shannon's (1948) communication-theoretic notion of information, which requires the estimation of the frequency of symbols drawn from a set. Thus, mathematical information theory based on Shannon's concept simply becomes irrelevant in this context (for a contrary position, see Yockey 1992). At the very least, any usable concept of biological information must refer to individual sequences and be symbolic, semantic, or semiotic, in the sense that it must capture the idea that the sequence is a "sign" for something else (Sarkar 2005, Ch. 10). As such, it must account for biological specificity.

The concept was central to two related theoretical interpretations within molecular biology:

(a) that the DNA/protein relation is a genetic code, typically extended to suggest that all phenotypic traits are encoded in the DNA of the genome; and
(b) that the genome constitutes a genetic program for the organism.

As discussed earlier, developments within eukaryotic genetics began to limit the scope of the genetic code in the 1970s. Any claim of the existence of a genetic program at the very least constitutes a claim of genetic reductionism, and at the very worst a claim of genetic determinism. Genetic reductionism must be clearly distinguished from physical reductionism (the physical explanation of properties of wholes from properties of parts, which was discussed earlier). Genetic reductionism is the claim that organismic features are satisfactorily explained by appeal to properties of genes or DNA (without recourse to properties of other molecules). It is, for instance, central to the project of developmental genetics. Such a reductionism was never very plausible; consequently, the metaphor of a genetic "program" was always troubled (Keller 2000). Nevertheless, the "program" metaphor was quite influential during the heyday of developmental genetics. As the next two sections will underscore, it does not survive even in a mitigated form in the postgenomic era. The failure of genetic reductionism makes any stronger claim of genetic determinism irrelevant.

Viewing information as sequence, Crick (1958) also proposed the *sequence hypothesis:* that the sequence of amino acid residues in a protein (also called its *primary* structure) determines its three-dimensional conformation (also called its *tertiary* structure). Attempting to show how this comes about came to be called the protein folding problem. It has never been successfully solved, and not for lack of effort (Sarkar 1998). Moreover, for many proteins, it is known that sequence alone is insufficient for specifying three-dimensional conformation. It may even be the case that the same sequence can lead to several different conformations. This failure casts additional doubts on the utility of the concept of biological information stored in the genome, at least in the sense Crick intended it. Even if the genetic code were as exceptionless and predictively successful as was believed in the 1960s, all that it would allow is the inference of an amino acid residue sequence from the DNA. If the protein sequence does not determine its conformation, ipso facto, the DNA sequence cannot. It follows that the information within the DNA cannot specify phenotypes even further removed from the genome.

To the extent that the genetic code still remains useful, a proper explication of the concept of biological information remains an unaccomplished philosophical task of some importance (see Biological Information).

The Modern Era

By the "modern era" of molecular biology is meant the period beginning with the production of large genomic sequences in the 1990s. It is also referred to as the Genomic, Postgenomic, or, less accurately, Proteomic era ("proteomic" is less accurate because, to date, there has been limited progress in proteomics; see below). What marks this era is the study of large genomic sequences, and not individual alleles that had been previously identified by their phenotypic effects.

Genomics and Postgenomics

Genomics was ushered in by the decision to sequence the entire human genome as an organized project (the Human Genome Project [HGP]), involving a large number of laboratories in the late 1980s. Subsequently, similar projects were established to sequence the genome of many other species. To date, genomes of over 150 species have been sequenced. Almost every month sees the announcement of the completion of sequencing for a new species. The sheer volume of sequence information that has been produced has spawned a new discipline of "bioinformatics" dedicated to the computerized analyses of biological data.

When the HGP was first proposed, there was considerable controversy among biologists about its wisdom (Tauber and Sarkar 1992; Cook-Deegan 1994). There were:

(i) doubts about its ability to deliver on the bloated promises made by proponents of its scientific and, especially, medical benefits;
(ii) questions whether such organized "Big Biology" projects were wise science policy because of their potential effect on the ethos of biological research; and
(iii) worries that society would be legally and medically ill-prepared to cope with the results of sequencing that came too rapidly, in contrast to the normal slower accumulation of human genomic sequence information. It was feared that legislation protecting genetic privacy and preventing genetic discrimination would not be in place; there would be a shortage of genetic counselors; and so on.

In one important respect, the critics were correct: There have been few immediate medical benefits from the HGP, and no significant such innovation seems forthcoming. Instead, recent work underscores the importance of gene/environment interactions that critics had routinely invoked to criticize the claims of the HGP (see Heredity and Heritability). However, in another sense, even the most acerbic critics should now accept that the scientific results of the sequencing projects, taken together, have been breathtaking.

Contrary to the expectations of the HGP's proponents, few successful predictions about organismic development have come from sequence information alone (Stephens 1998). However, genomic research is persistently throwing up surprises:

1. The most important surprise from the HGP is that there are probably only about 30,000 genes in the human genome, compared with an estimate of 140,000 as late as 1994 (Hahn and Wray 2002). In general, plant genomes are expected to contain many more genes than the human genome. Morphological or behavioral complexity is not correlated with the number of genes that an organism has. This has been called the G-value paradox (ibid).
2. The number of genes is also not correlated with the size of the genome, as measured by the number of base pairs. The fruit fly *Drosophila melanogaster* has 120 million base pairs but only 14,000 genes; the worm *Caenorhabditis elegans* has 97 million base pairs but 19,000 genes; the mustard weed *Arabidopsis thaliana* has only 125 million base pairs and 26,000 genes; while humans have 2,900,000,000 base pairs and 30,000 genes (Hahn and Wray 2002).
3. At least in humans, the distribution of genes on chromosomes is highly uneven. Most of the genes occur in highly clustered sites. Most of these genes are expressed in many tissues—the so-called "housekeeping" genes (Lercher, Urrutia, and Hurst 2002). However, the spatial distribution of cluster sites appears to be random across the chromosomes. (Cluster sites tend to be rich in C and G, whereas gene-poor regions are rich in A and T.) In contrast, the genomes of arguably less complex organisms, including *D. melanogaster*, *C. elegans*, and *A. thaliana,* do not have such pronounced clustering.
4. Only 2% of the human genome codes for proteins, while 50 % of the genome is composed of repeated units. Coding regions are interspersed by large areas of noncoding DNA. However, some functional regions, such as *HOX* gene clusters, do not contain such intervening sequences.
5. Scores of genes appear to have been horizontally transferred from bacteria to humans and other vertebrates, though apparently not to other eukaryotes. However, this issue remains highly controversial.
6. Once attention shifts from the genome to the proteome, or the protein complement of a cell (see below for more detail), a strikingly different pattern emerges. The human proteome is far more complex than the proteomes of the other organisms for which the genomes have so far been sequenced. According to some estimates, about 59% of the human genes undergo alternative splicing, and there are at least 69,000 distinct protein sequences in the human proteome. In contrast, the proteome of *C. elegans* has at most 25,000 protein sequences (Hahn and Wray 2002).
7. It now appears that noncoding DNA is routinely transcribed into RNA but not translated in complex organisms (Mattick 2003). It seems that these RNA transcripts form regulatory networks that are critical to development. Interestingly, the amount of noncoding DNA sequences in organisms appears to grow monotonically with the morphological complexity of organisms.
8. At least in *A. thaliana*, there is evidence of genome-wide non-Mendelian inheritance during which specifications from the grand-parental, rather than parental, generation are transmitted to organisms (Lolle et al. 2005).

An important task of modern molecular biology is to make sense of these disparate unexpected discoveries. One conclusion seems unavoidable: Any concept of the gene reasonably close to that in classical genetics will be irrelevant to the molecular biology of the future (see Genetics).

Proteomics

The term "proteome" was introduced only in 1994 to describe the total protein content of a cell produced from its genome (Williams and Hochstrasser 1997). Unlike the genome, the proteome is not even approximately a fixed feature of a cell (let alone an organism), but changes over time during development. Deciphering the proteome, and following its temporal development during the life cycle of each tissue of an organism, has emerged as the major challenge for molecular biology in the postgenomic era. This project has been encouraged by the discovery of unexpected universality of developmental processes at the level of cells and proteins (Gerhart and Kirschner 1997). For instance, even though hundreds of genes are known to specify molecules involved in transport across cellular membranes, there are only about twenty transport mechanisms in all living systems. The emergence of proteomics in the wake of the various sequencing projects signals an acceptance of the position that studying processes largely at the DNA level will not suffice to explain phenomena at the cellular and higher levels of organization. Even genomics did not go far enough; a sharper break with the past will be necessary.

Nevertheless, in one very important sense, the emergence of proteomics recaptures the spirit of early molecular biology, when all molecular types, but especially proteins, were the foci of interest, and the deification of DNA had not replaced a pluralist vision of the molecular basis for life. In the late 1960s, Brenner and Crick proposed "Project K, the complete solution of *E. coli.*" *E. coli* (strain K-12) was selected as a model organism because of its simplicity (as a unicellular prokaryote) and ease of laboratory manipulation. Project K included: (i) a "detailed test-tube study of the structure and chemical action of biological molecules (especially proteins)"; (ii) completion of the models of protein synthesis; (iii) work on the structure and function of cell membranes; (iv) the study of control mechanisms at every level of organization; and (v) the study of the behavior of natural populations, including population genetics. Once *E. coli* was solved, and biology was supposed to move on to more complex organisms (Crick 1973, 67).

Notice that DNA receives no preferential attention at the expense of other molecular components in Project K and that the centrality of proteins as the most important active molecules in a cell is recognized. Project K accepts that there is much more to the cell than DNA; it accepts that no simple solution of the cell's behavior can be read from the genomic sequence. After a generation of infatuation with DNA and genetic reductionism, the aims of proteomics return in part to the vision of biology incorporated in Project K. However, at least in one important way, that project went even beyond proteomics as currently understood: It emphasized all levels of organization, whereas the explicit aims of proteomics are limited to the protein level. The future will probably require further expansion.

Meanwhile, work on proteins has also generated unexpected challenges. In particular, the four rules of classical molecular biology have not survived intact, and at least the last three will require some modification. It now appears—though the essential idea goes back to the 1960s—that the fit between interacting sites of protein molecules is more dynamic than in the classical model, with the active site often "inducing" an appropriate fit (see e.g., Koshland and Hamadani 2002). It also appears that a more complicated model than the original allostery model will be required to account for many cases of cooperativity. A systematic philosophical appraisal of these developments is yet to be undertaken.

Philosophical Speculations

The developments described in the last section are so recent that any attempt to interpret their philosophical significance must remain partly speculative. Some of the empirical generalizations noted will undoubtedly be challenged by further work in the near future—if the recent past of molecular biology is any guide to its future. Moreover, there has been very little philosophical attention to these developments.

Beyond Reduction?

That the four rules of classical molecular biology are being challenged, at least to some extent, is not reason enough to generate any new skepticism about the reductionist interpretation of explanation in molecular biology. They do not bring the physical explanation of wholes by parts into question. However, if an RNA-based (or other) regulatory network turns out to be crucial to explaining development (and evolution, as Mattick 2003 argues), the reductionist interpretation may be in trouble. If network-based explanations are ubiquitous, it is quite likely that what will often bear the explanatory weight in such explanations is the topology of the network. As noted earlier, some classical phenomena such as dominance have already been known to resist straightforward reductionist explanation (Sarkar 1998).

Topological explanations have not received the kind of attention from philosophers they deserve, even though networks have lately entered the center stage of scientific attention (Mattick and Gagen 2005). Here "topology" refers to the connectivity properties of systems such as networks, which, without loss of generality, can be modeled as directed graphs. The vertices of such a graph represent components of a system, and edges (between vertices), with appropriate directionality and weights, represent interactions between such vertices. How topological an explanation is becomes a matter of degree: The more an explanation depends on individual properties of a vertex, the closer an explanation comes to traditional reduction (the components matter more than the structure) (see Reductionism). Conversely, the more an explanation is independent of individual properties of a vertex, the less reductionist it becomes. In the latter case, if explanations invoke properties of a graph that measure its connectivity, then these are topological explanations. Such connectivity measures include the number of edges in the graph, the distribution of edge degree between vertices (the "degree" of a vertex being the number of edges incident on it), and so on. (For a review of network theory, see Newman 2003). If topological explanations become necessary in molecular biology, it will mark a serious philosophical break with the reductionist classical era.

Beyond DNA Information?

As noted earlier, there is as yet no fully satisfactory account of biological information that is appropriate for molecular biology. However, the developments within eukaryotic genetics and, especially, genomics strongly suggest that the view that DNA is the sole carrier of information, however it is characterized, cannot be sustained at least for organisms more complicated than prokaryotes and perhaps not even for them. Most of the critical interactions that determine the future behavior of a cell seem to occur at the level of RNA: splicing, RNA editing, and so on. Because of this feature of cellular interactions, Sarkar (2005, Ch. 14) has speculated that the DNA genome consists of a relatively static set of sequestered modular templates (resulting in the "SMT model" of the cell), far from the classical view of the genome coding a program for development. The failure of the sequence hypothesis for many proteins only increases skepticism about the classical picture.

The routine generation of untranslated RNA transcripts from the genome also suggests that should cellular processes be viewed informationally, RNA networks form a parallel information-processing system partly independent from the genomic DNA (Mattick 2003). At present, it is unclear whether such information must also be viewed semiotically, though it seems likely, since the simplest way in which RNA sequences can be viewed as carriers of information is by the specification of information by the RNA sequences.

Similarly, the discovery of ubiquitous non-Mendelian genetic specification in *A. thaliana* (Lolle et al. 2005) also suggests that there is yet another parallel system of heredity that can also potentially be viewed informationally and, once again, is not specified through DNA. It is also possible that all such phenomena are best interpreted not informationally but using the more traditional—generally structural—conceptual apparatus of physics and chemistry. However, the distinction between the two frameworks becomes blurred in the case of RNA because the relation between the sequence and three-dimensional conformation seems to be relatively straightforward, at least much more so than in the case of proteins.

Finally, in these discussions of biological information, two issues should be distinguished:

- whether an informational framework for molecular biology is of any use; and
- whether, within any such framework, DNA (or, more restrictively, genomic DNA) is the sole repository of that information.

The problems mentioned here provide a forceful argument against the second claim, leaving open the status of the first.

Toward a Dynamic Account of the Organism

One problem with informational interpretations of molecular biology is that they have always been static: Time does not enter explicitly into accounts of biology based on the transfer of information, though, implicitly, such transfer must take place during some time interval. Recall that the proteome is not a static feature of the organism, let alone the cell: Proteomics requires a commitment to the characterization of cellular and organismic change over time. Moreover, the recent discoveries of potentially ubiquitous RNA network-based regulation also underscore the importance of dynamic accounts explicitly taking time into consideration. Moreover, new microarray techniques and their extensions are increasingly making temporal stages of cellular changes empirically accessible. The challenge remains to develop a theoretical framework to interpret the empirical information.

Any such framework can begin with either a physicalist or an informational characterization of cellular processes or a mixture of both, though prospects for a physicalist account do not seem particularly promising because of the sheer complexity of the molecular networks involved (Sarkar 2005, Ch. 10). But a dynamic informational account also leads to uncharted territory. In retrospect, what seems surprising is how successful the static framework for classical molecular biology has been, given that organisms are obviously dynamic entities undergoing development over time.

Conclusions: An Invitation

Molecular biology has not received the extent of philosophical attention it deserves, and the little it has received has been limited to the classical period (see Darden and Tabery 2005 for a more detailed summary than what has been presented here).

There are at least two reasons why philosophers should invest more work on the subject:

- Without at least a partial methodological commitment to molecular concepts and techniques, any subdiscipline within biology will likely soon be relegated to irrelevance. Philosophy of biology that does not take molecular biology into account will remain incomplete.
- Modern molecular biology raises fundamentally new epistemological questions, especially about the relevance of physical and semiotic informational accounts that have dominated discussions of biology for the last century. The deployment of philosophical (particularly formal) techniques may contribute significantly to the advancement of the field.

The most important task in the philosophy of biology for the next few decades will be to conceptualize the functional role of DNA within the cell so as to explain the surprising organization and other properties of the genome that were discussed earlier. Physical and informational accounts will probably have to interact in order to create a consistent satisfactory picture. As the last section indicates, any such attempt must necessarily begin with a clearer account than what is currently available of what 'information' means in a biological context. This is probably where philosophers have the most to contribute to the future of molecular biology (see Biological Information). Perhaps techniques from formal epistemology or semantics will enable progress where traditional biological tools have largely failed.

Sahotra Sarkar

References

Barrell, B. G., G. M. Air, and C. Hutchison III (1976), "Overlapping Genes in Bacteriophage PhiX174," *Nature* 264: 34–41.

Berget, S., C. Moore, and P. Sharp (1977), "Spliced Segments at the 5′ Terminus of Adenovirus 2 Late mRNA," *Proceedings of the National Academy of Sciences (USA)* 74: 3171–3175.

Berk, A., and P. Sharp (1978), "Structure of the Adenovirus 2 Early mRNAs," *Cell* 14: 695–711.

Cattaneo, R. (1991), "Different Types of Messenger RNA Editing," *Annual Review of Genetics* 25: 71–88.

Cavalier-Smith, T. (1978), "Nuclear Volume Control by Nucleoskeletal DNA, Selection for Cell Volume and Cell Growth Rate, and the Solution of the DNA C-Value Paradox," *Journal of Cell Science* 34: 247–278.

Chow, L., R. Gelinas, T. Broker, and R. Roberts (1977), "An Amazing Sequence Arrangement at the 5′ Ends of Adenovirus 2 Messenger RNA," *Cell* 12: 1–18.

Cook-Deegan, R. (1994), *The Gene Wars*. New York: Norton.

Crick, F. H. C. (1958), "On Protein Synthesis," *Symposia of the Society for Experimental Biology* 12: 138–163.

——— (1973), "Project K: "The Complete Solution of *E. coli*," *Perspectives in Biology and Medicine* 17: 67–70.

Culp, S. (1995), "Objectivity in Experimental Inquiry: Breaking Data-Technique Circles," *Philosophy of Science* 62: 438–458.

Darden, L., and J. Tabery (2005), "Molecular Biology," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2005/entries/molecular-biology

Fox, T. D. (1987), "Natural Variation in the Genetic Code," *Annual Review of Genetics* 21: 67–91.

Gehring, W. (1998), *Master Control Genes in Development and Evolution: The Homeobox Story*. New Haven, CT: Yale University Press.

Gerhart, J., and M. Kirschner (1997), *Cells, Embryos, and Evolution*. Oxford: Blackwell Science.

Griffiths, P. (2001), "Genetic Information: A Metaphor in Search of a Theory," *Philosophy of Science* 67: 26–44.

Hahn, M. W., and G. A. Wray (2002), "The G-Value Paradox," *Evolution and Development* 4: 73–75.

Hull, D. (1972), "Reduction in Genetics—Biology or Philosophy?" *Philosophy of Science* 39: 491–499.

Jacob, F. (1973), *The Logic of Life: A History of Heredity*. New York: Pantheon.

Keller, E. F. (2000), *The Century of the Gene*. Cambridge, MA: Harvard University Press.

Kitcher, P. (1984), "1953 and All That: A Tale of Two Sciences," *Philosophical Review* 93: 335–373.

Koshland, D. E., Jr. and K. Hamadani (2002), "Proteomics and Models for Enzyme Cooperativity," *Journal of Biological Chemistry* 277: 46841–46844.

Lercher, M. J., A. O. Urrutia, and L. D. Hurst (2002), "Clustering of Housekeeping Genes Provides a Unified Model of Gene Order in the Human Genome," *Nature Genetics* 31: 180–183.

Lolle, S. J., J. L. Victor, J. M. Young, and R. H. Pruitt (2005), "Genome-Wide Non-Mendelian Inheritance of Extra-Genomic Information in *Arabidopsis*," *Nature* 434: 505–509.

Mattick, J. (2003), "Challenging the Dogma: The Hidden Layer of Non-Protein-Coding RNAs in Complex Organisms," *BioEssays* 25: 930–939.

Mattick, J., and M. J. Gagen (2005), "Accelerating Networks," *Science* 307: 856–857.

Mayr, E. (1961), "Cause and Effect in Biology," *Science* 134: 1501–1506.

Monod, J. (1971), *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*. New York: Knopf.

Nagel, E. (1961), *The Structure of Science*. New York: Harcourt, Brace and World.

Newman, M. E. J. (2003), "The Structure and Function of Complex Networks," *SIAM Review* 45: 167–256.

Olby, R. C. (1974), *The Path to the Double Helix*. Seattle: University of Washington Press.

Pauling, L., and R. B. Corey (1950), "Two Hydrogen-Bonded Spiral Configurations of the Polypeptide Chains," *Journal of the American Chemical Society* 71: 5349.

Rheinberger, H. J. (1997), *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford, CA: Stanford University Press.

Sarkar, Sahotra (1989), "Reductionism and Molecular Biology: A Reappraisal," Ph. D. Dissertation, Department of Philosophy, University of Chicago.

——— (1996), “Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology,” in Sarkar (ed.), *The Philosophy and History of Molecular Biology: New Perspectives*. Dordrecht, Netherlands: Kluwer, 187–231.

——— (1998), *Genetics and Reductionism*. New York: Cambridge University Press.

——— (2005), *Molecular Models of Life: Philosophical Papers on Molecular Biology*. Cambridge, MA: MIT Press.

Schaffner, K. F. (1967), “Approaches to Reduction,” *Philosophy of Science* 34: 137–147.

Shannon, C. E. (1948), “A Mathematical Theory of Information,” *Bell System Technical Journal* 27: 379–423, 623–656.

Stephens, C. (1998), “Bacterial Sporulation: A Question of Commitment?” *Current Biology* 8: R45–R48.

Tauber, A. I., and S. Sarkar (1992), “The Human Genome Project: Has Blind Reductionism Gone Too Far?” *Perspectives on Biology and Medicine* 35: 220–235.

Thiéffry, D., and S. Sarkar (1998), “Forty Years under the Central Dogma,” *Trends in Biochemical Sciences* 32: 312–316.

Waddington, C. H. (1962), *New Patterns in Genetics and Development*. New York: Columbia University Press.

Waters, C. K. (1990), “Why the Anti-Reductionist Consensus Won’t Survive the Case of Classical Mendelian Genetics,” in A. Fine, M. Forbes, and L. Wessels (eds.), *PSA 1990: Proceedings of the 1990 Biennial Meeting of the Philosophy of Science Association*. East Lansing, MI: Philosophy of Science Association, 125–139.

Watson, J. D., and F. H. Crick (1953), “Molecular Structure of Nucleic Acids—A Structure for Deoxyribose Nucleic Acid,” *Nature* 171: 737–738.

Williams, K. L., and D. F. Hochstrasser (1997), “Introduction to the Proteome,” in M. R. Wilkins, K. L. Williams, R. D. Appel, and D. F. Hochstrasser (eds.), *Proteome Research: New Frontiers in Functional Genomics*. Berlin: Springer, 1–12.

Wimsatt, W. C. (1976), “Reductive Explanations: A Functional Account,” *Boston Studies in the Philosophy of Science* 32: 671–710.

Yockey, H. P. (1992), *Information Theory and Molecular Biology*. Cambridge, UK: Cambridge University Press.

See also **Biological Information; Biology; Explanation; Function; Genetics; Heredity and Heritability; Holism; Mechanisms; Neurobiology, Philosophy of; Genetics; Physicalism; Reductionism**

N

ERNEST NAGEL

(16 November 1901–20 September 1985)

Nagel was born in Bohemia and came to the United States when he was ten years old. He became a naturalized citizen of the United States in 1919. In 1923, he received a B.A. from the College of the City of New York, in 1925, a master's degree in philosophy from Columbia University, and in 1931, a Ph.D. in philosophy from Columbia. Most of his academic career was spent at Columbia, beginning with his appointment in 1931 and ending with his retirement in 1970. During his last three years at Columbia, he held the position of university professor. He died in New York City.

Nagel received many honors. He was a Guggenheim Fellow in 1934–1935 and 1950–1951. In 1954, he was elected to the American Academy of Arts and Sciences, and in 1962 to the American Philosophical Society. He was elected to the United States National Academy of Sciences in 1977.

During his more than forty years of active intellectual life at Columbia—he continued to participate in seminars and other activities after his retirement—Nagel played a central role in the intellectual life of Columbia and, in a more general way, of New York City. For several generations of students and colleagues, his critical philosophic spirit and his detailed attention to scientific methods made him an exemplar of how philosophy could be related to the sciences, both natural and social. His lecture courses and seminars were attended not merely by students of philosophy, but by a wide-ranging mixture of students from the natural and social sciences, as well as professional disciplines. These activities extended to a series of famous seminars with colleagues in other disciplines. Perhaps the best known was his long-standing seminar with Paul Lazarsfeld on methodology in the social sciences.

Nagel's own intellectual mentors were primarily Morris R. Cohen and John Dewey. Dewey was jointly appointed in philosophy and education at Columbia and was active there during the first decade or so of Nagel's years at Columbia. With Cohen, Nagel wrote what was probably the most influential textbook in logic and scientific method in the United States published in the first half of the twentieth century.

Major Works

The textbook that Nagel coauthored with Cohen was *An Introduction to Logic and Scientific Method* (Cohen and Nagel 1934). Nagel's (1939a) "Principles of the Theory of Probability," which was a contribution to the *International Encyclopedia of Unified Science* (Neurath, Carnap, and Morris 1939), was published separately (see Unity of Science Movement). Collections of Nagel's articles were published under the titles *Sovereign Reason* and *Logic without Metaphysics* (Nagel 1954 and 1956). His most important work was *The Structure of Science* (Nagel 1961); then, much later, *Teleology Revisited* (Nagel 1979) was published. What is important about his career is not only his teaching at Columbia and his role in New York City's intellectual life, but also the very large number of articles he published on a great variety of philosophical topics and, perhaps equally important, the extensive critical reviews, published mainly in the *Journal of Philosophy,* of many major philosophical works in the philosophy of science.

Some extended major critical analyses are to be found in his articles on Russell's philosophy of science (Nagel 1944; Russell 1944), Dewey's theory of natural science (Nagel 1950), and Carnap's theory of induction (Nagel 1963). In these three articles, Nagel shows many philosophical sympathies. But the striking thing about his approach is the carefulness of his critical appraisal of significant issues.

Criticism of Carnap

Nagel's critical spirit is reflected in his analysis of Carnap's use, in one form or another, of Laplace's ([1812] 1952) classical principle of indifference (see Carnap, Rudolf; Inductive Logic; Probability).

> I wish next to raise an issue that concerns not only c^* but also the whole continuum of inductive methods Carnap regards as possible candidates for explicating the notion of evidential support. Among the conditions he lays down which any reasonable c must satisfy, there are two that bear considerable resemblance to the notorious Principle of Indifference, often regarded as the Achilles heel of the classical theory of probability. The first of these stipulates that all the individuals are to be treated on par, the second introduces a similar requirement for the primitive predicates. (Nagel 1963, 797)

Here is Carnap's response:

> Nagel expresses doubts about the validity of those principles of my theory which are related to the classical principle of indifference. . . . Nagel raises objections especially against A7 [axiom of indifference] and in this context uses an illustration which refers to samples of water, taken either from different sources or from the same reservoir which is known to be homogeneous, and the like. What Nagel says about these situations and the attitude a scientist would take with respect to such samples is certainly correct, but it is no argument against A7. If the scientist X knows anything about the individuals a_1, a_2, a_4, a_5 other than that they come from the same reservoir, and if he knows either that the water in that reservoir is homogeneous or that it is not, then the knowledge of X is much stronger than the evidence e to which A7 refers. The special case of A7 formulated by Nagel is applicable only if, first, X does not know anything about the individuals a_1, a_2, a_4, a_5 other than that they have the property M and if, second, he does not know with regard to any other individual whether or not it has the property M. Nagel's error here is a case of what I shall later call the fallacy of incomplete evidence. (Carnap 1966, 991)

What is perhaps most interesting about Carnap's response to Nagel is that he does not say how to proceed if his axiom A7 of invariance is violated. Nagel, on his part, is not really suggesting a detailed alternative solution but is proposing a course of prudence in not endorsing too easily the principle of indifference.

Major Articles

Also to be mentioned is Nagel's (1955) presidential address to the American Philosophical Association, published as "Naturalism Reconsidered." It is equally worth mentioning some of the important and later much cited articles of Nagel. A reflection of his wide-ranging historical interests, as well as philosophical ones, is his influential article on the relation between the development of modern logic and the development of axiomatic methods in the nineteenth century (Nagel 1939b). Equally important is his still much cited, informal, but detailed, argument on how physicists conceive of the reduction of thermodynamics to statistical mechanics (Nagel 1949) (see Reductionism). This is a subject of great technical complexity. Nagel provides a clear analysis, showing the main ideas of the reduction, without losing the reader in the inevitable and complicated technical details. Other important works dealt with psychoanalytic theory (Nagel 1959), a much-debated topic at the time, and historical determinism (Nagel 1960).

The Structure of Science

General Issues

Nagel's (1961) most important work was his magisterial book on the philosophy of science, *The Structure of Science.* It is a mark of the depth and importance of this work that more than forty years later it is still a primary reference for students in the philosophy of science. In the introductory chapter, three broad areas are identified as those of major importance for analysis. They are the

1. Logical patterns exhibited by explanations in the sciences,
2. Construction of scientific concepts, and
3. Testing and validation of scientific inferences and their conclusions.

The next four chapters are general ones. Chapter 2 concentrates on patterns of explanation; Chapter 3 on the deductive pattern of explanations, in terms of both individual events and of laws; Chapter 4 focuses on the character of scientific laws, especially the questions of their universality and necessity, a topic that has a long history in philosophy, reaching back to Aristotle. Chapter 5 is concerned with experimental laws and theories. Nagel identifies three major components of theories. The first component is the abstract or systematic calculus; the second is a set of rules that assign an empirical content to the concepts of the abstract system; and the third is an interpretation or model for the abstract calculus. What is important about Nagel's treatment of these matters is that he provides many more detailed scientific illustrations than will be found in many comparable works. In this chapter he also provides a detailed treatment of the rules of correspondence for moving from the theory and its concepts to experimental data. Chapter 6 deals with the cognitive status of theories. What is important is the contrast between three views—the descriptive view, the instrumental view, and the realist view of theories. There is here, and in many other parts of Nagel's work, an important tension between the instrumental view, in which he is influenced by Dewey, and the realist view, which he sees as close to much of the language and thought of scientists.

Foundations of Physics and Biology

In broad terms, Chapters 7–11 deal with the foundations of physics. Chapter 7 focuses on the science of mechanics and is important in providing a clear account of why mechanical explanations have played such a prominent role in scientific thinking. Chapter 8 is on space and geometry, with reference to space in Newtonian or classical physics especially. Chapter 9 is on geometry in physics, particularly on the transition from classical physics to the geometric approach of general relativity theory. Chapter 10 focuses on causality and indeterminism in physical theory. Nagel gives a detailed analysis of the language, concepts, and laws of quantum mechanics. In this chapter, he also gives a careful and nuanced account of the way in which quantum mechanics is indeterministic, and also of the way in which it is not. Here is a good passage about the way in which quantum mechanics is deterministic:

> [A]n examination of the fundamental equations of quantum mechanics shows that the theory employs a definition of state quite unlike that of classical mechanics, but that relative to its own form of state-description, quantum theory is deterministic in the same sense that classical mechanics is deterministic with respect to the mechanical description of state. However, the state-description employed in quantum theory is extraordinarily abstract; and, although its formal structure can be readily analyzed, it does not lend itself to an intuitively satisfactory nontechnical exposition. (Nagel 1961, 306)

Chapter 11 is on the reduction of theories, and Nagel returns here to his well-known formulation of the reduction of thermodynamics to statistical mechanics. Important sections are added on emergence and wholes and sums and organic unities, which take us well beyond considerations of thermodynamics. The last section contains one of the most extensive discussions of scientific psychology in the book, with critical attention to the claims of Gestalt psychologists. The careful analysis of holism in this chapter is rightly regarded as one of the classical examples of critical thought in modern philosophy of science (see Emergence). The section begins by distinguishing eight senses of 'whole' and 'part.' Toward the end of the section, Nagel has this to say about organic unities:

> [L]et us turn to . . . what appears to be the fundamental issue in the present context. That issue is whether the analysis of "organic unities" necessarily involves the adoption of irreducible laws for such systems, and whether their mode of organization precludes the possibility of analyzing them from the so-called "additive point of view." The main difficulty in this connection is that of ascertaining in what way an "additive" analysis differs from one which is not. The contrast seems to hinge on the claim that the parts of a functional whole do not act independently of one another, so that any laws which may hold for such parts when they are not members of a functional whole cannot be assumed to hold for them when they actually are members. An "additive" analysis therefore appears to be one which accounts for the properties of a system in terms of assumptions about its constituents, where these assumptions are not formulated with specific reference to the characteristics of the constituents as elements in the

> system. A "nonadditive" analysis, on the other hand, seems to be one which formulates the characteristics of a system in terms of relations between certain of its parts as functioning elements in the system.
>
> However, if this is indeed the distinction between these allegedly different modes of analysis, the difference is not one of fundamental principle. We have already noted that it does not seem possible to distinguish sharply between systems that are said to be "organic unities" and those which are not. Accordingly, since even the parts of summative wholes stand in relations of causal interdependence, an additive analysis of such wholes must include special assumptions about the actual organization of parts in those wholes when it attempts to apply some fundamental theory to them. There are certainly many physical systems, such as the solar system, a carbon atom, or a calcium fluoride crystal, which despite their complex form of organization lend themselves to an "additive" analysis; but it is equally certain that current explanations of such systems in terms of theories about their constituent parts cannot avoid supplementing these theories with statements about the special circumstances under which the constituents occur as elements in the systems. (Nagel 1961, 394–395)

Chapter 12 is on mechanistic explanation in biology. This chapter anticipates the contents of Nagel's John Dewey lectures, which were delivered at Columbia in 1977 and published in the *Journal of Philosophy* in the same year, and also in Nagel's 1979 book, *Teleology Revisited.* This chapter and the later lectures provide a careful account of the importance of a scientific notion of teleology in biology, with a particular emphasis on the structure of teleological explanations. In the last part of the chapter Nagel clears a critical path between the rhetorical excesses of some organismic biologists and the unsupported dogmatism of some mechanistic biologists.

Social Sciences and History

Chapter 13 is on methodological problems in the social sciences, with an emphasis on work in sociology, but with comments as well on issues in psychology, economics, and anthropology. A notable feature of this chapter is Nagel's critique of the well-known view of John Stuart Mill that experimentation in the social sciences is not possible. Nagel shows that in fact Mill was not at all successful in trying to draw a sharp line between the possibility of experimentation in the natural sciences and in the social sciences. Chapter 14 is on explanation and understanding in the social sciences. Nagel concentrates on three important issues. The first is why statistical generalizations are to be expected as appropriate explanations of many social phenomena. The second is what the scientific status of functionalism is in the social sciences. Here "functionalism" refers to the doctrine that every social aspect of a culture of society has some purposive role to play, much in the spirit of teleological approaches in biology. The third issue concerns whether or not methodological individualism is the correct way to think about the methods and aims of the social sciences. One too simple view is that sociology should be reducible to psychology, group behavior to individual behavior. Nagel's detailed analysis of the subtle aspects of this controversy is among the best in the extensive literature. The final chapter, Chapter 15, is on problems in the logic of historical inquiry. Much of the focus is on philosophical problems of the nature of history that have been current for a very long time but remain controversial. To provide insight into how Nagel approaches these matters, two extensive quotations are cited. The first is on the selective character of historical data and the accompanying analysis:

> It is a platitude that research in history as in other areas of science selects and abstracts from the concrete subject matter of inquiry, and that however detailed a historical discourse may be it is never an exhaustive account of what actually happened. Curiously enough, although natural scientists have rarely been agitated by parallels in their own branches of study to these obvious features of historical inquiry, the selective character of historical research continues to be a major reason historians give for the sharp contrast they frequently draw between other disciplines and the study of the human past, as well as the chief support for the skepticism many of them profess concerning the possibility of achieving "objective" historical explanations. . . . Were this doctrine sound, every historical account that could be constructed by a finite intelligence would have to be considered a necessarily mutilated version of what actually happened; indeed, all science and all analytical discourse would have to be condemned in an identical manner. But the claim that all historical explanations are inherently arbitrary and subjective is intelligible only on the assumption that *knowledge* of a subject matter must be *identical with* that subject matter or must *reproduce it* in some fashion; and this assumption, as well as the claim accompanying it, must be rejected as absurd. Thus, a map cannot be sensibly characterized as a distorted version of the region it represents, merely because the map does not coincide with the region or does not mention every item that may actually exist in that region; on the contrary, a "map" which was drawn to scale and which omitted nothing would be a monstrosity utterly without purpose. (Nagel 1961, 576–577)

Nagel's vivid map analogy is a characteristic feature of both his lectures and his writing—finding something concrete and familiar, but serious, to illuminate the argument.

The second problem concerns historians' use of counterfactuals:

> [N]o mention has thus far been made of a familiar special form in which historians frequently assign an order of relative importance to events, namely, when they assert contrary-to-fact conditionals about the past. . . . To cite a famous example, many historians believe that the battle of Marathon in 490 B.C. was one of the decisive military conflicts in human history; and they support this belief by the contrary-to-fact judgment that, had the Persians been victorious, an Oriental theocratic-religious culture would have been established in Athens, with the consequence that Greek science and philosophy, in which Western civilization has its roots, would not have been developed. . . . Contrary-to-fact judgments are unavoidable except by eschewing all judgments of relevance and all attempts at explaining what has happened. We had occasion to note much earlier [in chapter 4] the intimate connection between scientific laws and counterfactual statements; and, since historical explanations require at least the tacit use of general assumptions, such explanations thereby assert at least by implication contrary-to-fact conditionals. . . . Nevertheless, it is in general by no means an easy task to provide reasonably firm grounds for contrary-to-fact judgments in human history. The task is undoubtedly more difficult than the analogous task in many other disciplines, partly because (as has so often been noted) it is impossible to perform experiments on nonrecurrent events, but in large measure because of the paucity of relevant data on most of the questions about which historians make such judgments. Despite these disadvantages, the task is not quite so hopeless as is frequently claimed. (Nagel 1961, 588–589)

Nagel's way of doing philosophy is nicely illustrated by this quotation. He is skeptical of bold philosophical claims of absolute distinctions—for example, between the methods of physicists and those of historians. But he is happy to focus on distinctions or similarities that have serious conceptual or empirical support. It is sweeping, overly general pronouncements about science, its methods or its structure, that spur his critical spirit to dig into the details whatever the subject matter, be it motion of atoms or battles of the past.

The sweep of this work, with detailed analysis ranging from quantum mechanics to history, is unique among major works in the philosophy of science published in the second half of the twentieth century. In considering mathematically developed parts of science, such as quantum mechanics, he was usually, but not always, successful in conveying a definite sense of the major conceptual issues without using explicitly the mathematical concepts on which specific results depended. The goal, with philosophical readers in mind, was to talk about details but to minimize mathematical formulations and computations. How Nagel felt about the desirability of entering into the intricacies of any scientific discipline on which one wished to make philosophical remarks is well exemplified by the following quotation on the theory of natural science by Dewey, a philosopher whom Nagel admired but of whom he was appropriately critical:

> But there are also less external reasons for the hesitations which even those in full sympathy with Dewey's aims and over-all conclusions have experienced with his account of natural science. The great William Harvey is reported to have said of Francis Bacon that he wrote about science like a Lord Chancellor. Of Dewey it can be said with equal justice that he writes about natural science like a philosopher, whose understanding of it, however informed, is derived from second-hand sources. With rare exceptions, the illustrations he supplies for his major theses on the nature of physical science and its methods come from everyday inquiries of a fairly elementary kind, or from popularized versions of the achievements of theoretical physics. It is indeed curious that a thinker who has devoted so much effort to clarifying the import of science as has Dewey, should exhibit such a singular unconcern for the detailed articulation of physical theory. (Nagel 1950, 247)

Writing this summary of *The Structure of Science*, almost a half a century after it was first published, it seems appropriate to end by some comments on aspects of science and the philosophy of science that were not so evident in that earlier period but are now salient.

The first is that the treatment of causality by Nagel is too centered on determinism. At the time he was writing, one could scarcely find mention of the word *cause* in a standard statistical analysis of data, even if that were implicit in the design of the experiment from which the data arose. The situation is very different now. There is a large and complicated literature on probabilistic causality, but much of what Nagel has to say about causality is not affected by this move from deterministic to probabilistic conceptions.

The second and related point is that the discussions of statistical laws and statistical generalizations, especially in the social sciences, seem too purely empirical after half a century of building probabilistic or stochastic models of all kinds of psychological, economic, and social behavior. The origin of such models can be traced back to before the second half of the twentieth century, but the renaissance certainly did not occur until then.

Their impact has been profound and has made probabilistic modeling and ways of thinking an integral part of the social sciences and also, at the beginning of the twenty-first century, of biology. But there is little in this new emphasis that goes against any fundamental tenets of Nagel's view of the structure of science or the place of probability in it. If it had happened earlier, it is a move he would have applauded.

Third, the various chapters on the social sciences, with their emphasis on sociology, anthropology, and history, seem, in many ways, out of kilter with the main theoretical developments in the social and behavioral sciences over the past several decades. These developments have centered on the increasing use of mathematically formulated models and theories, in economics especially, and also, to a lesser degree, in psychology. If new chapters were to be added, one on the structure of modern theories of economics and another on psychology, it would be very appropriate. In terms of the most recent events, the one on psychology would also move, in a detailed way, toward the intimate involvement with the neurosciences, which will, in the rest of this century, surely have a profound impact on our scientific conception of human nature, and on the way that psychologists formulate their theoretical ideas and philosophers of science modify their conceptions about the nature of language and mental representation. But, again, Nagel would be the last to be surprised at such developments, and they would not disturb, in a deep way, his insistence that what he was after in *The Structure of Science* was to give the general framework of the scientific method, not the current details of specific disciplines.

Patrick Suppes

References

Carnap, Rudolf (1966), "Replies and Expositions," in Paul Arthur Schilpp (ed.), *The Philosophy of Rudolf Carnap* (Library of Living Philosophers, Vol. 11). LaSalle, IL: Open Court, 859–1016.

Cohen, Morris Raphael, and Ernest Nagel (1934), *An Introduction to Logic and Scientific Method*. New York: Harcourt, Brace, Jovanovich.

Laplace, Pierre Simon, Marquis de ([1812] 1952), *A Philosophical Essay on Probabilities*. Translated from the 6th French edition by Frederick Wilson Truscotte and Frederick Lincoln Emory. New York: Dover.

Nagel, Ernest. (1939a), "Principles of the Theory of Probability," in O. Neurath, R. Carnap, and C. Morris (eds.), *International Encyclopedia of Unified Science* (Vol. 1, No. 6). Chicago: University of Chicago Press.

——— (1939b), "The Formation of Modern Conceptions of Formal Logic in the Development of Geometry," *Osiris* 7: 142–224.

——— (1944), "Russell's Philosophy of Science," in Paul Arthur Schilpp (ed.), *The Philosophy of Bertrand Russell* (Library of Living Philosophers, Vol. 5). Chicago: Northwestern University Press, 317–350.

——— (1949), "The Meaning of Reduction in the Natural Sciences," in Robert C. Stouffer (ed.), *Science and Civilization*. Madison: University of Wisconsin Press, 99–135.

——— (1950), "Dewey's Theory of Natural Science," in Sidney Hook (ed.), *John Dewey, Philosopher of Science and Freedom*. New York: Dial Press.

——— (1954), *Sovereign Reason*. Glencoe, IL: Free Press.

——— (1955), "Naturalism Reconsidered," *Proceedings and Addresses of the American Philosophical Association* 28: 5–17.

———. (1956), *Logic without Metaphysics*. Glencoe, IL: Free Press.

——— (1959), "Methodological Issues in Psychoanalytic Theory," in Sidney Hook (ed.), *Psychoanalysis Scientific Method and Philosophy*. New York: NYU Press, 38–56.

——— (1960), "Determinism in History," *Philosophy and Phenomenological Research* 20: 291–317.

——— (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World.

——— (1963), "Carnap's Theory of Induction," in Paul Arthur Schilpp (ed.), *The Philosophy of Rudolf Carnap* (Library of Living Philosophers, Vol. 11). LaSalle, IL: Open Court, 785–826.

——— (1979), *Teleology Revisited*. New York: Columbia University Press.

Neurath, Otto, Rudolf Carnap, and Charles W. Morris (1939), *International Encyclopedia of Unified Science* (Vol. 1, No. 6). Chicago: University of Chicago Press.

Russell, Bertrand (1944), "Reply to Criticisms," in Paul Arthur Schilpp (ed.), *The Philosophy of Bertrand Russell* (Library of Living Philosophers, Vol. 5). Chicago: Northwestern University Press, 679–742.

NATURAL KINDS

See **Induction, Problem of; Species**

NATURAL NECESSITY

See **Laws of Nature**

NATURAL SELECTION

In modern evolutionary biology, a set of objects is said to experience a selection process precisely when those objects vary in *fitness* (see Fitness). For example, if zebras that run fast are fitter than zebras that run slow (perhaps because faster zebras are better able to avoid lion predation), a selection process is set in motion. If the trait that exhibits variation in fitness is *heritable*—meaning, in our example, that faster parents tend to have faster offspring and slower parents tend to have slower offspring—then the selection process is apt to change trait frequencies in the population, leading fitter traits to increase in frequency and less fit traits to decline (Lewontin 1970). This change is the one that selection is "apt" to engender, rather than the one that must occur, because evolutionary theory describes processes other than natural selection (e.g., mutation, recombination, migration, drift, inbreeding) that can change trait frequencies and can nullify the effects that selection is disposed to bring about (see Evolution). This is why heritable variation in fitness is neither necessary nor sufficient for evolution (Brandon 1990).

The logical schema just described is very abstract; the zebra example involves conspecific organisms, but it also is possible that the "objects" in a selection process might be genes, or groups of conspecific organisms, or communities of organisms from different species. The schema also leaves open how often selection actually brings about the effects that it would cause if there were no counteracting forces. Thus, the question of how the schema applies to the living world gives rise to many empirical questions, some of which have interesting philosophical dimensions.

Adaptationism

At the close of his introduction to *On the Origin of Species*, Darwin [1859] 1964, 6 says that natural selection is "the main but not the exclusive" cause of evolution. In reaction to misinterpretations of his theory, Darwin felt compelled to reemphasize, in the book's last edition, that there was more to evolution than natural selection. It remains a matter of controversy in evolutionary biology how important natural selection has been in the history of life. This is the point of biological substance that presently divides adaptationists and anti-adaptationists. The debate over adaptationism also has a separate methodological dimension, with critics insisting that adaptive hypotheses be tested more rigorously (Gould and Lewontin 1979; Sober 1993).

Although it is widely agreed that natural selection has been an important cause of the similarities and differences that characterize the living world, the question remains of how important nonselective processes have been. For example, the underlying genetic system can "get in the way" of natural selection, preventing the fittest of the phenotypes found in a population from evolving to fixation. The simplest example of this is heterozygote superiority: If there are three genotypes at a locus and the heterozygote is the fittest, the genetic system will prevent that genotype and its associated phenotype from evolving to 100% representation. Adaptationists tend to minimize the practical import of this theoretical possibility (saying, for example, that heterozygote superiority is rare), whereas anti-adaptationists often take it very seriously indeed.

Gould and Lewontin (1979) criticized adaptationists for inventing "just-so stories," in which claims about adaptive significance are accepted only because they seem intuitively plausible. They also complained that adaptationism was unfalsifiable, since a new adaptive explanation could be invented if an old one were empirically disconfirmed; unfortunately, this point also applies to the pluralism about evolutionary processes that anti-adaptationists have favored. Gould and Lewontin also criticized adaptationists for taking a naively atomistic approach to how traits are individuated. Although few biologists would regard five fingers on the left hand as a different trait from five fingers on the right (because they do not evolve independently), adaptationists have argued that female orgasm in humans evolved independently of male orgasm, subject to its own selection pressures. It is a characteristic anti-adaptationist suggestion that female orgasm is to male orgasm as male nipples are to female nipples; the traits in each pair are products of the same developmental processes. Both evolved because there was selection in one sex for the trait, and the trait emerged in the other as a correlated consequence (Lloyd 2003).

Adaptationists have replied to these criticisms in several ways. One reply has been to insist that the idea of natural selection is an indispensable tool for biological investigation (Dennett 1995). A second has been to assert that selection is the only natural process that can account for adaptive complexity (Dawkins 1982). It is noteworthy that these adaptationist replies do not address the methodological objections that the critics advanced.

One positive outcome of the controversy has been the development of more rigorous methods for testing adaptive hypotheses—for example, controlling for the influence of nonselective processes (Harvey and Pagel 1991; Orzack and Sober 2001). It is to be hoped that biologists will recognize that global affirmations or denials of the importance of natural selection are not required *before* one studies the evolution of a particular trait in a particular group of organisms. Adaptationism and anti-adaptationism, as general biological claims, are summary *conclusions* that might be drawn after the evolution of a range of traits is understood; they are not needed as *premises.*

The Units of Selection Problem

Although Darwin usually thought of natural selection in terms of different organisms in the same species competing with each other, he also thought there were traits in nature that should be explained by postulating a process of *group selection*, wherein different groups in the same species compete with each other. This idea came in for severe criticism during the 1960s, and it remains controversial to this day (see Altruism). The logical schema for the process of evolution by natural selection also has been applied to the genes that exist in a single organism; this is the process of *intragenomic conflict.* The genes in a single organism often sink or swim together—they are equal in fitness, in that each has the same chance of finding its way to the next generation. However, genes in the same organism sometimes compete. For example, in the process of *meiotic drive*, heterozygotes produce gametes that bear one allele disproportionately more than the other. *Multilevel selection theory* (Sober and Wilson 1998) is the idea that there are different "units of selection"—that natural selection occurs among genes in the same organism, among organisms in the same group, among groups in the same species, and perhaps even among species in the same monophyletic taxon (Gould 2002). This idea contrasts with the doctrine of the *selfish gene*, which says that natural selection should be thought of as a process that exists exclusively at the genetic level (Dawkins 1976; Sterelny and Kitcher 1988).

When Darwin discussed the evolution of altruistic characteristics, he saw that there could be a conflict of interest between what was good for the individual and what was good for the group. More recent work has shown that this type of conflict can arise at levels of organization that Darwin was unable to consider. For example, driving genes in the house mouse are favored at the intragenomic level but are selected against at the level of whole organisms, since they render males sterile when found in double dose. There also is selection against the driving gene at the group level, since groups whose males are all homozygotes go extinct, and the copies of the gene found in females are thereby taken out of circulation (Lewontin 1970). It is intuitive to think of selection processes at different levels as component vectors that serve to increase or reduce the frequency of a trait; the net effect of selection at all levels is the result of combining these vectors into a single resultant.

Gradualism

Darwin thought of natural selection as acting on variations that have small effects—a complex adaptation, like the vertebrate eye, does not appear all at once. This point pertains to the question of how a trait first originates in a single individual—whether its parents had 99% of an eye, or no eye at

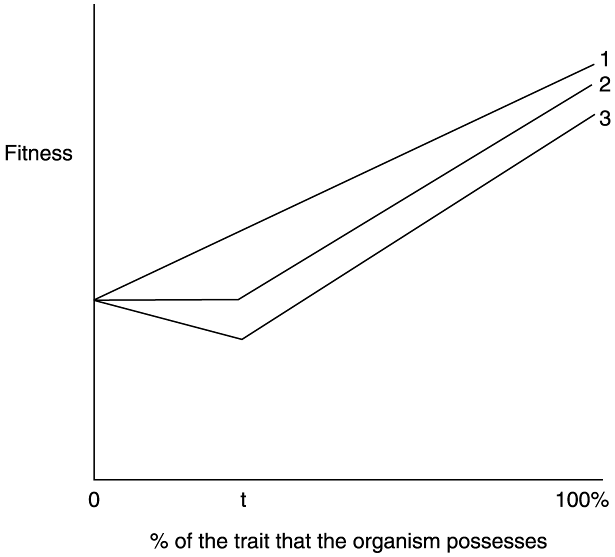


Fig. 1 What are the fitness consequences of having *n*% of a wing or an eye, as opposed to having (*n* – 1)%? According to line 1, each small increase represents an increase in fitness. According to line 2, having more of the trait makes no difference in fitness until a threshold (*t*) is crossed. Line 3 also depicts a threshold effect, but here having more of the wing or eye is deleterious, not neutral, until the threshold is crossed. Evolution via the pure process of Darwinian gradualism requires the monotonic increase that line 1 exhibits and cannot occur if the fitnesses are those represented by lines 2 or 3. However, evolutionary theory countenances processes additional to that of "pure Darwinian gradualism," so, in fact, the theory says that it *is* possible for a trait to evolve under all three scenarios.

all. In either case, once the complex trait is found in a single individual, the question must be faced of how the trait is to become common in the population. Both saltationists (those who believe that evolution involves "jumps") and gradualists may want to invoke the process of natural selection to account for this. Saltation, therefore, is not an alternative to natural selection. Although Darwin did not know about Mendelian genes, much subsequent work in genetics has vindicated his gradualism. However, recent work indicates that single mutations sometimes produce large effects—for example, the *HOX* gene can cause a leg to appear on the head of a fruit fly. Contemporary biologists usually do not see this as diminishing the importance of natural selection.

If Darwinian gradualism, strictly construed, is to transform an ancestral population in which the organisms have no eyes at all into a descendant population in which all organisms have 100% of an eye, it is essential that *n*% of an eye represent an advantage compared with (*n* – 1)%. In terms of the fitness functions depicted in Figure 1, Darwinian gradualism can occur when the fitnesses obey the pattern in line 1, but not when they conform to lines 2 or 3. However, recall that contemporary evolutionary theory describes processes other than strict Darwinian gradualism (for example, drift, to be described later, allows a less fit trait to replace one that is fitter), and so the evolution of traits whose fitness profiles conform to lines 2 and 3 is by no means impossible.

In thinking about whether a trait can evolve in conformity with the rules of Darwinian gradualism, it is important to consider the possibility that a trait might start evolving for one reason and then continue evolving for another. Perhaps rudimentary and complex eyes both evolved for the same adaptive reason, because they help the organism process information about the environment that is contained in light (Dawkins 1996). However, this pattern seems less plausible for the case of wings. Even if later wing evolution was driven by the usefulness of flying, it is hard to see how the evolution of rudimentary wings could have proceeded in this way, since 5% of a wing provides no lift at all; being able to fly is a threshold effect. Kingsolver and Koehl (1985) argue that insect wings began evolving as devices for regulating temperature and continued to evolve as devices for flying.

Progress

Is the process of natural selection an instrument of progress? This question must be divided in two, distinguishing the issue of moral progress from that of improvement in fitness. With respect to the former, Darwin clearly recognized that the process of natural selection involves a mountain of suffering and death; indeed, he sometimes expressed revulsion at the adaptations that natural selection produces. One grisly example that repeatedly drew his attention was the ability of parasitic wasps to paralyze their hosts and lay eggs in them; when the eggs hatch, the young feed on the living caterpillar, leaving its brain for last (Gillespie 1979). At the same time, Darwin often expressed approval of the changes that selection brings about in nature, and he often did so as well in connection with the workings of selection in human evolution. Whatever moral ambivalence Darwin may have felt about natural selection, it is interesting that Darwin's "bulldog," Thomas Henry Huxley (1893) thought that there was a profound conflict between what is good for us in terms of fitness and what is good for us in terms of morality, with morality obliging us to control the instincts that natural selection has put in place. These critical assessments of the moral significance of natural selection contrast starkly with the wholesale endorsement of the process that the political movement called Social Darwinism supplied. Social Darwinism used Darwin's theory to construct a moral justification for ruthless capitalism, according to which the weak had to be crushed to make way for the strong. Many modern biologists, sobered by the political misuses of Darwinism, now draw back from this ideological endorsement. They want their science to be *value free;* that is, a scientific theory has the job of saying what changes occur in nature and why, but it is not a scientific problem to say whether those changes represent good news or bad. (See Ruse 1997 for more on how the concept of progress figures in the thinking of different evolutionists.)

With respect to the second sort of progress—the effect of selection on fitness—Darwin thought of the process as an improver. He says that "as natural selection works solely by and for the good of each being, all corporeal and mental endowments will tend to progress towards perfection" (Darwin 1859, 489). Modern evolutionary theory makes it clear, however, that the process of natural selection need not improve the fitness of organisms. Quite apart from whether improving fitness is a good thing, selection can reduce the average fitness of the organisms in a population, even when the physical environment is static. If altruists compete against selfish individuals in a single persisting population, selfishness will go to fixation, with the result that the individuals in the population at the end of the process are less fit than their ancestors were when the process began. Adam Smith's optimistic picture of an invisible hand increasing the wealth of nations has been supplemented (though not replaced) by the pessimistic picture of the tragedy of the commons. Selection can improve average fitness, but it also can reduce it (Sober 1993, 97–99).

The Propensity Interpretation of Fitness

Alfred Russel Wallace, the codiscoverer of the theory of evolution by natural selection, suggested to Darwin that he drop the expression "natural selection" because it misleadingly suggested conscious choice; Wallace preferred Herbert Spencer's phrase "the survival of the fittest" to characterize the theory (Hodge 1992). Darwin embraced this summary slogan; however, it gave rise to the criticism that the theory is a tautology. If the fit are defined as those who survive, one cannot explain why one set of organisms survived to reproductive age while another did not by saying that the former were fitter. The first step in replying to this criticism is provided by the propensity interpretation of fitness (Mills and Beatty 1977; Brandon 1990; Sober 1984). Fitness is to reproductive success as solubility is to dissolving—fitness is a dispositional property. In particular, it is a probabilistic disposition, a propensity. A fair coin is disposed to land heads more often than one that is biased in favor of tails. If one organism is fitter than another, then the first will *probably* be more reproductively successful. The natural way to represent this mathematically is in terms of the idea of a probabilistic expectation. If an organism has a probability p_i of having exactly i offspring ($i = 0, 1, 2, 3, \ldots$), then its expected number of offspring is $3i(p_i)$. The expected number of offspring is not the exact number one should expect the organism to have; rather, it is the average number the organism would have if it got to live its life again and again under identical circumstances. Thus, the first line of reply to the charge that the statement "*a* is fitter than *b* if and only if *a* is more reproductively successful than *b*" is a tautology is to point out that the statement is not even true. If the rejoinder comes that it is then a matter of definition that "*a* is fitter than *b* if and only if *a* has a higher expected number of offspring than *b*," the reply here is that every theory contains definitions. Even if this is the proper definition of fitness, it does not follow that the entire theory is tautologous; evolutionary biology is full of empirical claims. (For discussion of whether fitness should be defined as a probabilistic expectation, see Sober 2001.)

Chance in Evolution

One reason that fitness should be understood as a probabilistic quantity and not as an organism's actual degree of reproductive success is that evolutionary theory describes a nonselective process that can lead organisms to enjoy different degrees of reproductive success. This is the process of random genetic drift, which Motoo Kimura (1983) developed to explain the huge amounts of molecular variation observed in natural populations. Drift occurs when traits change frequency by random walk; this occurs when they are identical, or nearly identical, in fitness. Here we find a disanalogy with a deterministic propensity like solubility: If *a* and *b* are both immersed and *a* dissolves while *b* does not, then *a* must have been soluble and *b* must have been insoluble. But if *a* is more reproductively successful than *b*, it is not inevitable that *a* was fitter than *b*.

When should a difference in reproductive success be attributed to natural selection? If two identical twins are on a mountaintop and one is killed by a lightning strike while the other is not, should we conclude that the second twin was fitter (Beatty 1991)? One way to answer this question in the negative is to argue that there is no phenotypic difference between the two twins that could form the basis for saying that they differed in fitness. This brings out another feature of the propensity interpretation—just as a tossed coin has a given propensity to land heads by virtue of its physical makeup, so an organism has whatever degree of fitness it has in a given environment by virtue of its genetic and phenotypic characteristics. However, the fact remains that the first twin was standing in one place while the second was standing in another when the lightning struck. Is this not a phenotypic difference? It is not relevant that the property of standing in a given place on a given day is not heritable; selection does not require heritability, though evolution by natural selection does. A different approach to the twins problem is to think of it statistically. If two coins are each tossed once, and the first lands heads and the second tails, standard statistical practice does not allow one to reject the null hypothesis, which says that they are identical in their probabilities of landing heads. A similar conclusion can be drawn about the twins. Notice that it does not matter that the two individuals happen to be genetically identical.

To understand the relationship between selection and drift, it is important to distinguish process from product. Evolution in finite populations always includes the process of drift, whatever change in trait frequencies may result. In similar fashion, the possibility of sampling error exists when a fair coin is tossed ten times, regardless of whether the outcome is nine heads and one tail or five of each. And no matter which outcome occurs, it is a mistake to ask "how much" of the outcome was due to the coin's fairness and how much was due to finiteness of sample size. It is useful to have both selection and drift in evolutionary theory because both categories are needed to describe relevant similarities and differences. Two populations may be characterized by the same suite of trait fitness values even though they differ in size, and two populations may have the same size even though they are characterized by different suites of fitness values.

Chance is said to enter evolutionary theory in two ways. First, mutations are said to occur "by chance." Second, random genetic drift is described as a "chance process." The term "chance" has different meanings in these two remarks. The point about mutations is just that they do not arise because they would be useful. This has nothing to do with whether mutations are deterministically caused or arise by an irreducibly probabilistic process. The relation of random genetic drift to the possibility of an underlying determinism raises issues that are more subtle. When we talk about different organisms having different probabilities of surviving and different expected numbers of offspring, how are these probabilities to be interpreted? Reasoning about Newtonian theory, Laplace ([1814] 1951) famously opined that if determinism is true, then probabilities (other than 0 and 1) are merely subjective—they reflect an agent's lack of information, not the objective chanciness of events. If this is right, then interpreting the probabilities used in evolutionary theory depends on facts about microphysics (Rosenberg 1994). However, it is worth contemplating a possibility not dreamt of in Laplace's philosophy. Perhaps nonextreme probabilities at the macro level can be objective even if determinism is true at the micro level. The actual relative frequency interpretation of probability allows for this possibility, although it is inadequate in other respects as an interpretation of the probability concepts used in science. Perhaps other, more adequate, interpretations of probability can allow macroprobabilities to be both objective and independent of whether microdeterminism is true.

What Does Natural Selection Explain?

If selection (in the form of lion predation) over many generations has favored fast zebras over slow ones, selection can explain why all present-day zebras run fast. But does it explain, in addition, why this or that individual runs fast? Sober (1984) answered this question in the negative: Selection explains only the frequencies of traits in a population, not why individuals have the traits they do. Selection is like an entrance exam—if you are required to speak English to gain admission to a room, the test explains why the room is composed entirely of English speakers. However, the test does not explain why the individuals in the room (Sam, Aaron, etc.) speak English. The phenotypes that individuals develop are to be explained by their genes and environment, not by the process of natural selection. Neander (1995) criticized this position on a number of grounds. One criticism involves an appeal to transitivity: If selection can explain why all the individuals in a given generation have a trait, and if the individuals in the next generation have the traits they do because they inherited them from the previous generation, then by transitivity, selection helps explain why the offspring have the traits they do.

Are There Laws in Evolutionary Biology?

Beatty (1991) argued that biological regularities hold only because this or that contingent event occurred in the evolutionary process. For example, if a population obeys Mendel's "law" of assortment, which says that *Aa* heterozygotes produce equal numbers of *A*- and *a*-bearing gametes, this is because there has been sufficiently strong selection against meiotic drive in ancestral populations for this fair Mendelian mechanism to evolve. However, it is a historical contingency that this type of selection pressure actually occurred. Sober (1997) replied that even if a regularity of the form "All *H*s are *F*" is a contingent consequence of the earlier evolutionary event *E*, it still can be the case that "if *E* is true earlier, then it will be true later on that all *H*s are *F*" is an evolutionary law. Rosenberg (1994) develops a different set of reasons for thinking that there are no biological laws other than the principle of natural selection. A separate puzzle about the status of laws in evolutionary biology concerns the fact that they appear to be a priori mathematical truths when spelled out carefully (Sober 1984). This distinguishes them from physical laws like the law of universal gravitation, which is empirical. A possible explanation for why the dynamical laws of evolution should be a priori may be found in the fact that fitness and other biological properties are multiply realizable (see Sober 1999 for discussion).

Elliott R. Sober

References

Beatty, J. (1991), "Random Drift," in E. Fox Keller and E. Lloyd (eds.), *Keywords in Evolutionary Biology*. Cambridge, MA: Harvard University Press, 273–281.

Beatty, J. (1995), "The Evolutionary Contingency Thesis," in G. Wolters and J. Lennox (eds.), *Concepts, Theories, and Rationality in the Biological Sciences: The Second Pittsburgh-Konstanz Colloquium in the Philosophy of Science*. Pittsburgh: University of Pittsburgh Press, 45–81.

Brandon, R. (1990), *Adaptation and Environment*. Princeton, NJ: Princeton University Press.

Darwin, C. ([1859] 1964), *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Cambridge, MA: Harvard University Press.

Dawkins, R. (1976), *The Selfish Gene*. Oxford: Oxford University Press.

——— (1982), "Universal Darwinism," in D. Bendall (ed.), *Evolution from Molecules to Men*. Cambridge: Cambridge University Press.

——— (1996), *Climbing Mount Improbable*. New York: Norton.

Dennett, D. (1995), *Darwin's Dangerous Idea*. New York: Simon and Schuster.

Gillespie, N. (1979), *Charles Darwin and the Problem of Creation*. Chicago: University of Chicago Press.

Gould, S. (2002), *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.

Gould, S., and R. Lewontin (1979), "The Spandrels of San Marco and the Panglossian Paradigm—A Critique of the Adaptationist Programme," *Proceedings of the Royal Society of London B* 205: 581–598.

Harvey, P., and M. Pagel (1991), *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.

Hodge, M. J. S. (1992), "Natural Selection: Historical Perspectives," in E. Fox Keller and E. Lloyd (eds.), *Keywords in Evolutionary Biology*. Cambridge, MA: Harvard University Press, 212–219.

Huxley, T. (1893), *Evolution and Ethics*. London: Macmillan.

Kimura, M. (1983), *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Kingsolver, J. G., and Koehl, M. A. R. (1986), "Aerodynamics, Thermoregulation, and the Evolution of Insect Wings: Differential Scaling and Evolutionary Change," *Evolution* 39: 488–504.

Laplace, P. ([1814] 1951), *A Philosophical Essay on Probabilities*. New York: Dover.

Lewontin R. (1970), "The Units of Selection," *Annual Review of Ecology and Systematics* 1: 1–18.

Lloyd, E. (2003), *Something about Eve: Bias in Evolutionary Explanations of Women's Sexuality*. Cambridge, MA: Harvard University Press.

Mills, S., and J. Beatty (1977), "The Propensity Interpretation of Fitness," *Philosophy of Science* 46: 263–288.

Neander, K. (1995), "Pruning the Tree of Life," *British Journal for the Philosophy of Science* 46: 59–80.

Orzack, S., and E. Sober (2001), "Adaptation, Phylogenetic Inertia, and the Method of Controlled Comparisons," in *Adaptationism and Optimality*. Cambridge: Cambridge University Press, 45–63.

Rosenberg, A. (1994), *Instrumental Biology or the Disunity of Science*. Chicago: University of Chicago Press.

Ruse, M. (1997), *Monad to Man: The Concept of Progress in Evolutionary Biology*. Cambridge, MA: Harvard University Press.

Sober, E. (1984), *The Nature of Selection*. Cambridge, MA: MIT Press.

——— (1993), *Philosophy of Biology*. Boulder, CO: Westview Press.

——— (1997), "Two Outbreaks of Lawlessness in Recent Philosophy of Biology," *Philosophy of Science* 64: S458–S467.

——— (1999), "Physicalism from a Probabilistic Point of View," *Philosophical Studies* 95: 135–174.

——— (2001), "The Two Faces of Fitness," in R. Singh, D. Paul, C. Krimbas, and J. Beatty (eds.), *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*, vol. 2. Cambridge: Cambridge University Press, 309–321.

Sober, E., and D. S. Wilson (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Sterelny, K., and P. Kitcher (1988), "The Return of the Gene," *Journal of Philosophy* 85: 339–361.

See also **Adaptation; Altruism; Evolution; Fitness; Population Genetics; Species**

NATURALISM

See **Epistemology; Evolutionary Epistemology; Quine, Willard Van**

NATURALIZED EPISTEMOLOGY

See **Epistemology; Evolutionary Epistemology**

JOHN VON NEUMANN

(28 December 1903–8 February 1957)

It is widely acknowledged that John von Neumann was a unique scientific genius. His contributions covered the most diverse domains of pure and applied mathematics: from the axiomatization of set theory to the numerical analysis of nonlinear partial differential equations using computers. He chiefly influenced the style of modern mathematical physics and laid the mathematical foundations of quantum physics; his work on game theory opened up a whole new branch of mathematical economics; he played a major role in the design and theoretical understanding of the first computers; and his last writings have stimulated debates in artificial intelligence. During and after the Second World War, in particular under the Eisenhower presidency, von Neumann played an increasingly important role on a large number of military advisory committees.

Von Neumann's importance for philosophy of science, both historically and systematically, emerges from the foundational nature of some of the theories he developed, the universal outlook of his writings, the philosophical significance of a mathematically rigorous analysis of basic theoretical concepts, and a few explicitly methodological papers. Philosophers have often been intrigued by the visionary or utopian nature of some of his ideas. Yet these utopias were typically coined in rather precise mathematical terms, so that they amounted to highly motivating mathematical conjectures of the greatest possible scope. No wonder that the organizers of the 1954 International Congress of Mathematicians invited von Neumann to deliver "an address of the same nature as Hilbert's famous address in 1900" because they considered him as "probably the only active mathematician in the world who is master of mathematics to such a degree" (Rédei and Stöltzner 2001, 227). Von Neumann's answer was simultaneously modest and utopian. Instead of twenty-three problems, he mentioned a single field close to his heart—the role of operator theory in quantum mechanics—and he developed it into a quest for a new relationship between logic and probability.

In his later years, von Neumann's far-reaching optimism about the role of mathematics in the sciences—from numerics to Hilbert's axiomatic method—was complemented by the then widespread idea that technology was able to control practically all aspects of politics and society. In this vein, he advocated game theory as a tool for strategic analyses and contemplated the idea of global weather control (von Neumann 1955).

From Mathematical Wunderkind to Military Strategist

Von Neumann usually impressed his company by his exceptional calculational powers and his encyclopedic memory, which reached far beyond his scientific interests. It has often been observed that his greatest pleasure was simply thinking.

János (Jáncsi) Neumann was born in Budapest as the first of three brothers into the family of the banker Max Neumann, who was given a title by the emperor Franz Josef I in 1913. At first educated privately, above all in foreign languages, he entered the Lutheran gymnasium, where his extraordinary mathematical abilities prompted his teacher to arrange tutoring by university professors. Still at school, he coauthored his first mathematical paper. Of great importance to the young von Neumann were the table conversations in his father's house, where he came to meet eminent Budapest scientists, artists, and businessmen (cf. Macrae 1992; Vonneumann 1987).

In addition to his mathematical studies at Berlin, Göttingen, and Budapest (until 1926), von Neumann took a degree in chemical engineering at the Technical University in Zurich (in 1925). Apart from the Budapest mathematicians, he was influenced by Erhard Schmidt, Weyl, and the Hilbert school. In 1927 he was appointed as a Privatdozent at the University of Berlin, where he interacted with Schrödinger. In 1929, he moved to Hamburg, but he began to believe that his prospects to obtain a chair in Germany, let alone in Hungary, were dim. After visiting Princeton University in the United States the following year, he was offered a professorship there. In 1933, von Neumann became one of the four founding permanent faculty members of the Institute for Advanced Study and remained there until his untimely death from bone cancer.

Shortly after his naturalization as a United States citizen in 1937, von Neumann became actively involved in military research into problems of ballistics, explosion shock waves, the implosion bomb, the hydrogen bomb, and nuclear missiles, among others. The respective nonlinear problems occupied the various computer projects he was associated with. In 1943, he arrived at Los Alamos to work on the Manhattan Project. He was also spending most of his time on an increasing number of advisory committees. In 1954, President Eisenhower appointed him to the Atomic Energy Commission. In 1956, some of the military leaders of these committees gathered around his bed in Walter Reed Hospital.

Among the reasons for this striking influence, his friend Ulam counts his ability "to commune with the physicists, understand their language, and to transform it almost instantly into a mathematician's schemes and expressions. Then, after following the problems as such, he could translate them back into expressions in common use among physicists" (Oxtoby, Pettis, and Price 1958, 37). Admiral Elliott B. Strauss praised his "invaluable faculty of being able to take the most difficult problem, separate it into its components, whereupon everything looked brilliantly simple, and all of us wondered why we had not been able to see through to the answer as clearly as it was possible for him to do" (Oxtoby et al. 1958, 4). During the Cold War years, von Neumann, appalled by Stalinism and totalitarianism, advocated a strategy of military strength; he criticized the opinions of those of his former colleagues on the Manhattan Project who campaigned for global disarmament.

Set Theory and the Foundational Program

Von Neumann's (1925 and 1928) first major breakthrough was his axiomatization of set theory. While Zermelo-Fraenkel (ZF) set theory used an axiom scheme that generated infinitely many axioms, von Neumann made do with finitely many. Unlike Zermelo, von Neumann held that the naive intuition of a set was of minor importance for the choice of the proper axiomatization.

Basic to von Neumann's axiom system were the concept of class and the element relation. This was in contrast to the program of naive set theory advocated by Cantor and by Hausdorff, for whom all predicate-extensions were considered as sets. Von Neumann's axiom system avoided the set-theoretic antinomies because the classes formed in them need not be sets. Thus von Neumann did not restrict the scope of predicate-extensions but allowed certain predicates to specify only classes. His works also motivated the set theories of Bernays and Gödel (with classes and sets as two different entities), which are intimately related in the sense that either all three are consistent or none is. The theory of classes is only a conservative extension of ZF, and no axiomatization of naive set theory has yet been given—as, for instance, Quine's New Foundations intended.

A crucial step in Hilbert's axiomatic method was to prove the consistency of an axiom system, be it set theory or relativity theory, relative to the consistency of arithmetic (see Hilbert, David). In order to establish the latter, and thus to provide an absolute justification for mathematical knowledge, Hilbert developed a new proof theory, which he called *meta-mathematics.* Von Neumann (1927a) wrote a long paper that discussed the general aspects of this program and brought some new rigorous results. Seen in retrospect, some of them came close to indicating the actual limits of Hilbert's original program.

At the 1930 Königsberg meeting organized by the Vienna Circle, von Neumann (1931b) acted as the advocate of Hilbert's formalism. When in discussions there, Gödel first mentioned his incompleteness result in public, and von Neumann quickly admitted that Hilbert's foundationalist program in its original form had become infeasible, though not based on wrong intentions. However, Gödel's results, which von Neumann quickly reproduced and extended, did not tarnish the value of the axiomatic method in mathematics and the empirical sciences because, as von Neumann (1947) would put it, the concepts of classical mathematics, "stood on at least as sound a foundation as, for example, the existence of the electron. Hence, if one is willing to accept the sciences, one might as well accept the classical system of mathematics" (6). Mathematics did not amount to just abstraction and absolute rigor. It was instead characterized by its peculiar relationship to the empirical sciences. "Some of the best inspirations of modern mathematics (I believe, the best ones) clearly originated in the natural sciences" (2). In virtue of this intimate connection, mathematics and the empirical sciences shared many pragmatic criteria of success, among them unificatory power and simplicity, but mathematics additionally required "elegance" in its "'architectural," structural makeup (Stöltzner 2001).

Quantum Physics

In a 1954 questionnaire for the National Academy of Sciences, von Neumann listed as his three most important contributions: operator theory, the foundations of quantum mechanics (QM), and the ergodic theorem. The first two originated in his 1927 works with Hilbert and Nordheim and were elaborated in the 1932 book *Mathematical Foundations of Quantum Mechanics.* On the surface, the book's aim was to give a mathematically satisfactory formulation of QM in terms of operators in Hilbert space without resorting to Dirac's then ill-defined delta functions. But its lasting importance was to supply a concise and mathematically rigorous reference frame for subsequent debates about the interpretations of quantum mechanics.

Most influential were the "no-hidden-variable theorem" and the theory of measurement. Taking causality as tantamount to complete determination, von Neumann phrased the notorious causality problem of QM as a question of fact that could be resolved mathematically, accepting certain axioms. Since the dispersion of quantum mechanical ensembles was beyond doubt, the only way to restore causality was to posit hidden parameters, not contained in QM, that permitted a further subdivision of the ensembles. All such attempts contradicted the axioms assumed by von Neumann, and he consequently surmised that any causal modification of QM would amount to a drastically different theory of atomic phenomena. Contrary to his expectations, such modifications can recover quantum mechanical predictions, and Bell's analysis (1966) pinpointed the part of von Neumann's axiom system that prevented this from happening (see Locality). It was too restrictive to assume that if the physical quantities $R, S, \ldots$ have the operators $\boldsymbol{R}$, $\boldsymbol{S}$, ..., then the quantity $R + S + \ldots$ has the operator $\boldsymbol{R} + \boldsymbol{S} + \ldots$ Physically, this was reasonable when results of measurement were identified with properties of isolated systems, but "quite unreasonable when one remembers with Bohr 'the impossibility of any sharp distinction between the behaviour of atomic objects and the interaction with the measuring instruments which serve to define the conditions under which the phenomena appear'" (Bell 1966, 447).

Von Neumann's theory of measurement was based on the distinction between the irreversible measurement process and the deterministic unitary time evolution in between. Let I be the object system, II the measuring device, and III the observer. Von Neumann showed that the boundary, the Heisenberg cut, could be placed either between I+II and III or between I and II+III. Von Neumann emphasized that III remains outside any calculation:

> [f]or, subjective perception leads us out of the latter, or more precisely: it leads into the intellectual inner life of

> the individual, which is uncontrollable since it must be taken for granted by any attempt of [empirical] control. . . . Nevertheless, it is a fundamental requirement of the scientific world view—the so-called principle of the psycho-physical parallelism—that it must be possible so to describe the (in reality) extra-physical process of the subjective perception as if it would occur in the physical world, i.e., to assign to its parts equivalent physical processes in the objective environment, in ordinary space. ([1932] 1955, 223f.).

That the Heisenberg cut could be shifted arbitrarily between I and III was interpreted as suggesting a psycho-physical principle. The "as if" in the above quotation has sparked debates as to whether von Neumann took the collapse of the wave packet to be a physical process (Barrett 1999) or not (Becker 2004), and what the philosophical content of the principle of psycho-physical parallelism was. To some extent, such an alternative ascribes too realist an aspiration to the early von Neumann. As strange as it seems in retrospect, he advocated both a descriptivist approach in the style of logical empiricism ([1932] 1955, Chs. 1–4) and a metaphysical subject/object distinction (ibid, Chs. 5–6) that blatantly contradicted the empiricist criterion of meaning (See Quantum Measurement Problem; Vienna Circle).

Motivated by problems with a frequentist interpretation of quantum probabilities, shortly after 1932, von Neumann became dissatisfied with the Hilbert-space framework and investigated what are today called "type II^1 von Neumann algebras" (cf. Rédei 1997). Physically, this amounted to considering the quantum theory of systems with infinitely many particles as more fundamental than QM. Von Neumann also developed—though this was initially hardly noticed by physicists—the mathematical means for a more general algebraic approach to quantum physics: the infinite-dimensional tensor product, operator algebras or rings of operators, nondistributive lattices or quantum logic, and continuous geometry. The Stone–von Neumann uniqueness theorem rigorously established that in QM all representations of the canonical commutation relations for a finite number of degrees of freedom are equivalent up to isomorphism. This also gave a rigorous justification to the equivalence of the Heisenberg and Schrödinger version of the theory. Von Neumann (1931a) worked with the Weyl form of the position and momentum operators

$$\left(U(s) = \exp(-\frac{i}{\hbar}sx), U(r) = \exp(-\frac{i}{\hbar}rp); s, r \in \Re\right)$$

because the unbounded operators themselves give rise to a plethora of domain problems, which might even yield inequivalent representations that are not physically pathological (see Summers 2001; Thirring 1981). It took some time until physicists realized that the existence of inequivalent representations is quite the standard case in quantum statistical mechanics, which involves infinitely many degrees of freedom (e.g., the theory of ferromagnetism), and quantum field theory, where the Fock representation in which the theory is built up by creating and annihilating particles in the vacuum is not equivalent to most interacting interpretations (see Quantum Field Theory).

The great value of the operator algebras developed by Murray and von Neumann (1936) became clear only during the renaissance of mathematical physics in the 1960s. They not only exhibit a rich variety of algebraical and topological features but also permit a unified approach to problems of quantum mechanics, quantum statistical mechanics, and quantum field theory. As the Hepp-Bell debate showed (Hepp 1972; Bell 1975), this approach could also venture a stand on the measurement problem. Over the years the C*-algebraic quantum field theory developed by Haag and Kastler has produced many rigorous results about the conceptual structure of quantum field theory, among them the issues of locality, spontaneous symmetry breaking, and gauge invariance. An increasing number of philosophers of science avail themselves of this approach in analyzing the structure of quantum field theory (e.g., Clifton and Halvarson 2001) (see Quantum Field Theory).

In the same year, Birkhoff and von Neumann (1936) published their seminal paper on quantum logic. In contrast to von Neumann algebras, quantum logic, or the study of nondistributive orthomodular lattices, took off much quicker and has since become a research field in its own right (see Quantum Logic). While the subsequent developments would emphasize the logical aspects, von Neumann's own conception of quantum logic remained intimately related to his research program in axiomatic quantum physics. Already in the *Mathematical Foundations*, he held that "that the concept of 'simultaneous decidability' represents a refinement of the concept of 'simultaneous measurability'" (von Neumann [1932] 1955, 134). With von Neumann algebras in place and the frequency interpretation of probability abandoned at the end of the 1930s, von Neumann kept searching for a way to "interpret the algebraic structure representing quantum logic as the algebra of random events in the sense of a non-commutative probability theory" (Rédei and Stöltzner 2001, 154).

Ever since Boltzmann's statistical derivation of the second law of thermodynamics, the problem of ergodicity has enjoyed philosophical significance. Historically, the ergodic hypothesis was the indispensable link between the deterministic kinetic theory of gases and the indeterministic behavior macroscopically observed in phenomena, such as Brownian motion, that were governed by the second law of thermodynamics. When the precise nature of the problem became clearer through the work of the Ehrenfests, it quickly turned out that Boltzmann's original version was too strong for realistic systems. Boltzmann demanded that the orbit of a mass point densely cover the energy shell, so that the mean value of a physical quantity on this shell (mathematically speaking, any real-valued function on phase space) was equal to the time average. G. D. Birkhoff (1931) and von Neumann (1932) found a more appropriate definition of ergodicity based on the partition of states following which a dynamically invariant state was ergodic if it could not be decomposed further into invariant states, and proved the pointwise and mean ergodic theorems, respectively. Also, the stronger condition of mixing, when the time average is replaced by the limit, goes back to von Neumann. Subsequently, other notions of ergodic behavior were introduced, and ergodic theory has become an important measure-theoretic tool for analyzing all sorts of dynamical systems (Mackey 1990). Von Neumann (1927b; [1932] 1955) also gave a definition for the entropy of a quantum mechanical system that showed the difference between the classical and quantum concepts. The field has since developed in various directions involving the relationship between entropy and information, the characterization of chaotic behavior, and a rigorous analysis of large quantum systems. Von Neumann's operator-theoretic methods, for instance, permitted an extension of the concept of equilibrium to infinite-dimensional systems (see Thirring 1983).

Game Theory and the Expanding Economy Model

"Among the many areas of mathematics shaped by this genius, none shows more clearly the influence of John von Neumann than the Theory of Games" (Kuhn and Tucker 1958, 100). When von Neumann ([1928] 1959) published his first paper on game theory, there existed, in contrast to quantum mechanics, neither well-entrenched scientific concepts nor a ready-to-use mathematical theory such as Hilbert's theory of integral equations. The introduction to *Theory of Games and Economic Behavior,* coauthored with the economist Morgenstern (von Neumann and Morgenstern 1944), was quite explicit that "mathematical discoveries of a stature comparable to that of calculus will be needed in order to produce decisive success in this field" (6). Rather than importing an unsatisfactory theory formulated in terms of mathematical concepts useful in physics, economists should strive for a mathematically precise formulation of elementary facts about simple games and, constantly comparing them with empirical evidence, approach realistic situations.

Von Neumann's most important achievement in game theory was the *minimax theorem.* It originally stated that every two-person, zero-sum game with finitely many pure strategies has a determined solution. Although in virtue of further developments its reputation changed from being "considered the elegant centerpiece of game theory" to an overly special case, it is beyond doubt that "the most fundamental concepts of the general theory—extensive form, pure strategies, strategic form, randomization, utility theory—were spawned in connection with the minimax theorem" (Aumann 1987, 6f). The *Theory of Games* focused on games of 3, 4, and ultimately *n* players, of both zero-sum and non-constant sum varieties. It took cooperation as a given and developed a solution concept, the stable set, which showed how any game could give rise to various stable coalition formations, depending on the payoffs available and assuming that a coalition played a 2-person, zero-sum, and thus minimax, game against its nonmembers (see Game Theory).

Through the work of von Neumann and John Nash, game theory became a general methodology that in principle applied to all kinds of interactive situations, from the formation of an industrial oligopoly to the emergence of the social contract, and had consequences for philosophical ideas about rationality (especially decision making) and ethics. More than his contributions discussed so far, von Neumann's works on game theory were embedded into the sociopolitical context of the day. His development of game theory in the late 1930s was sparked by sociopolitical developments in his native Hungary, and parts of the theory soon found application to operations research problems during World War II (see Leonard 2005).

Von Neumann also strongly furthered the development of mathematical economics by his expanding economy model. It is a closed pure production model with a profitless economy in which overproduced goods are free, and inefficient processes are not used. Assuming that every good appears either as output or input and that the total value of all goods is positive, von Neumann could prove the existence of a constantly expanding economy.

Mathematically, the problem corresponded to a system of inequalities and required a generalization of Brouwer's fixed point theorem. By interpreting the quantities and prices as vectors of a mixed strategy in a zero-sum game, von Neumann's model can be related to the minimax theorem. There is some disagreement in the literature as to whether von Neumann's expanding economy was "unlike any other economic model that preceded it" (Thompson 1987, 248) or whether "his contribution is fundamentally a technical one" (Dore, Chakravarty, and Goodwin 1989, 6) because essential features had been discovered before. Scholars also disagree whether it extends the tradition of classical political economy or represents a special case of the Arrow-Debreu-McKenzie model of pure exchange.

Computers, Automata, and the Brain

There are disagreements about the relative credits for the development of the first modern electronic computer. Goldstine (1983) puts von Neumann on top because he not only made substantial contributions, among them the stored-program concept, but also integrated all of them into a coherent whole. Although the computer at the Institute of Advanced Study (IAS) was never the most powerful machine, its logical design and the IAS reports deeply influenced a rapid and decentralized development around the world (cf. Aspray 1990; Heims 1980).

"Von Neumann is the father of modern scientific computing. His work in the major areas of this field—numerical analysis, numerical algorithms, computations, mathematical modeling, and asymptotic analysis—stands today as vital and seminal" (Glimm, Impagliazzo, and Singer 1990, 185). This gave applied mathematicians important tools to study realistic nonlinear problems instead of simplified toy models, and led to a substantial reconfiguration of the terrain between fundamental theory and actual experiment. New topics of interest to philosophers of science thus emerged; they include chaos theory, design and evaluation of large experiments on the computer, and the questions of what a simulation actually shows.

Von Neumann also analyzed the abstract foundations of computing. In this he could build on ideas that had emerged in the foundational debates, such as Gödel numbering or the universal Turing machine (see Turing, Alan). A von Neumann machine, as it was later called, consists of five components: the finite uniform memory in which instructions are stored, a central control unit that interprets instructions, a central arithmetic unit that executes operations on the data contained in memory, and the input and output organs. Memory contents can be either instructions, if executed, or data, if processed. As the machine can thus modify its own program during execution, von Neumann (1993) thought it to be brain-like. During the last years of his life, von Neumann attempted to develop a general theory of information processing automata that embraced both the technological and the biological realm and could thus serve as a rigorous and highly abstract basis to study the similarities and dissimilarities between computers and the brain. In this project he also employed the neural networks of McCulloch and Pitts (see Artificial Intelligence).

An important part of von Neumann's work was the theory of reliable computation with unreliable components, which he modeled by adjoining a probability space to a network model. In this way the failure of the components was determined in the mean. Although this thermodynamic approach did not quite succeed, it provided important insights into how complex structures could still function reliably even if their components are as unreliable as nerve cells.

Complexity studies also triggered von Neumann's (1966) theory of self-reproducing automata. Its aim was not just to produce identical copies of an automaton but to find conditions under which machines were capable of evolution, that is, of interacting with the environment and producing machines of increasing complexity. "A real difficulty here is that of striking the proper balance between formal simplicity and ease of manipulation, on the one hand, and approximation of the model to real physical machines, on the other hand" (Shannon 1958, 126). Artificial life, as this subject is called nowadays, could be trivial or intractable. Von Neumann investigated both the idea of a universal machine capable of constructing any automaton with finite means and a more specific model based on reproducing cells that had 29 possible states and interacted with their nearest neighbors. Today there exists a variety of self-reproducing automata, most famous among them the Game of Life (see McMullin 2000).

Already suffering from his fatal illness, von Neumann (1958) completed a small booklet in which he compares a serial digital computer and the neural machinery in the brain from the standpoint of information theory. Although neurons appear to operate digitally, their rather complex structure and their dependence on certain global potentials make them take on features of an analog machine as well. By a series of comparative numerical estimates, he concluded that the neural machinery compensates missing logical depth, and poor reliability of its components by logical breadth, that is by a highly parallel architecture.

Assuming that the brain contains both a language proper based on logic and a mathematical language based on numbers, von Neumann (1958) advanced a provocative thesis on the foundations of mathematics. Since the arithmetical part of neuronal activity was of an essentially statistical character and was short of arithmetical depth, "the nervous system appears to be using a radically different system of notation from the ones we are familiar with in ordinary arithmetics and mathematics" (79). Distinguishing the complete code (machine language) and the short code (high-level programming language), he adopted the idea that a Turing machine can imitate the behavior of any other machine. Mathematics thus may well be just a short code, "a secondary language, built on the primary language truly used by the central nervous system" (82). "Just as languages like Greek or Sanskrit . . . it is only reasonable to assume that logics and mathematics are similarly historical, accidental forms of expression" (81).

To be sure, von Neumann advocated neither the contingency of mathematical truth nor a radically empiricist stand in the foundations of mathematics. One must rather view his "systematized set of speculations" (von Neumann 1958, 1) as a continuation of the pragmatist or opportunist conception of the axiomatic method outlined in *The Mathematician* (von Neumann 1947).

Michael Stöltzner

References

Aspray, William (1990), *John von Neumann and the Origins of Modern Computing*. Cambridge, MA: MIT Press.

Aumann, R. (1987), "Game Theory," in John Eatwell, Murray Milgate, and Peter Newman (eds.) *The New Palgrave Dictionary of Economics*. New York: W. W. Norton, 460–482.

Barrett, Jeffrey A. (1999), *The Quantum Mechancs of Minds and Worlds*. Oxford: Oxford University Press.

Becker, Lon (2004), "That von Neumann Did Not Believe in a Physical Collapse," *British Journal for the Philosophy of Science* 55: 121–135.

Bell, John S. (1966), "On the Problem of Hidden Variables in Quantum Mechanics," *Reviews of Modern Physics* 38: 447–452.

——— (1975), "On Wave Packet Reduction in the Coleman-Hepp Model," *Helvetica Physica Acta* 48: 93–98.

Birkhoff, G. D. (1931), "Proof of the Ergodic Theorem," *Proceedings of the National Academy of Sciences USA* 17: 656–660.

Birkhoff, Garrett, and John von Neumann (1936), "The Logic of Quantum Mechanics," *Annals of Mathematics* 37: 823–843.

Clifton, Robert, and Hans Halvorson (2001), "Entanglement and Open Systems in Algebraic Quantum Field Theory," *Studies in History and Philosophy of Modern Physics* 32: 1–31.

Dore, Mohammed, Sukhamoy Chakravarty, and Richard Goodwin (eds.) (1989), *John von Neumann and Modern Economics*. Oxford: Clarendon Press.

Glimm, James, John Impagliazzo, and Isadore Singer (eds.) (1990), *The Legacy of John von Neumann*. Providence, RI: American Mathematical Society (Proceedings of Symposia in Pure Mathematics, vol. 50).

Goldstine, Herman H. (1983), "The Role of John von Neumann in the Computer Field," in Rutherford Aris, H. Ted Davis, and Roger Stuewer (eds.), *Springs of Scientific Creativity: Essays on Founders of Modern Science*. Minneapolis: University of Minnesota Press, 308–327.

Heims, Steve J. (1980), *John von Neumann and Norbert Wiener: From Mathematics to the Technologies of Life and Death*. Cambridge, MA: MIT Press.

Hepp, Klaus (1972), "Quantum Theory of Measurement and Macroscopic Observables," *Helvetica Physica Acta* 45: 237–248.

Kuhn, H., and Tucker, A. (1958), "John von Neumann's Work in the Theory of Games and Mathematical Economics," in J. C. Oxtoby, B. J. Pettis, and G. B. Price (eds.), *John von Neumann 1903–1957*. Special issue of the *Bulletin of the American Mathematical Society* 64: 100–122.

Leonard, Robert (2005), *From Red Vienna to Santa Monica: Von Neumann, Morgenstern, and Social Science, 1925–1960*. Cambridge and New York: Cambridge University Press.

Mackey, G. W. (1990), "Von Neumann and the Early Days of Ergodic Theory," in James Glimm, John Impagliazzo, and Isadore Singer (eds.), *The Legacy of John von Neumann*. Providence, RI: American Mathematical Society (*Proceedings of Symposia in Pure Mathematics,* vol. 50), 25–38.

Macrae, Norman (1992), *John von Neumann*. New York: Pantheon.

McMullin, Barry (2000), "John von Neumann and the Evolutionary Growth of Complexity: Looking Backward, Looking Forward," *Artificial Life* 6: 347–361.

Murray, F. J., and John von Neumann (1936), "On Rings of Operators," *Annals of Mathematics* 37: 116–229 (*CW* vol. 3, 6–119).

Oxtoby, J. C., B. J. Pettis, and G. B. Price (eds.) (1958), *John von Neumann 1903–1957*. Special issue of the *Bulletin of the American Mathematical Society* 64.

Rédei, Miklós (1997), "Why von Neumann Did Not Like the Hilbert Space Formalism of Quantum Mechanics (and What He Liked Instead)," *Studies in History and Philosophy of Modern Physics* 27: 493–510.

Rédei, Miklós, and Michael Stöltzner (eds.) (2001), *John von Neumann and the Foundations of Quantum Physics*. Dordrecht, Netherlands: Kluwer.

Shannon, C. E. (1958), "Von Neumann's Contributions to Automata Theory," in J. C. Oxtoby, B. J. Pettis, and G. B. Price (eds.), *John von Neumann 1903–1957*. Special issue of the *Bulletin of the American Mathematical Society* 64: 123–129.

Stöltzner, Michael (2001), "Opportunistic Axiomatics: Von Neumann on the Methodology of Mathematical Physics," in Miklós Rédei and Stöltzner (eds.), *John von*

Neumann and the Foundations of Quantum Physics. Dordrecht, Netherlands: Kluwer, 35–62.

Summers, S. J. (2001), "On the Stone–von Neumann Uniqueness Theorem and Its Ramifications," in Miklós Rédei and Michael Stöltzner (eds.), *John von Neumann and the Foundations of Quantum Physics*. Dordrecht, Netherlands: Kluwer, 135–152.

Thirring, Walter (1981), *A Course in Mathematical Physics*, Vol. 3: *Quantum Mechanics of Atoms and Molecules*. New York and Vienna: Springer.

——— (1983), *A Course in Mathematical Physics*, Vol. 4: *Quantum Mechanics of Large Systems*. New York and Vienna: Springer

Thompson, G. L. (1987), "von Neumann, John," in John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave Dictionary of Economics*. New York: W. W. Norton, 818–822.

von Neumann, John (1925), "Eine Axiomatisierung der Mengenlehre," *Journal für reine und angewandte Mathematik* 154: 219–240. (All articles by von Neumann listed in this bibliography have been reprinted in Abraham H. Taub [ed.] [1961–1963], *Collected Works* (*CW*), 6 vols. Oxford, UK: Pergamon Press.)

——— (1927a), "Zur Hilbertschen Beweistheorie," *Mathematische Zeitschrift* 26: 1–46.

——— (1927b), "Thermodynamik statistischer Gesamtheiten," *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 273–291.

——— ([1928] 1959), "On the Theory of Games of Strategy," in A. W. Tucker and R. D. Luce (eds.), *Contributions to the Theory of Games*, vol. 4. Princeton, NJ: Princeton University Press, 13–43. Originally published as "Zur Theorie der Gesellschaftspiele," *Mathematische Annalen* 100: 295–320.

——— (1928), "Die Axiomatisierung der Mengenlehre," *Mathematische Zeitschrift* 27: 669–752.

——— (1931a), "Die Eindeutigkeit der Schrödingerschen Operatoren," *Mathematische Annalen* 104: 570–578 (*CW* vol. 2, 221–229).

——— (1931b), "Die formalistische Grundlegung der Mathematik," *Erkenntnis* 2: 116–121.

——— ([1932] 1955), *Mathematische Grundlagen der Quantenmechanik*. English translation. Princeton, NJ: Princeton University Press. (In some key philosophical passages, this translation misrepresents von Neumann's original wording.)

——— (1932), "Proof of the Quasi-Ergodic Hypothesis," *Proceedings of the National Academy of Sciences USA* 18: 70–82.

——— (1937), "Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes," *Ergebnisse eines mathematischen Kolloquiums* 8: 73–83. English translation: "A Model of General Economic Equilibrium," *Review of Economic Studies* 13 (1945): 1–9.

——— (1947), "The Mathematician," in *The Works of the Mind*. Edited by R. B. Heywood. Chicago: University of Chicago Press, 180–196.

——— (1955, June), "Can We Survive Technology?" *Fortune* 51: 106–108 and 151–152.

——— (1958), *The Computer and the Brain*. New Haven, CT: Yale University Press.

——— (1966), *Theory of Self-Reproducing Automata*. Edited and completed by Arthur W. Burks. Urbana: University of Illinois Press.

——— (1993), "First Draft of a Report on the EDVAC," *IEEE Annals of the History of Computing* 15: 27–75.

von Neumann, John, and Oskar Morgenstern (1944), *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Vonneumann Nicholas (1987), *John von Neumann As Seen by His Brother*. Meadowbrook, PA: Author.

OTTO NEURATH

(10 December 1882–22 December 1945)

Otto Neurath was born in Vienna, the son of Wilhelm Neurath, the political economist and social reformer. After initially studying mathematics and physics, and then history, philosophy, and economics, he followed Ferdinand Toennies' advice to move to Berlin, where he received a doctoral degree in the history of economics in 1906. He studied under Eduard Meyer and Gustav Schmoller, and he was awarded the degree for two studies of economic history of antiquity, one on Cicero's *De Officiis* and the other with an emphasis on the nonmonetary economy of Egypt. With a grant from the Carnegie Endowment for International Peace, a subsequent study of the Balkan wars before World War I led to his theory of war economy as a natural (nonmonetary) economy. In 1919, the short-lived Bavarian socialist government appointed him head of the Central Planning Office. His program for full socialization was based on his theory of natural economy and a holistic requirement to bring different institutions and kinds of knowledge together in order to understand, predict, and control the complex phenomena of the social world: unity of science at the point of action.

From 1920 to 1934, Neurath participated actively in the development of Vienna's socialist politics, especially in housing and adult education. He founded the Social and Economic Museum of Vienna, where he developed and applied the "Vienna method" of picture statistics and the ISOTYPE language (International System of Typographic Picture Education). Like the thought of other Viennese philosophers such as Wittgenstein and Popper, Neurath's philosophy was inextricably linked to pedagogical theory, and also, as in Popper's case, political thought (see Popper, Karl Raimund). In 1928, he helped found the *Verein Ernest Mach*, which, with the publication in 1929 of an intellectual manifesto, became the public face of the Vienna Circle (see Vienna Circle). Subsequently he created the International Foundation for Visual Education in The Hague and spearheaded the International Unity of Science Movement. The latter, inspired by a tradition culminating in the French Encyclopedists of the Enlightenment, launched the project of an Encyclopedia of Unified Science (see Unity of Science Movement). As with the pictorial languages, the scientific encyclopedia was supposed to promote scientific and social cooperation and progress at an international level. Neurath fled from the Nazis, first to The Hague, and then, in 1940, to England, where after nine months in an internment camp, he resumed activities related to the ISOTYPE language and the unity of science. He died in Oxford, England (see Cartwright, Cat, Fleck, and Uebel 1996 for more detail).

Science and Society: Empiricism in the Social Sciences (1910–1931)

Neurath championed the "scientific attitude." He denied any value to philosophy over and above the pursuit of work on science, within science, and for science. From this naturalistic viewpoint, philosophy investigates the conditions of the possibility of science as they appear within science itself, namely, in terms of physical, biological, social, psychological, linguistic, logical, mathematical, and other conditions. His views on the language, method, and unity of science were motivated throughout by his interest in the social life of individuals and their well-being. To theorize about society is inseparable from theorizing for and within society. Science is in every sense a social and historical enterprise. It is as much about social objectives as it is about physical objects, and about social realizations as much as about empirical reality.

Neurath drew from the spirit, if not the letter, of two major turn-of-the-century thinkers: Ernest Mach and Karl Marx. Mach introduced a radical antimetaphysical approach to the analysis of science, which Neurath embraced in part because he believed that metaphysical obscurantism, whether in German philosophy or theology, underwrote social institutions that attacked Enlightenment values such as equality, freedom, and progress (see Mach, Ernest). Neurath's aim was to apply an empirical attitude in the social sciences. He attacked the distinction drawn by Dilthey, Rickert, and Weber between the natural and the cultural (or social) sciences, which he thought rested on metaphysical concepts and nonempirical methods (the "empathic method" of understanding). A purely empiricist language would represent a big step toward unified science. Methodologically, the job of the social sciences, like the natural sciences, would be to establish empirical correlations, statistical when possible, about the behavior of social wholes or complexes or, failing that, partial correlations between aspects or parts thereof, to determine the limits of their validity, and to infer predictions about the future (including, as a distinctive feature, the possibility of self-fulfilling or self-refuting prophecies).

Neurath saw in Marxism a model of empirical social science without metaphysics, of a tool for social reform and also for the unification of sociology, political theory, and economics. He never considered himself an orthodox Marxist. However, Marx inspired in him a belief in historical holism, that there is an ineliminable social and historical context of language, concepts, and beliefs. He also inspired a belief in pragmatic holism, the Enlightenment idea that the scientific attitude and the compilation of knowledge possess a practical and socially redemptive (revolutionary) value—the social scientist is also a social engineer. This influence is reflected in Neurath's work in natural economy and his activities as a social planner. During the 1910s, he articulated an ecological program based on central planning and a nonmonetary economy. It was directed not toward the increase of monetary wealth, but toward replacing the anarchy of wasteful capitalistic production with the nonwasteful allocation of exhaustible resources and goods in a way that increased the standard of living of society as a whole ("Epicurean socialism").

Such views led to the so-called calculation debates about rational choice. In 1920, in the earliest one, Ludwig von Mises objected to socialism on the grounds that it precluded the possibility of rational economic action in the absence of a single unit of comparison and the commensurability of different values in terms of that unit. Neurath dismissed as "pseudorational" the view that all decisions are or even can be the outcome of algorithmic, unambiguous procedures of technical calculations in an ideal language. Instead, his method acknowledged the limits of rules of reasoning and required an empirical judgment of needs and goals; it suggested, for instance, "qualitative exactness" as exhibited by the logic of relations. His method was supposed to be based on rationality in place of pseudorationality. Comparison between wholes (plans) could not always rest on a commensurability of their parts:

> The question might arise, should one protect coal mines or put greater strain on men? The answer depends for example on whether one thinks that hydraulic power may be sufficiently developed or that solar heat might come to be better used, etc. If one believes the latter, one may "spend" coal more freely and will hardly waste human effort where coal can be used. If however one is afraid that when one generation uses too much coal thousands will freeze to death in the future, one might use more human power and save coal. Such and many other non-technical matters determine the choice of a technically calculable plan. . . . [W]e can see no possibility of reducing the production plan to some kind of unit and then to compare the various plans in terms of such units. (Neurath 1973, 263)

In general, because reasons underdetermine actions, these could not always be technical decisions, the outcome of a calculus of reasons without attention to local contexts and to ethical and political judgments. In a later debate on calculation, Neurath argued that a centralized program required the coordination of available knowledge, in his case in the form of a plan for a practical, cooperative unification of the sciences.

Neurath's ideas also reflect the influence of the French conventionalist thinkers Poincaré (on underdetermination in geometry) and Duhem (on holism in physics) (see Conventionalism; Duhem Thesis; Poincaré, Henri). Their work was a recurrent topic of discussion among Neurath, Hans Hahn, and Philip Frank in the 1910s, in the so-called First Vienna Circle. Neurath's theoretical holism for the natural and social sciences is the view that a multiplicity of theories are equally supported by the same data (underdetermination) and individual hypotheses cannot be tested in isolation but only accompanied by auxiliary ones (holism). From a purely logical point of view, scientific methodology is open-ended. Just as Neurath rejected idealized languages and procedures in the practice of science, he also rejected overidealistic aims. For Neurath the objective was not to determine which theory was true, but which theory, or combination of theories, should be used in a given context for a given purpose. This may involve a decision that is ethical or political: The social scientist should examine all possible theories, scenarios, and predictions that fit the available data, in order (like an engineer) to design social machines that had never been built. In the 1910s, Neurath conceived of both economic and historical theorizing in a combinatorial fashion, *viz.*, as exploring all possible combinations of given elements.

Neurath rejected all pictures of "ideal" science as gross metaphysics. His attention to the inextricable link between science and society had several implications:

1. The amount of knowledge available can never be exhaustive.
2. The uncertainty involved in justifying decision making "scientifically" can be honestly and rationally eliminated only through the introduction of nonempirical "auxiliary motives" (to deny this limitation on the power of scientific logical justification amounts to pseudorationalism).
3. The empirical language of coordinated scientific practice cannot be precise and atomic, since the introduction of vague terms is as inevitable as it can be useful.
4. Nor can such an empirical language be private, since language is essentially social.
5. Abstract analysis of complex phenomena prompts the adoption of a multiplicity of idealized concepts such as social factors or indicators, which are value laden and historically contingent.

Scientific Language and Scientific Method: The Vienna Circle and Neurath's Logical Empiricism (1931–1935)

Neurath's later views on the language and method of science expressed his simultaneous response to problems in the social sciences and to philosophical issues addressed by the Vienna Circle between 1928 and 1934 and by Karl Popper (see Popper, Karl Raimund; Vienna Circle). A primary aim of the Vienna Circle was to account for the objectivity and intelligibility of scientific method and concepts. Their philosophical approach was to take the so-called *linguistic turn,* that is, to investigate the formal framework of scientific knowledge (the emphasis on language was familiar to Neurath from Toennies' formal approach to sociology and social signs). The dominant, and later popularized, position was that revisable theoretical scientific statements should stand in appropriate logical relations to unrevisable statements about elementary observations (data), called "control sentences" or "protocol sentences" (see Protocol Sentences). Such relations would provide theoretical terms with cognitive meaning or sense, and theoretical statements with verification.

Neurath took the objectivity of scientific knowledge to be provided by the public and social nature of its representations and rules for acceptance. By 1931, his proposal for unifying scientific language was "radical physicalism": Scientific statements must speak of material events and things in space and time (not necessarily in the language of physics) (Neurath 1983, 52–90) (see Physicalism). Physicalism was stimulated by Marx's materialism and also by Neurath's links to the *Neue Sachlichkeit* (New Objectivity or New Factuality), the movement around the Bauhaus School in Dessau (where he and Carnap lectured). Objectivity and the avoidance of metaphysical nonsense require that statements be compared with statements, not with "reality." Protocol statements preserve linguistic empiricism by replacing talk of reality (*contra* Popper and Schlick) and of subjective sense experience (*contra* Schlick and early Carnap). With regard to the latter, he anticipated Wittgenstein's private language argument as central to the controlling function of protocol statements. Neurath motivated the social dimension of language by considering the case of the isolated Robinson Crusoe, whose own successful use of control statements requires intersubjective features of language. The correspondence between Neurath's epistemology and social thought is no surprise. Connections between private experience and private property had been drawn by Berkeley and others. More directly, Marx himself had argued that language was essentially social, and had used the Crusoe example to conclude that Crusoe's planning is a model of control in a moneyless socialized economy.

Neurath's protocol sentences are not atomistic reports of private experiences—Machian "red here now" (Neurath 1983, 91–99). They are syntactically complex; the complexity being in part the result of the physicalist or public status of the information, including the physicalist description of the fact, the spatiotemporal coordinates of the recorded event, a public reference to the observer, and, possibly, the physiological characterization of the experience. Likewise, they are also "rich in theory" and include vague unanalyzable "cluster concepts" (*Ballungen*) from ordinary languages. Lacking the ideal precision of some theoretical terms, scientific language becomes a historically shaped "jargon."

Neurath rejected both Carnap's method of verification and Popper's method, and logic, of falsification. He also rejected the foundationalist tradition of seeking a secure basis for knowledge, whether in a priori principles or in sense data. His Duhemian and historical holism extended to all sciences as well as to logic and mathematics. More importantly, it also extended to the empirical data, that is, to the protocol sentences. Even though protocol sentences provided the necessary stability to scientific inquiry that theoretical hypotheses lacked, their scientific status, insofar as they had any, was underwritten by their testability and revisability when confronted with other protocol sentences. The structure of the protocols captures the possibility of testing. This leads to the Neurath Principle: In case of conflict between a theoretical prediction and a protocol sentence, either could in principle be rejected. The adequacy of this rule is borne out by actual scientific practice. Neurath's naturalism, holism and antifoundationalism are best illustrated by his image of a boat: "We are like sailors who have to rebuild their boat on the open sea, without ever being able to dismantle it in dry-dock and reconstruct it from its best components" (Neurath 1983, 92).

Unity of Science and the Encyclopedia Model (1931–1945)

The boat image also illustrates Neurath's conception of the unification of the sciences as a historical, nonfoundational, and communal enterprise. Despite popular approaches suggesting a hierarchical/pyramidal structure, due to, among others, Comte and Ostwald (and, later, Carnap), from 1910, Neurath's approach to unification was thoroughly antireductionist: cognitively, logically, and pragmatically. Electron talk is irrelevant to understanding and predicting the complex behavior of social groups. He also dismissed the idea of *one* method and *one* ideal language—for instance, of mathematics or physics, to be followed by all the other sciences. In other words, he opposed the ideal of a "system model": an axiomatic, deductively closed, and complete hierarchy. Instead he proposed the "encyclopedia model": a more or less coherent totality of scientific statements at a given time, in flux, incomplete, and with linguistic imprecisions and logical gaps, unified linguistically by the jargon of physicalism, the cooperative and empiricist spirit, and the acceptance of a number of methods or techniques

(Neurath 1983, 145–158, 213–229). Neurath spoke of a "mosaic," an "aggregation," and an interdisciplinary "orchestration" of the sciences (Neurath 1983, 230–242). Correspondingly, his later political writings emphasized internationalism, democracy, and plurality of institutional loyalties.

Jordi Cat

References

Cartwright, N., J. Cat, L. Fleck, and T. E. Uebel (1996), *Otto Neurath: Philosophy between Science and Politics*. Cambridge: Cambridge University Press.

Neurath, O. (1909), *Antike Wirtschafgeschichte*. Leipzig: Teubner.

——— (1910), *Lehrbuch der Volkwirtschaftslehre*. Vienna: Hoelder.

——— (1939), *Modern Man in the Making*. New York: Knopf.

——— (1944), *Foundations of the Social Sciences*, in O. Neurath, R. Carnap, Charles Morris (eds.), *International Encyclopedia of Unified Science*, vol. 2, no.1. Chicago, University of Chicago Press.

——— (1973), *Empiricism and Sociology*. Edited by R. S. Cohen and M. Neurath. Dordrecht, Holland: Reidel.

——— (1981), *Gesammelte philosophische und methodologische Schriften*. Edited by R. Haller and H. Rutte. Vienna: Hoelder-Pichler-Tempsky.

——— (1983), *Philosophical Papers, 1913–1946*. Edited by R. S. Cohen and M. Neurath. Dordrecht, Holland: Reidel.

——— (1991), *Gesammelte bilpaedagogische Schriften*. Edited by R. Haller and R. Kinross. Vienna: Hoelder-Pichler-Tempsky.

——— (1994), *Otto Neurath oder Die Einheit von Wissenschaft und Gesellschaft*. Edited by P. Neurath and E. Nemeth. Vienna: Boehlau.

——— (1998), *Gesammelte oekonomische, soziologische und sozialpolitische Schriften*, vols. 1 and 2. Edited by R. Haller and U. Hoefer. Vienna: Hoelder-Pichler-Tempsky.

——— (2003), *Selected Economic Writings*. Edited by R.S. Cohen and T. E. Uebel. Dordrecht, Holland: Kluwer.

Neurath, O., R. Carnap, and Charles Morris (eds.) (1970), *Foundations of the Unity of Science: Towards an International Encyclopedia of Unified Science*, vols. 1 and 2. Chicago, University of Chicago Press.

See also **Carnap, Rudolf; Conventionalism; Duhem Thesis; Logical Empiricism; Physicalism; Popper, Karl Raimund; Schlick, Moritz; Unity and Disunity of Science; Unity of Science Movement; Vienna Circle**

NEUROBIOLOGY

Most of the issues found in traditional philosophy of science are recapitulated in the philosophy of neuroscience. In particular, philosophers of neuroscience worry about what counts as appropriate empirical justification for a theoretical claim, how to determine which level of organization is the correct one for a scientific explanation, what explanations should look like, whether all explanations will or should reduce to some primitives, and how what is learned about the mind/brain should affect larger social, economic, and political decisions (Bechtel et al. 2001; Schaffner 1993) (see Explanation; Reductionism). In addition, philosophers of neuroscience concern themselves with some traditional aspects of philosophy of mind, including how it is a brain can represent, if it does, and how and whether this representation ties to other notions of representation in cognitive science and beyond (Bechtel et al. 2001) (see Cognitive Science). It is difficult to focus on only one of these concerns to the exclusion of the rest. Most likely, as some particular aspect of the practice of neuroscience becomes understood, others will be as well. What follows discusses these areas of concern as they *differ* from traditional arguments. This discussion therefore should be seen as a complement to the very rich literature in traditional philosophy of science and of mind.

Theory-Laden Observations and Single-Cell Recordings

It is almost a truism in philosophy of science that there is no real distinction between observation and theory. That is, all scientific observations are filtered through and by a prior theoretical framework. Raw data become observations as they are interpreted regarding how they either fit or belie hypotheses (Woodward 1989). In short: what counts as an observation and how that observation functions in the business of science is heavily mediated by theory. In neuroscience in particular, it is easy to change the fundamental nature of observations using accepted methodological techniques for manipulating raw data. The line between good data tinkering and fudging the data is quite thin.

Good data allow scientists to discriminate among competing claims about phenomena (Suppe 1989). The particular practices of the scientific subfield define how to judge whether data are good. Sometimes these practices involve explicit calculations and formal derivations; sometimes they involve matters of personal judgment and skill. The cases in neuroscience involve both. In particular, it is a matter of personal judgment in single-cell recording when to employ certain computational procedures. Different sorting techniques give rise to different data, so which techniques to employ is an important question. But it is not a question for which any good algorithm exists to answer it.

In 1791, Luigi Galvani was the first person who demonstrated the link between neural communication and electrical impulses when he stimulated frogs' legs with electricity and made them twitch. But not too much could be made of this discovery until the 1920s, when scientists developed the technology that allowed them to measure nerve impulses directly using amplified signals sent via electrodes. Not surprisingly, single-cell recording devices have improved much since then. Most importantly, perhaps, E. V. Evarts (1968) developed techniques for recording from single nervous cells in alert moving animals. But it has been only during the last decade or so that neuroscientists have been able to record from the extracellular space of a large number of neurons from awake and behaving animals.

When scientists record with an electrode near a single cell, they do pick up the cell's action potentials, which is what most people think of when they think of neuronal communication. But they also record things that look like action potentials but are actually voltages generated by axonal bundles, or the field potentials from parallel sets of dendrites. Moreover—and especially if the microelectrode has a relatively low impedance—extracellular electrodes pick up signals from several neurons at the same time, recording from all the cells in a nearby area.

The problem is how to differentiate the contributions of the different cells and cell parts with single lump recording. In many cases scientists care only about one particular action potential; the rest, from their perspective, is background noise. The challenge is how to separate what they want from all the electrical signals they do not want. The challenge is how to move from the recordings of the electrode's output to genuine, reliable, and informative data.

This challenge is compounded by the noisy nature of the recordings themselves. Some of the noise is mechanical and comes from the amplifiers, but some is biological and comes from the neurons. Brain cells jitter around constantly (cf. Connors and Gutnick 1990). Neurons are not quiet until they fire off a spike, as some might think. Instead, they are always producing some activity or other. All in all, scientists have to cull their data from quite a din.

Finally, because scientists cannot assume that anything in a recording remains constant, it is difficult to get a theoretical hook into the waveform. Spike shapes can change over time, electrodes can drift during recording sessions, changing position relative to the cells, which would also alter the spike amplitudes, and the electrical properties of electrodes vary with changes in tip condition or background impedance. Gathering data from single-unit activity presents neuroscientists with a serious technical challenge.

In order to get usable data (to get genuine observations) out of what the electrode transmits, scientists must isolate each neuron's contributions to the recorded waveform. They first need to ascertain exactly how many neurons the recorded waveform reflects. How can they do this if they have a mess of overlapping action potentials and field potentials from a variety of cells at different and unknown distances from the electrode? This question becomes particularly vexing if other neurons in the same area have spikes of the same or similar shape and amplitude.

There are several decomposition algorithms, albeit imperfect ones (Lewicki 1998). Each represents a different way to move from raw output to interpreted and interpretable data, giving scientists different ways of refining the waveforms they have recorded so that they can later interpret them. Each is what philosophers are thinking about when they talk about the theory-ladenness of data. Scientists have to choose what to do with their measurements in order to get something that can be scientifically useful. And how they choose is determined by previously accepted theories.

But even with all these advanced sorting techniques, it is still hard to predict the number of neurons eliciting the data. Ideally, scientists would like to claim that one neuron generates each cluster of spikes they have identified, but if the cells are firing in complex bursts, or if there is nonstationary noise, or if the spike trains overlap one another, they cannot get accurate classifications at all. It is simply an unsolved problem how to decompose coincident action potentials with variable spike shapes. The best scientists can do at this point is guess. Their guesses are informed by their years of experience, but they are guesses nonetheless.

Guessing is not quite what philosophers of science have in mind when they talk about the theory-ladenness of observation. Their vision of creating data is one of a more "scientific method." That is, to pull data out of the dial movements or changes in color or squiggles on the page, philosophers generally hold that there is some explicit background theory, devised in some other scientific inquiry, that scientists learn and then use to interpret what they are seeing or measuring as something useful for their studies. But there is a theoretical gap in the move from raw recordings to genuine data, which cannot be filled with any sort of decision-making algorithm. The best scientists can do at this point is simply leap across the gap, on blind faith, with an eye to where they want to go.

Neurophysiology travels in a cognitive circle; scientists use what they know to cull data that support what they believe to be the case. Nevertheless, progress is not stymied. Knowledge accrues in small increments, with each set of single-cell recordings altering the face of what is known a wee bit at a time. Because neurophysiological sorting techniques rely so heavily on previously accepted neurophysiological hypotheses, there will likely never be an abrupt or dramatic conceptual revolution. But what is known can evolve slowly but surely until the final resting position is quite far removed from where it started.

Localization and Reduction

When scientists do single-unit recordings from a set of neurons, they assume that they are busy examining a discrete system. They have been wildly successful using this strategy, identifying at least 36 different topographical visual processing areas in cortex (De Gelder 2000), differentiating the "what" from the "where" object processing streams (DeYoe and Van Essen 1988; Mishkin, Ungerleider, and Macko 1983) and distinguishing motion detection from contour calculations (Barinaga 1995), to name but a few examples. Maps of brain function are getting more and more complicated as more and more is learned about the processing capacities of individual cells. And all these projects are founded on the belief that brains have discrete processing streams that feed into one another.

Yet, the most neurons scientists have ever been able to record from simultaneously are around 150; the most cells they can ever see summed local field potential activity over are a few thousand. But brain areas have hundreds of thousands of neurons, several orders of magnitude more than they can access at any given time. And these neurons are of different types, with different response properties and different interconnections with other cells, including other similar neurons, neurons with significantly different response properties, and cells of other types completely. Any conclusions scientists draw about the behavior of whatever cells they are recording from are going to be limited to very basic stimulus-response and correlation analyses of whatever neuronal subtype they are currently examining. Hence, the functionality they ascribe based on these relatively meager sorts of experiments might be much more restricted than what the cells are actually doing.

They insert an electrode in or near a cell and then record what it does as they stimulate the animal in some fashion. They record from a cell in a vestibular nucleus and then move the animal's head about to see if that changes the activity of the neuron. If it does, then they move it some more or they move it differently and see how that changes the neuronal output. If it does not, then they try either another nearby cell or some other stimulus. But what they cannot do is record from all the neurons in some isolated area, even if the area is very small. And what they cannot do is test any given cell for all the known functional contributions of brain cells in general. So, what they conclude about any cell will reflect only the cells they or others have actually recorded from using stimuli they or others have actually used. This research strategy systematically underestimates when neurons actually respond and under what conditions.

This sort of unit study attempts to combine scores, hundreds, or even thousands of single-unit recordings together to try to analyze the population. Theoretically, perhaps a nervous system region could be stereotaxically delineated if it had reproducible correlations between afferent and efferent connections such that researchers could ultimately articulate the neurophysiological function of the defined region. However, the likelihood of success for this type of study decreases as the complexity of the organism increases. Scientists can draw functional conclusions regarding the activities of neurons in the abdominal ganglia of *Aplysia* or the segmental ganglia of the leech. But the architecture of these organisms' central nervous system is so different from mammals' that the probability of successfully using similar techniques for understanding humans is very low.

In addition, the actual processing of information that goes on in those cells involves lots of different kinds of excitatory and inhibitory inputs from other areas in the brain stem, cerebellum, and cerebral cortex. The dorsal horn is supposed to integrate afferent nociceptive information from the periphery and pass it on to the motor system, but it does not do that segregated from the rest of the brain's activities and ongoing processes. It is integrating and passing on any number and variety of data as the animal is trying to pursue prey or flee from an enemy. Moreover, the brain regions that perform these tasks are often connected to the very area scientists are recording from. The motor system feeds back down into the dorsal horn, as does the thalamus and significant parts of the cortex.

The impact on cognitive processing of such rampant feedback connections in the brain is only now starting to be explored in neuroscientific research, though exactly how to do this remains a difficult question. Neurophysiologists design their experiments keeping in mind the known anatomic connections between and among the relevant structures. At the same time, any actual experimental observations of all the remote influences on the dorsal horn, for example, are impossible, no matter how many individual neurons scientists record from. They simply do not have any way of conducting such extensive, invasive tests on live animals. At best, the particular influences assumed in any particular recording series are a matter of previously accepted gospel, dogma, and faith.

Ideally, neuroscientists try to conjoin their single-cell studies with some sort of lesion experiment. Once scientists construct a general flowchart of the relevant structures based on anatomy experiments and have estimated normal unit behavior from a series of single-cell studies, they then try to knock out the hypothesized functions by placing lesions in otherwise normal animals. They run these experiments based on the assumption that these lesions, placed in regions known to be important, will change the unit behavior of cells they are studying in a consistent fashion. If they witness such a change, they use that information to explain the relative functional contributions of the lesioned region to the cells under scrutiny. In other words, they are using lesion studies to try to derive a functional boxology for the brain, just as cognitive psychologists use reaction time distributions and error measurements to find one for the mind.

But there is a larger theoretical concern. What neuroscientists know, but generally ignore, is that any functional change in the central nervous system will lead to compensatory changes elsewhere. Because it is highly plastic, the brain will compensate elsewhere for a lesion made in one of its sections (e.g., Merzenich et al. 1983). Usually these compensatory sites are not components in the system or region being studied. But even if they are, neuroscientists ignore plasticity of the brain in favor of assuming a consistent functional alteration as caused by the lesion and nothing more. How are investigators supposed to evaluate some observed functional change when the difference they see might have been evoked by the brain's attempt to compensate for its loss and not by any specific deficit induced by the lesion?

The short answer is that they cannot if they are restricted to single-cell recordings and lesion studies. To answer this question, scientists need to be able to see the activity of the entire brain at once and over time. The excitement over functional magnetic resonance imaging and other imaging techniques concerns exactly this point: Researchers do have a way of looking at the activity of the whole brain at one time as tied to some cognitive activity or other. But magnetic resonance imagining, the best noninvasive recording device currently available, has a spatial resolution of only about 0.1 millimeter, and each scan samples about a half second of activity. This imprecision forecloses the possibility of directly connecting single-cell activity—which operates three to four orders of magnitude smaller and faster—with larger brain activation patterns.

Here is how most functional imaging studies work. Experimenters pick two experimental conditions that they believe differ with respect to the cognitive or perceptual process under investigation. They then compare brain activity recorded under one condition with what happens in the second condition, looking for regions whose activity levels differ significantly across the two. These areas, they believe, comprise the neural substrates of the task under scrutiny.

This so-called subtraction method has no way of determining whether the differences found are actually tied to the cognitive process or to something else occurring concurrently but coincidentally. Methodological difficulties with current imaging techniques are now well known and shall not be rehearsed here due to space limitations (see Bechtel 2000; Cabeza and Nyberg 1997 and 2000). Notice that how well the subtraction method will work depends upon the sensitivity of the measuring devices—the worse the instrument is, the better the method seems to be for localization studies. A low signal-to-noise ratio (SNR) means that neuroscientists will find only a few statistically significant differences across conditions, which is just the sort of result they need in order to bolster any claims identifying particular cognitive processes with discrete brain regions.

But as the imaging technology improves and the SNR increases, more and more sites will differ across trials. The more sites that differ, the more it looks as though essentially the entire brain is involved in each cognitive computation. Thus, any assumption of functional specificity in the brain can be justified.

Neuroscience is a victim of imprecise instrumentation. If scientists extrapolate from what they might learn with more sensitive measures, it can easily be seen that there will come a time when this whole approach just will not work anymore. Put in the harshest terms, brain imaging seems to support reductionism because it is not very good yet.

For example, Brodman area 6 appears significantly active after subtraction in studies of phonetic speech processing, voluntary hand and arm movements, sight-reading of music, spatial working memory, recognizing facial emotions, binocular disparity, sequence learning, idiopathic dystonia, pain, itch, delayed response alternation, and category-specific knowledge, to list only a subset of activities in which it is significantly and differentially active. It could be the case that if neuroscientists keep on doing the sort of subtraction studies they are doing currently, then eventually they will find a unifying and pithy way to describe what premotor cortex is doing. In this instance, neuroscience would be on the right track to determining brain function, but scientists still have a long way to go yet. It could also be true that how a region functions depends heavily on the "neural context." Its functional role in a cognitive economy depends on how it is connected to other areas and how those other areas are responding. (The function of these areas would also be dependent on their particular connectivity and the current patterns of activation.) If this is correct, then searching for *the* function of particular areas is misguided, for different brain regions play different roles depending upon the cognitive tasks at hand.

Neuroscientific Theories

Brains are complicated and messy affairs; theories about brains share these same traits. The difficulty is that in order to make a simple generalization about how some aspect of the brain functions, scientists have to retreat to such a broad level of abstraction that their assertions become almost empirically meaningless. In order to make their claims testable in a laboratory, neuroscientists have to confine their ideas to particular animals, to particular experimental tasks, or to both. As a result, they end up with neuroscientific "theories" that contain two distinct parts: a broad statement of a theoretical principle and a set of detailed descriptions of how that principle plays out across different animal models and experimental tasks. Though the detailed descriptions fall under the general principle, they are not immediately derivable from it. Moreover, as described below, the detailed descriptions can be incompatible with one another, though each will maintain a family resemblance with the others.

At a gross level, mammalian brains are remarkably similar to one another. Indeed, the central nervous system (CNS) in invertebrates is not all that different from the mammalian CNS either. There are innumerable homologous areas, cell types, neurotransmitters, peptides, chemical interactions, and so forth. However, there are important differences beyond these surface resemblances.

For example, consider the semicircular canal of the ear. All mammals have roughly the same five end organs in their ears to support their auditory and vestibular systems, which all work to keep the lateral semicircular canals parallel to the horizontal plane relative to the Earth; keeping it in that position allows mammals to get the best possible information about head position in space. (The lateral canal is maximally excitatory to a yaw [left to right] head motion; keeping the canal in line with the horizontal plane allows the organ to detect this motion with the greatest accuracy.) But rodents ambulate with their necks extended, which keeps their heads in an extreme dorsal position, while humans incline their heads about twenty degrees when walking naturally. In general, the differences in the shape of the semicircular canals in the ear with skull shape are correlated with the position that an animal's head is normally in.

For another example, consider the retina. There are striking differences between herbivores and predators in brain structure, for creatures who munch on grasses and trees require much less precise environmental information than those who hunt moving targets in order to survive. As a result, rodents have no foveae. To maintain visual fixation on a point, they move their necks, using what is known as the vestibular-colic response. The vestibular system in their ears tells them how their head is oriented and they use that information to reorient their heads in order to keep whatever object currently fascinates them in their line of sight.

In contrast, primates have foveae and move their eyeballs to keep their target within the foveal area, using the vestibular-ocular response. This is a much more precise orienting mechanism, which allows them to move their eyes to compensate for changes in head position such that they can keep objects foveated for as long as they wish. For some indication of how important computing horizontal eye motion is to primate brains, consider that the abducens (or sixth) nerve in humans, which controls horizontal eye abduction, feeds into one of the biggest motor nuclei in the brain stem. This ocular nucleus, which controls only one very tiny muscle, is only slightly smaller than the nucleus that controls all of the twenty or so facial muscles.

In more striking contrast still, bats do not maintain ocular position in the same fashion as the rest of the mammals. Because they fly and so have greater freedom to move in three-dimensional space, maintaining body position relative to the horizontal is not an easy option. As a result, they use other sense organs, primarily hearing (the other half of the eighth nerve), to determine how their eyes should be oriented. Consequently, they need not rely on vestibular-ocular responses, as primates do, even though their bodies are equipped with such reflex machinery.

All of these anatomical and physiological differences are important when neuroscientists want to investigate something, like how the brain learns to compensate for damage to the vestibular pathways. What may seem as small and insignificant differences from a broad mammalian perspective becomes hugely important as scientists seek to understand the particular mechanisms of brain plasticity. Can they use animals with no foveae and a vestibular-colic response to learn about how foveated mammals recover their vestibular-ocular response? More generally, how well do particular animal models translate across the animal kingdom? Should scientists be allowed to generalize from experiments on a single species (or set of species) to how nature functions?

In all vertebrates, a unilateral labyrinthectomy (UL), or a lesion of the labyrinthine structure in one ear, gives rise to two types of ocular motor disorders. There are static deficits, such as a bias toward looking toward the lesioned side when the head is not moving, and dynamic deficits, such as abnormal vestibular-ocular reflexes, which occur in response to head movements. In only two or three days following the UL procedure, the brain starts to compensate for its loss and the static deficits disappear. Since labyrinthine structures do not regenerate, and peripheral neurons continue to fire abnormally, whatever the brain is doing to recover has to be a central effect. Single-neuron recordings from a variety of animals indicate that the vestibular nuclei on the same side of the brain as the lesion start to show normal resting rate activity as the brain learns to compensate for its injury. Scientists do believe that whatever the mechanism is, it is also likely to be a general procedure the brain uses for recovery, for there are similar resting rate recoveries of the sort seen with the ipsilateral vestibular nuclei following denervation in the lateral cuneate nucleus, the trigeminal nucleus, and the dorsal horn, among other areas. Exactly how an argument to defend these convictions is supposed to run, though, is unclear, since it is fairly easy to find significant differences in how organisms recover and compensate for vestibular damage across the animal kingdom.

Frogs, for example, appear to rely on input from the intact labyrinth to regulate the resting activity of the vestibular nuclei. Mammals, however, do not. The recovery of their vestibular nuclei occurs independent of transcommissural inputs. In addition, static symptoms follow different time courses in different animals. In rats, spontaneous nystagmus disappears within hours after UL, while in the rabbit and guinea pig, it persists for several weeks. In humans, it may continue in one form or another for several years. There is a fundamental tension in neuroscience between the big picture and what is found in particular instances. All sciences strip away features of the real world when they devise their generalizations. Physicists neglect friction; economists neglect altruism; chemists neglect impurities, and so on. However, what neuroscientists are doing is not analogous to what the physicists, economists, and chemists are doing. In each of the other cases, the scientists are simplifying the number of parameters they must consider in order to make useful and usable generalizations. In contrast, if neuroscientists were to ignore the differences they find across species, then they would have no data left to build a theory with. There is nothing left over, as it were, once neuroscientists neglect the anatomical and physiological differences found in the brain across the animal kingdom. There is much left over when physicists neglect friction—most of classical mechanics is left, in fact. In contradistinction to the other sciences, there is a tension in neuroscience between the general rules one hopes to find that describe all brains and the particular cases neuroscientists happen to study.

What should the scope and degree of generalization for neuroscientific theories be? It is an unpleasant choice. Either scientists settle for large-scale abstract generalizations, which gloss over what may be important differences, or they focus on the differences themselves, at the expense of what may be useful generalizations. However, despite appearances, it is not an either–or proposition that has to be resolved before scientists can move ahead, for a proper neuroscientific theory contains both general (and fairly vague) abstractions, as well as detailed comments on specific anatomies and physiologies. The paradigmatic theories of physics are simple elegant equations with universal scope. Theories in neuroscience read more like a list of general principles plus detailed commentaries. One feels the tug of the dilemma posed above only if one is operating with a restricted notion of what a scientific theory is. Some theories are pithy and succinct; some are not. Neuroscientific theories are not.

In neuroscience, what scientists start with is a theoretical description at the most general level; it might be called the "theoretical framework"—the most general component in a neuroscientific theory. Once they adopt the framework, they can make more precise hypotheses as a way of filling out their theoretical proposal. These claims can be local to particular phyla or species; hence, they are not intended as a more detailed specification of the general framework. Instead, they can be thought of as instances or examples of how the framework might be cashed out in particular cases.

However, it is not the case that all "fillings out" fail to generalize. For example, the dynamic symptoms of unilateral labyrinthectomy recover using a different mechanism (probably). One hypothesis is that brains use a form of sensory substitution to compensate for the vestibular-ocular reflex. In this case, the brain uses internally generated signals from the visual or somatosensory systems to compensate for the vestibular loss. It may substitute computations from the saccadic or a visual pursuit system, both of which (probably) reconstruct head velocity internally, for vestibular throughputs. Data drawn from experiments on frogs, cats, and humans indicate that they all apparently use the same mechanism, though it remains to be seen whether this proposal will be applicable to all creatures and can be generalized much beyond vestibular reflexes.

There are different degrees of abstraction one might use once some theoretical framework is adopted. Some discussions are going to be restricted to a single species, or maybe even one developmental stage within a species; others will include several unrelated species or phyla. Both are legitimate ways of cashing out the framework in particular instances, and neither is to be preferred over the other. The data will dictate the scope of subhypotheses, and scope can vary dramatically.

And this is how theories in neuroscience are built and structured. Detailed conclusions regarding a single animal model give rise to general theoretical principles. These principles inspire new experiments done with other animal models, which in turn provide new (and probably incompatible) details but also new general principles. These new principles then connect to other detailed studies using different protocols on still other animals, and so on it goes.

At the end of the day, there is a set of related theoretical principles that jointly compose a general theoretical framework. These principles are held together by the detailed data coming out of a wide variety of animal studies. Neuroscience continually moves between two different ways of understanding the nervous system, first in broad and sweeping strokes and second by being submerged in the minutiae. General theoretical principles arise out of and then feed back into particular animal experiments done on different animal models. Because physiology differs across species, specific experimental protocols are appropriate only for specific models. Sometimes the data arising out of the different animal models or different experimental procedures overlap, but largely they do not. Hence, sometimes the detailed conclusions are consistent, but sometimes (a lot of times) they are not. Neuroscientists weave a story through their animal models and experimental protocols united by a common guiding theoretical thread. They both find commonalties and define differences. This entire exercise, taken together, fashions the theoretical structure of neuroscience.

Representation in the Brain

Important to keep in mind is that brains are evolved products, formed to help animals feed, flee, fight, and reproduce. This contrasts with digital computers, say, which are designed to compute. Early cognitive science generally took minds to be importantly analogous to computers and so tried to build theories of representation that spanned what both computers and humans did (see Cognitive Science). Contemporary neuroscientists think of representation in terms of brains only.

Most operate implicitly under the assumption that individual neurons are the representational engine that drives brains. This assumption is not universally accepted in neuroscience, but it does provide a good starting point nonetheless. As indicated above, the difficulty with looking at the behavior of individual neurons is that it rapidly becomes extremely complicated.

Hence, many working in the area of brain representation carry out their research looking at artificial neural nets, which have the advantage of being much simpler and easier to control than what nature has provided, so it is much easier to design and carry out experiments on them. But this very usefulness becomes a disadvantage because scientists cannot be sure that the behaviors they get out of them are relevantly similar to what is seen in the brain. What representations look like in them are multidimensional phase spaces, which is quite different from what traditional philosophers envision when they speak about representations. Whether a phase-space approach will replace traditional approaches to representation in philosophy remains to be seen.

Part and parcel of the problem of understanding representations is understanding learning, or how creatures get representations in the first place. Brain organization is remarkably constant across structures. Cortex is cortex is cortex. Therefore, most specialized areas have to be carved out via experience. The postulated mechanism for producing such changes is nothing more than Hebbian learning, a mechanism first articulated in the 1940s: Repeated activation will cause future activation to be easy; decreased activation will make future responses more difficult (Hebb 1949).

But to get directed learning, one must combine a learning mechanism with some sort of reward system. Animals need a reason to repeat an event in order to learn about it. Things that feel good they repeat. Things that do not they avoid. Much is now known about reward networks, especially those involving fear conditioning in rats. These give scientists some clues about how all sorts of reward-based learning might be going on in brains, though whatever the final story is, it will be unquestionably more complicated than what is known now.

Neuroethics

As progress is made into understanding how the brain works and how to influence brain functioning, serious ethical questions arise concerning how leaders in such fields as medicine, government, and insurance should react to new information and possibilities (Marcus 2004). This is a newly burgeoning area of research, with national attention only now being focused on the issues. Particular questions that philosophers of neuroscience will have to answer concern how and whether scientists should alter normal functioning brains, how and whether scientists should use brain technology to track individuals' social behavior, and how and whether what is learned about the brain changes how philosophers think of being human.

Scientists know a lot about how memory works and, more importantly, how it fails. Ways in which memory can be imperfect include:

1. Decreased accessibility to memories over time,
2. Lapses in attention,
3. Temporary inability to access stored information,
4. False recognition of things,
5. False remembrance of things,
6. Contamination of stored information by current beliefs, and
7. Recollection of items at inappropriate times.

All of these processes are perfectly normal and occur in everyone at some time or another. Suppose there is some way of correcting some or all of these deficits? Should doctors do it? Or should they accept less-than-perfect memories as the way humans are?

Neuroscientists are already tracking where and how moral decisions are made in the brain; they are also looking at brain differences between normal and sociopathic, psychopathic, and violently impulsive individuals, who respond to violent or otherwise disturbing situations with increased activity in the amygdala and decreased activity in the frontal lobes relative to normal. Scientists can now identify such trends in individuals before they actually commit any crime. Should they? And what should they do with such information once they have it?

If it is concluded that violence is biologically based, as are all other behavioral decisions, then what does this say about notions of self or free will? How might this alter the court systems, since they operate under the assumption that one is guilty if one could have done otherwise in a situation but chose not to? Similar questions arise regarding gender differences in the brains. Female brains differ from males'. What effect, if any, should this fact have on educational systems, social expectations of gendered behavior, or men's and women's professional lives?

Philosophers and neuroscientists are only beginning to confront these sorts of questions, and technology is only beginning to allow scientists to understand and change the brain to any significant degree. The questions philosophers need to confront in the next decade will differ greatly from these, as knowledge of the mind/brain continues to increase exponentially.

VALERIE GRAY HARDCASTLE

References

Barinaga, M. (1995), "Remapping the Motor Cortex," *Science* 268: 1696–1698.

Bechtel, W. (2000), "From Imaging to Believing: Epistemic Issues in Generating Biological Data," in R. Creath and J. Maienschein (eds.), *Epistemology and Biology*. Cambridge, MA: Cambridge University Press, 138–163.

Bechtel, W., P. Mandik, J. Mundale, and R. Stufflebeam (eds.) (2001), *Philosophy and the Neurosciences: A Reader*. New York: Blackwell Publishing.

Cabeza, R., and L. Nyberg (1997), "Imaging Cognition: An Empirical Review of PET Studies with Normal Subjects," *Journal of Cognitive Neuroscience* 9: 1–26.

——— (2000), "Imaging Cognition II: An Empirical Review of 275 PET and fMRI Studies," *Journal of Cognitive Neuroscience* 12: 1–47.

Connors, B.W., and M. J. Gutnick (1990), "Intrinsic Firing Patterns of Diverse Neocortical Neurons," *Trends in Neuroscience* 13: 99–104.

De Gelder, B. (2000), "More to Seeing Than Meets the Eye," *Science* 289: 1148–1149.

De Yoe, E. A., and D. C. Van Essen (1988), "Concurrent Processing Streams in Monkey Visual Cortex," *Trends in Neuroscience* 11: 219–226.

Evarts, E. V. (1968), "A Technique for Recording Activity of Subcortical Neurons in Moving Animals," *Electroencephalography and Clinical Neurology* 24: 83–86.

Hebb, D. O. (1949), *The Organization of Behavior*. New York: Wiley.

Lewicki, M. S. (1998), "A Review of Methods for Sorting: The Detection and Classification of Neural Action Potentials," *Network: Computational Neural Systems* 9: R53–R78.

Marcus, S. J. (ed.) (2004), *Neuroethics: Conference Proceedings*. New York: Dana Press.

Merzenich, M. M., J. H. Kaas, M. Sur, R. J. Nelson, and D. J. Felleman (1983), "Progression of Change Following Median Nerve Section in the Cortical Representation of the Hand Areas 3b and 1 in Adult Owl and Squirrel Monkeys," *Neuroscience* 10: 639–665.

Mishkin, M., L. G. Ungerleider, and K. A. Macko (1983), "Object Vision and Spatial Vision: Two Cortical Pathways," *Trends in Neuroscience* 6: 414–417.

Schaffner, K. F. (1993), *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.

Suppe, F. (1989), *The Semantic Conception of Theories and Scientific Realism*. Chicago: University of Illinois Press.

Woodward, J. (1989), "Data and Phenomena," *Synthese* 79: 393–472.

See also **Biology, Philosophy of; Cognitive Science; Explanation; Reductionism**

NEUTRALIST/SELECTIONIST DEBATE

See **Evolution; Fitness; Natural Selection**

O

OBSERVATION

Empiricism is the doctrine that all our ideas are based on observation. In his classic statement of the doctrine, Hume (1967) maintained that ideas are built from sense impressions by habits of association based on constant spatiotemporal conjunctions. Twentieth-century logical empiricists criticized Hume as having focused too narrowly on psychological association as the way ideas "come from" experience. For them, the problem was to provide a logical rather than a psychological construction of ideas. Their interest in science led them to emphasize this concern by speaking of "observation" rather than "impressions." Their approach through language led them to speak of "observation terms" rather than "observation." These terms were to be distinguished from the terms based on them, which came to be called "theoretical."

A major attempt to carry out this modern version of the empiricist program is well illustrated in the career of Rudolf Carnap, arguably the most influential proponent of what came to be called logical empiricism. In his first major work, *Der logische Aufbau der Welt*, written in 1922–1925, Carnap (1967) attempted to develop a "constructional system" that would be a step-by-step derivation, or construction, of all concepts from "a few fundamental concepts" (1). Such a construction could, he argued, be achieved by building on either a physical or a psychological basis. The advantage of a physical approach would be that "from the standpoint of empirical science the constructional system with a physical basis constitutes a more appropriate arrangement of concepts than any other" (95). But Carnap chose a psychological approach because of his intention "to have the constructional system reflect not only the logical-constructional order of the objects, but also their epistemic order" (101). He called this basis "autopsychological" or "solipsistic," declaring that it "could also be described as *the given*" (102). In later discussions, the basis on which meaningful concepts could be built (from which they could be derived by logical construction) was widely spoken of as "the given."

It is important to note that the basic elements on which the rest of our concepts are built—what later would be called "observation"—were taken implicitly as synonymous with the deliverances of the senses or some subset thereof. Thus Carnap says, "The basic elements, that is, the experiences of the self as units . . . we call *elementary experiences*" (108). Hempel, writing later, implicitly made the same point: In regard to an observational term, it is possible, under suitable circumstances, to decide by means of direct observation whether the term does or does not apply to a given situation.

Observation may here be construed so broadly as to include not only perception, but also sensation and introspection; or it may be limited to the perception of what in principle is publicly ascertainable, that is, perceivable also by others (Hempel 1958, 42). Hempel's broadest conception of observation was still limited to "perception" and "sensation and introspection."

Defense of this empiricist doctrine required dealing with three major tasks.

1. *Arriving at an appropriate interpretation of what is to be considered a sense perception*. That is, classical empiricism had to make clear exactly how what is purely given in sense perception, free of interpretation, is to be distinguished from other ideas that are supposed to be based thereon. These other ideas include all concepts and beliefs that go beyond the purely given in sense perception, *viz*., the theoretical ones.
2. *Showing in what sense, and precisely how, those other ideas are based on sense perceptions*. Here we must distinguish between two kinds of ideas: concepts (terms) and propositions (statements having truth values). The focus of this article is on concepts, particularly that of observation; propositions and their testing are relevant only insofar as beliefs established as putative knowledge will be found to shape the formulation of the concept of observation. As to the testing and establishment of propositions, there were two distinct schools of empiricism. The first, often taken to be represented by Bacon, Hume, and Mill, held that our propositional beliefs are based on sense perception by being derived from (or constructed out of or otherwise induced from) sense perceptions. (In the case of Hume, this is a logical reconstruction of his psychologistic view.) The second view of propositions was the hypothetico-deductive (H-D) view, taken by most logical empiricists, according to which our propositional beliefs are based on sense perception in that they must be tested in terms of their implications regarding sense perceptions. Thus, each school had its task laid out. Traditional empiricists would have to show the exact way or ways in which the derivation or induction is made in accordance with specifiable rules whose origin and justification can be understood. Adherents of the H-D view would have to say something about how hypotheses are obtained and show precisely how those hypotheses are to be tested.
3. *Clarifying whether or not ideas are to be counted as knowledge, and why*. That is, the task was to show, first, whether some beliefs legitimately count as knowledge, as opposed to others that do not; and second, if so, why they count as knowledge. The latter task is generally taken to entail showing whether putative knowledge is about a world existing independently of sense perception or is only a coherent or usable way, perhaps one of many, of organizing or relating items of sense perception.

Failures of the Empiricist Program

It is widely admitted today that empiricism, logical or traditional, failed to deal successfully with its three basic tasks.

As to the first task, no clear specification was ever provided of what is to count as a sense perception (or linguistic observation term, as logical empiricists preferred to say), as distinguished from what is based thereon. To guarantee objectivity, the basis would have to be completely free from all interpretation and presupposition whatever: Interpretations should be constructions, conclusions from our perceptions, not assumptions buried within and potentially biasing them. But such a pristine, raw, brute "given" was never found, and the search for it is now generally recognized to have been an utter failure.

This failure carried over to afflict treatments of the second issue: If it remained unclear what was supposed to be the basis of all our other ideas, it also remained unclear whether all our beliefs might be infected by presupposed interpretations that would destroy any claims we might make to having knowledge. But even when a particular vocabulary was taken for granted as the basis (for example, as consisting of unanalyzable sense data, or as a language of objects), it was not clear how the nonbasic terms—the theoretical ones—were supposed to be based on the given. Even in *Der logische Aufbau*, Carnap realized the impossibility of giving all theoretical terms explicit definitions formulated by means of the basis; some terms could be understood only by "definitions in use" (65–67). In *Testability and Meaning*, first published in 1936–1937, Carnap (1954) developed this idea in terms of "reduction sentences," which were implicit rather than explicit definitions of theoretical terms. In *The Methodological Character of Theoretical Terms* (Carnap 1956), he introduced yet another view of the meanings of theoretical terms, as partially dependent on the meanings of other theoretical terms within a theory.

Responses to the third issue were equally inadequate. For example, the acceptability or rejectability of beliefs appeared not to be understandable in terms of pure, given sense perceptions. That issue also faced further objections, independent of the failures with the first two. Usually on technical grounds, no attempt to develop an inductive logic achieved consensus among philosophers, whether it attempted to show how beliefs could be obtained directly from such sense perceptions or how they could be tested (confirmed or disconfirmed) by sense perceptions.

The development of Carnap's thinking, as outlined above, encapsulates the broader development of empiricism during the twentieth century, and of the concept of observation in particular. Four decades of intense critical analysis showed that, at least in any sense compatible with the programs of traditional and logical empiricism, the attempt to clarify the "logical structure of the world" had failed.

This summary provides a general characterization of attempts to deal with these issues; however, it will not discuss further attempts, or the criticisms they provoked. Discussion of them is omitted here not simply because these arguments and criticisms are by now familiar, or because it is so widely accepted that they show the indefensibility of empiricism, in either of its two versions, but because they do not confront the really fundamental difficulties of the empiricist program. They show only that all efforts to provide an adequate version of the doctrine have failed *so far*, but they offer no indication whether that quest might still succeed, or whether it is wrong in principle. Though they show *where* classical empiricism went wrong, they fail to show *why* it did so—why, for example, its apparently genuine motivation, to avoid any subtle infusion of bias into inquiry, to let nature speak for itself, was misguided (if indeed it was), or why the effort to show how knowledge is based on sense perception failed. Further, the criticisms offer no guidance as to whether or how classical empiricism might be defended, or, if it could be, then in what directions it requires alteration. For such reasons, putting those criticisms at the center of the diagnosis of what was wrong has meant that in recent times, inquiry into the basis and implications of science has wandered without clear direction.

We need to gain more positive insight into two aspects of the failures: first, the fundamental reasons for them, thus answering the question of whether the program was rightly conceived but difficult, or wrong in principle; and second, the positive directions in which a more adequate view of science might be sought. A first step toward such positive insight can be taken by examining the following question: Even if the program of classical empiricism could be successfully fulfilled, what would have been omitted from our understanding of science? This question can be answered through close examination of a case of scientific research that exposes central aspects of the nature and role of observation in sophisticated modern science. Those aspects lead us to a view of observation, and of science generally and how to understand it, that is very different from that of traditional and logical empiricism and reveal far more fundamental reasons for the failures of those views than have been discussed above.

An Alternative View of Observation: Example and Formulation

The question of the source of energy that makes the stars shine arose in the nineteenth century from thermodynamics. The sun was known to be a typical star; its total energy output could be calculated from the amount received on Earth. None of the proposed sources of this energy (e.g., gravitational contraction, chemical burning, meteoritic impact, radioactive decay of heavy elements) were adequate to produce that amount. It was not until the advent of nuclear physics in the early 1930s that a promising theory of stellar energy production could be advanced. Hans Bethe and Carl von Weizsäcker proposed a detailed series of nuclear reactions taking place in the center of the sun at the high temperatures and pressures expected to exist there; the series begins with hydrogen and ends with the production of helium with the release of energy. Once produced, the energy would work its way to the solar surface and from there be emitted, some of it arriving on Earth.

By the time the energy reaches the surface of the sun, it has been degraded and transformed by interactions along the way and is released as electromagnetic energy, through electromagnetic interactions that are very different from the postulated nuclear processes in the deep interior. Thus neither the information obtainable through observing the solar surface nor the inferences made on the basis of them could give direct information about what went on in the central core. How then could the theory be tested in a more direct way?

The possibility of such a direct test lay in another major area of physics in the twentieth century, which developed the theory of weak interactions. A by-product of the Bethe–von Weizsäcker nuclear processes would be the production of neutrinos, or weakly interacting particles. Unlike the electromagnetic photons degraded by interactions while passing through the sun, these weakly interacting neutrinos would travel through it with extremely low probability of interacting with a single particle of the sun's bulk. They would pass directly to Earth (in the clear sense of not being interfered with and altered on the way) exactly as they were when created in the solar core. Of the enormous number of neutrinos that reach the Earth, most would pass right through it, just as they passed through the sun; but a very tiny number would interact with an appropriate target substance (chlorine in a container the size of a railroad car filled with ordinary cleaning fluid) to produce radioactive argon, with a half-life of roughly 35 days. Despite the infinitesimal number of argon atoms produced, it was shown that these atoms could be extracted from the tank before decaying. Their decays could be monitored and counted, to reveal the number of occurrences of neutrino captures, that is, if account was taken of other possible processes mimicking the production of radioactive argon. The total number of neutrinos produced in the center of the sun by the Bethe–von Weizsäcker processes could then be calculated. That number would constitute *unaltered* information about the nuclear reactions in the solar core and therefore would provide a direct test of the nuclear theory of solar (and stellar) energy production.

Though much of the reasoning involved must be omitted here, a few of the complexities of this test are evident even from this brief summary:

- the body of failed theories of stellar energy production;
- the new theories entering into the conception of the experiment (and their variety, including not only the electromagnetic, weak, and nuclear forces and their modern quantum-theoretic background but also the knowledge that the sun is a star);
- the feasibility of the experiment (e.g., What would be a good target substance, with a product having a convenient half-life? Could the "neutrino telescope" be constructed? Could the product, argon, be removed efficiently from the enormous volume of liquid? How could mimicking processes be guarded against?); and
- the interpretation of the results (What would be the implications of the numbers of neutrinos found?).

But there are dozens of other pieces of what can be called "background information" for the conception, execution, and interpretation of the experiment: Where should it be built? How can one clean the tank so as not to contaminate it? and so on. (See Shapere 1982 for further details about the experiment and its historical roots, and also the theory of observation discussed below.) Without a great deal of background information, not all of it plausibly classified as theoretical, the experiment would have been, in the most literal sense, inconceivable; without a great deal of background information (often different from that involved in conceiving the experiment), it would not have been executable; and without still other background information, its results would not be interpretable.

With only minor qualifications (Shapere 1982), scientists refer to this experiment as an observation (or an experiment set up to make an observation) of what goes on in the center of the sun, made in order to test the theory of stellar energy production in a direct way, which according to physics, cannot be done by relying on electromagnetic information received from the sun's surface. Calling this an observation is fundamentally at odds with the concept of observation in traditional and logical empiricism. But the usage has justification, partly as follows.

Empiricism identified two distinguishable aspects of observation: evidential and perceptual. It thus maintained that the problem of *what counts as evidence* is identical with the problem of *what is given in sense perception*. Further, it held that this given was free of any interpretation whatever (including background information), so that to explain observation was to explain what is given in experience, and simultaneously to explain what counts as evidence in testing a hypothesis or theory. The attempt to fill this out failed; but the point is that, for understanding science, *the attempt was misguided from the start*. In sophisticated cases of observation in modern science, the evidential and perceptual functions are separated, and the focus is on the former. There are good reasons for this. In its concern with testing, the focus of modern science is on observation as evidence, not on observation as perception. Correspondingly, sense perception is limited and unreliable in important ways; the perceptual aspect appears as an interference with the evidential. Therefore, for the purposes of scientific observation, the human eye, which is subject to error, is replaced by detectors that are not. True, to be useful for scientific purposes, the data must be finally put into a form accessible to human senses, but that is only because we are the ones who use the evidence. The real work of collecting evidence, the scientifically relevant aspects of observation, has already been done independently of the senses. The role left to sense perception is minimal; what is crucial is possession of the background information required to interpret the results.

The analysis of this experiment suggests a new view of observation in science, summarized as follows.

Statement *O*: An entity *X* is said to be *observable* if (1) information can be received by an appropriate receptor and (2) that information can be transmitted directly, that is, without interference, to the receptor from the entity *X* (which is the source of the information). Correlatively, something can be said to be *observed* if it meets these two conditions, with "is" substituted for "can be." This account of observation is heavily dependent on the accepted content of the science (the background information) of a particular era, particularly regarding what is counted as an appropriate receptor, as information, as the types of information there are, and as interference. (Indeed, although this cannot be shown here, Statement *O* in its entirety depends on the way things are.) Each of these is specified in light of the background information that makes possible the conception, execution, and interpretation of the experiment.

This dependence on background information might suggest that the present view resembles the views of such writers as Norwood Russell Hanson, Paul Feyerabend, and Thomas Kuhn. In contrast to the empiricist tradition, those authors argued that observation is "theory-laden" (Hanson 1958), or that "the meaning of observation sentences is determined by the theories with which they are connected" (Feyerabend 1965, 213) or that "no experiment can be conceived without some sort of theory" (Kuhn 1970, 87). However, the view outlined in the present article differs from those in a number of fundamental ways.

For one thing, in contrast to Kuhn, the view of this article is that there is not simply one basic idea (paradigm) or small set of ideas (elements of a disciplinary matrix) operating in all scientific problem situations. As science develops, a pool of background information also develops, and some items from that pool may be used in one problem stuation (such as an experiment) and different ones used in another.

Second, whereas Feyerabend frequently claimed (e.g., 1978) that in science "anything goes," this article holds that the items used in specific research must be specifically relevant to the situation at hand.

Third, unlike both Feyerabend and Kuhn, this article maintains that the background information used in an observation or experiment (as well as in other aspects of scientific research) is not always a theory (much less a paradigm or a disciplinary matrix). Many separate items of background information are brought to bear in particular problem situations; some of those items are naturally called theories, but others, such as how to clean the tank in the solar neutrino experiment, are not. Furthermore, many different theories enter in as background information (nuclear physics, electromagnetic theory, weak interaction theory, etc.), and so do many nontheoretical items.

Finally, as regards Hanson: Dealing with particular problems of observation, experiment, and theoretical research does require background information; but being "laden" with the sorts of background information described in this article does not violate objectivity. (Hanson himself did not say that it does, but his view has often been taken to imply this). Only beliefs that have been established are eligible to serve as background information. (This point holds against Feyerabend also: It is not the case that anything goes.) How this establishment comes about, however, cannot be discussed in the brief space available here.

Fundamental Weaknesses of the Classical Empiricist Program

We now return to the question of whether the empiricist program failed merely because of its difficulty or because it was fundamentally misguided. The points to be made are these:

1. Even if all our knowledge were shown to be ultimately based (in whatever sense) on sense perception, the nature of knowledge and of knowledge seeking would still not be understood by focusing only on that ultimate. That is, even if it were really the ultimate, much more would still need to be understood.
2. Furthermore, not only is the classical empiricist approach insufficient to understand science; its methods and problems are largely irrelevant and are even a hindrance.

In the sixteenth through eighteenth centuries, when classical empiricism was first presented systematically, it was reasonable to try to consider scientific concepts as being exhaustively interpretable in terms of sense perception, at least under some views of that term. Even though it was initially plausible, however, such a program has become more and more difficult to maintain as science has developed. If one were not too strict about adherence to the directness of sense perception, some problematic cases (e.g., the concepts of force, inertia, and space; forms of electromagnetic radiation beyond the visible) might still be assimilated within the program. But in an increasing proportion of cases in the nineteenth century and (especially) the twentieth, the selection of what was to be observed, how it was to be observed, its description, and the interpretation of the significance of the observation all grew increasingly distant from what was perceptible by the senses.

Again, the solar neutrino experiment provides an example of this departure. It purports to give direct observational evidence of processes occurring in the center of the sun, which is certainly not accessible to sense perception. Or if we say that it is neutrinos we observe, they are not accessible to sense perception either. The role of concepts such as these (e.g., force, neutrino) in science would still remain to be analyzed even if they were, as classical empiricists maintained, based on sense perceptions.

But those roles, and the reasoning involved in them, are crucial to science. The assumption of classical empiricism was that reflection on what could be constructed on the basis of sense perception was both necessary and sufficient for understanding what happens in scientific investigation. It is neither.

First, it is not sufficient, since no matter how sense perception is interpreted, its analysis (even as conceived in mature logical empiricism, with a theory of definition or logical construction) would still not make possible an understanding of the process of seeking and attaining knowledge, or the nature of knowledge, or the implications of the existence of knowledge, by considering solely the ultimate perceptual basis of knowledge or of claims to knowledge.

Second, in modern science, sense perception is only barely necessary, relegated to the periphery of inquiry as it is, and indeed as it must be, in some of the more sophisticated areas of modern science. As noted above, the role of sense perception was reduced dramatically during the nineteenth and twentieth centuries. Sense perception has been found to be too unreliable, in understandable ways, to depend on for the precision and the sophistication required in modern science. New instrumentation has been required to gather information that we have reasons to believe is available but is beyond the grasp of human sense perception. Sense perception still has a role to play in such investigations, but at least in the most sophisticated areas of science, that role now lies in the last stage of the inquiry, when the information received by the instruments (the receptors that replace our senses) is transformed into humanly accessible form. Otherwise, sense perception is relied on as little as possible in carrying out the important stages of obtaining the information and of interpreting it as information. Surely there is something paradoxical in saying that sense perception is fundamental to the scientific enterprise while at the same time it has come to be regarded as so peripheral to that enterprise.

Conclusions and Implications

Two conclusions must be drawn. First, *even if* we assume that a clear understanding of the idea of interpretation-free sense perception could be obtained, and *even if* it could be shown that all scientific concepts and propositions (and particularly those that we are justified in calling knowledge) are ultimately based on such sense perception, the most important reasoning in science will still have been neglected. For example, the central body of reasoning by which scientific ideas have been arrived at, especially in the twentieth century, would have been completely ignored.

Second, even assuming that classical empiricism were to obtain a clear interpretation of what is to count as pure sense perception, we would still be forced to recognize that this is only peripheral, not central, to understanding science and its processes. It is basically unsuitable for the investigations that modern science undertakes, and it appears only when the data achieved in an experiment have to be translated into humanly accessible form.

From the perspective of an attempt to understand modern science, the almost exclusive focus of empiricism on the role of sense perception in constructing or testing scientific ideas was a red herring. It called attention away from the centrally important aspects of scientific reasoning, which has led us to our present views of the universe and our place in it. It deflected attention away from the role of background information in scientific reasoning and activity and from the sources of the background ideas that are used, both specific ones and the general fact that their use almost defines what it is to be "sophisticated" modern science. Since that focus was a red herring, distracting attention from the truly important problems, criticism of classical empiricism on the ground that it never succeeded in specifying what counts as a sense perception, free of all interpretation or assumptions whatever, is largely beside the point. Such criticism is certainly correct, and its correctness certainly undermines the assumption, made above for the sake of argument, that the objection might have been surmountable. But the arguments just made show that even if it had been surmounted, classical empiricism would have given only the most peripheral and superficial insight into the workings of science. Thus not only is the objection profoundly unilluminating, its negative quality giving no direction to better understanding; it also distracts us from seeking the fuller and better understanding that is needed. Finally, and worse still, focus on this feature of the failure of classical empiricism has opened the way to abandoning any version of the project of empiricism—traditional, classical, or other—and, with that abandonment, giving up the claim that science can achieve, and indeed in many cases has achieved, knowledge.

The standard criticisms of classical empiricism are only symptoms of deeper problems. The classical empiricist program was misguided from the start by fundamental misunderstandings of what is involved in the process of seeking knowledge and in the nature of the knowledge sought. Its fundamental motivations were thus misguided in principle. In particular, the classical empiricist tradition took a wrong turn in insisting that background beliefs ought not to be appealed to in scientific inquiry—that the hazards of bias and error, so dreaded by classical empiricists, would result from the use of any sort of background beliefs in inquiry, whatever those background beliefs might be. All these ideas and worries were phantoms, products of basic misconceptions about the nature of inquiry and knowledge.

The worst part of the story is that even while logical empiricism is widely admitted to be defunct, its more potent phantoms persist. Its anthropomorphic concepts of observation and theory, together with its denial and consequent neglect of the centrally important roles of background information, have continued to haunt, dominate, and confuse discussion in the philosophy of science.

Some writers still defend a "constructive empiricism," which "requires theories only to give a true account of *what is observable*" (van Fraassen 1980, 3). Constructive empiricism does not require belief in the existence of unobservables; science requires only minimal empirical adequacy. The thrust of van Fraassen's arguments is directed against realistic treatments of theoretical purported entities. Limitations of space prevent full discussion of the implications of the present view for that topic here; however, one aspect is indicated by van Fraassen's (1980) remark that "[s]cience presents a picture of the world which is much richer in content than what the unaided eye discerns. But science itself teaches us also that it is richer than the unaided eye *can* discern" (59).

He distinguishes our being able to observe Jupiter through a telescope from our ability to observe particles moving in a cloud chamber (16–17); in the former case, astronauts will someday be able to get close to Jupiter and perceive it, whereas in the latter case, the inference from track to particle is based only on an analogy with seeing a jet trail in the sky and inferring the existence of a plane up there. As van Fraassen notes, this clearly distinguishes the perceptible from the nonperceptible. But is this relevant to the issue of realism regarding the particle? This entry has contended that the evidential, not the perceptual, is scientifically relevant. Scientists do talk—and for good reason, as was argued earlier—of observing particles in a cloud chamber just as they do of being able to observe the core of the sun by means of neutrinos. Thus if the question of observation is relevant to the question of realism, the latter must be resolved not by considering what is perceptible to the senses but by examining the evidential aspects of observation.

Despite all its criticisms of classical empiricism, the present view does not depart, at least in spirit, from that doctrine. It agrees with classical empiricism that *all our knowledge of the world (universe) is based on interactions with that world and its contents*. As the case of the solar neutrinos brings out, what counts as an interaction is determined by our knowledge and can differ in different fields, though there is always, at least in sophisticated science, a relationship of ancestry and descent (or cousinhood) between the different usages in different fields. We must learn, through investigation, what to count as an interaction. We also learn what to count as information, and as appropriate receptors, which may make use of more types of interaction than sense perception and may even exclude or severely limit appeal thereto. And we apply what we learn, as background information, in our subsequent inquiries.

In short, the view presented here is a *rational* descendant of classical empiricism. While still maintaining (on the basis of reasoned arguments) that we must interact with the world around us if we are to learn about it, this view nevertheless departs from classical empiricism in maintaining that we learn how to investigate nature by learning what it is to be an interaction, a piece of information, an appropriate detector, and much else. By building on what we have learned (i.e., by using background information) we learn how to learn, to think, and to talk about nature and how to assess what we have achieved when we have learned these things. This doctrine therefore emerges as a corrective to classical empiricism, a transition brought about in the development of science from an initially necessary reliance on sense perception, and the equating of sense perception with observation, to a distinction between the two and a focus on observation in the sense of evidence as the basis of inquiry and knowledge.

If we must have a name for the view, it can be called *interactional empiricism*, to distinguish it from the perceptual focus of classical, logical, and constructive empiricism, which together could equally well be called perceptual empiricism. Kosso (1989) calls his very similar view the "interaction-information account" of observation. This is fine, since information is an essential part of the account presented here.

But the important contrast with classical empiricism, captured in the name "interactional empiricism," lies in the idea of our interaction with nature. The present view is both a generalization of and a departure from classical empiricism. It is a generalization of that view in the sense that it involves considering sense perception as having to do with only one minor region of the electromagnetic spectrum, a manifestation of one of four fundamental forces (types of interaction) with nature. But it departs from classical empiricism in holding that we learn to go beyond the dictates of sense perception. In this double sense, of generalization and departure *on the basis of reasons*, the view presented here is a rational descendant of classical empiricism, just as the concept of Statement *O* is a rational descendant of sense perception.

DUDLEY SHAPERE

References

Carnap, R. (1954), *Testability and Meaning*. New Haven, CT: Whitlock's.

——— (1956), "The Methodological Character of Theoretical Concepts," in Herbert Feigl and Michael Scriven, *Minnesota Studies in the Philosophy of Science*, vol. 1: *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minneapolis: University of Minnesota Press, 38–76.

——— (1967), *The Logical Structure of the World and Pseudoproblems in Philosophy*. Translated by Rolf A. George. Berkeley and Los Angeles: University of California Press.

Feyerabend, Paul K. (1965), "Problems of Empiricism," in Robert Colodny (ed.), *Beyond the Edge of Certainty*. Englewook Cliffs, NJ: Prentice-Hall, 145–260.

——— (1978), *Against Method*. London: Verso.

Hanson, Norwood W. (1958), *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press.

Hempel, Carl (1958), "The Theoretician's Dilemma: A Study in the Logic of Theory Construction," in Herbert Feigl, Michael Scriven, and Grover Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 2: *Concepts, Theories, and the Mind-Body Problem*. Minneapolis: University of Minnesota Press, 37–98.

Hume, David (1967), *A Treatise of Human Nature*. Oxford: Clarendon.

Kosso, Peter. (1989), *Observability and Observation in Physical Science*. Dordrecht, Netherlands: Kluwer Academic.

Kuhn, Thomas S. (1970), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Shapere, Dudley (1982), "The Concept of Observation in Science and Philosophy," *Philosophy of Science* 49: 585–526.

van Fraassen, Bas C. (1980), *The Scientific Image*. Oxford: Clarendon.

See also **Carnap, Rudolf; Empiricism; Hempel, Carl Gustav: Hanson, Norwood W.; Induction, Problem of; Instrumentalism; Kuhn, Thomas; Logical Empiricism; Particle Physics; Phenomenalism; Realism; Scientific Domains**

OPERATIONALISM

See **Bridgman, Percy; Cognitive Significance; Verifiability**

P

PARADIGM

See **Kuhn, Thomas**

PARSIMONY

The principle of parsimony—or simplicity (treated here as an equivalent concept)—is also known as Occam's (or Ockham's) razor, after William of Occam, the medieval philosopher who said that plurality is not to be assumed without necessity and that what can be done with fewer assumptions is done in vain with more (Wood 1996, 20–22).

Scientists and philosophers often claim that the *parsimony* of a theory is relevant to deciding whether the theory is true or approximately true or would make accurate predictions. How this can be so is a central puzzle in the epistemology of science. It is not puzzling that people find parsimonious theories aesthetically attractive and easy to understand and manipulate. What requires elucidation is not the pragmatic value but the *epistemic* value of parsimony.

Just as people are said to be parsimonious when they are abstemious in how they spend money, a theory is parsimonious when it is tightfisted with respect to the entities, processes, or events it postulates. There is no cutoff separating theories that are parsimonious from theories that are not; rather, the difference is a matter of degree. The fundamental idea is comparative: One theory is more parsimonious than another. For example, if one theory postulates causes A and B to explain an observed effect E, while a second theory postulates only cause A and does not mention B, the latter theory is more parsimonious.

One epistemically significant feature of this difference is that if *A* and *B* are mutually independent, then according to probability theory the conjunction $(A \wedge B)$ will be less probable than *A*. Does this mean that parsimony and probability always coincide? In what follows, it will be seen that a number of philosophers have strenuously denied this. And even in the case at hand, there is reason to be careful about the suggestion. The second theory is "agnostic" about the relevance of *B*. But now consider a third theory, which asserts that *A* is a cause of *E* and denies that *B* is a cause. This third theory is "atheistic" about *B* and is more parsimonious than the first theory. However, probability theory does not say that $(A \wedge \neg B)$—that is, $(A \wedge \neg B)$—is more probable than $(A \wedge B)$. The hypothesis that there is at least one cause of *E* is more probable than the hypothesis that there are at least two causes, but there is no a priori reason to think that exactly one cause is more probable than exactly two. Parsimony has an obvious link with probability when a logically stronger hypothesis is compared with a hypothesis that is simpler and logically weaker; however, when two theories are mutually incompatible, the connection is anything but obvious.

The giants of the Scientific Revolution frequently referred to the importance of parsimony and its cognates. In *De revolutionibus orbium caelestium*, Copernicus emphasizes that his heliocentric theory differs from Ptolemy's geocentric theory in that the Ptolemaic system requires an independent model for the motion of each planet, whereas the Copernican system unifies the models for the different planets by including a common Earth/sun component in each. Copernicus remarks that his approach "follow[s] Nature, who producing nothing vain or superfluous often prefers to endow one cause with many effects" (Kuhn 1957, 176–179). Newton ([1686] 1953), in *Principia mathematica*, states as his first rule of reasoning in philosophy that

> we are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances. To this purpose the philosophers say that Nature does nothing in vain, and more is in vain when less will serve; for Nature is pleased with simplicity and affects not the pomp of superfluous causes. (3)

Leibniz ([1686] 1973, 11) defended parsimony as a criterion in scientific reasoning by appeal to his doctrine that God created the best of all possible worlds; our world is perfect because it is "at the same time the simplest in hypotheses and the richest in phenomena." For all these thinkers, the methodological principle rests on an ontological foundation. The principle of parsimony should be used in reasoning because nature is simple, and nature is simple because God made it so.

With the falling away of divine design as an acceptable justification of methodological principles, a fissure appeared in the foundation of scientific inference. If the principle of parsimony cannot be justified by tracing it back to a parsimonious creator, what could its justification be? Does the justification of the principle require any substantive assumptions about the natural world? Or is the principle just part and parcel of what it means to be "rational," which we are required to be no matter what the world is like? If the theological account is the thesis, its antithesis is the idea that the principle of parsimony is purely methodological. In between these two extremes, there is much room for accounts that reject both.

Local Versus Global Accounts

Most attempts to explain the epistemic relevance of parsimony treat the problem globally. They assume that if parsimony is epistemically relevant across a range of problems involving inference, the reason for its relevance must always be the same. However, it is worth pondering the possibility that the justification for using a principle of parsimony may vary from problem to problem. Perhaps parsimony needs to be understood not globally but locally (Sober 1990).

As an example, consider the long-standing use of parsimony as a criterion for inferring phylogenetic relationships in evolutionary biology (Sober 1988). Given a set of observed similarities and differences that characterize a set of species, how are these data to be used to figure out which species are closely related and which are related more distantly? A standard procedure is to find the phylogenetic tree that requires the smallest number of changes in "character state" to explain the data. This methodology assumes that the species are genealogically related and proceeds to identify the most parsimonious hypothesis concerning what the pattern of relatedness is. However, there is a prior question about phylogeny: Why think that the observed species have any common ancestors? Perhaps each species can be traced back to a separate origin.

The role of parsimony in answering this question can be understood by examining Crick's (1968) argument that the near universality of a single genetic code among the organisms now on Earth is evidence that they are all genealogically related. Crick says that the shared genetic code is arbitrary—one among a large number of viable mappings of nucleotide triplets onto amino acids. However, once an organism uses a given code, its fitness is likely to be compromised if it or its descendants modify the code already in place. Stabilizing selection then makes it highly probable that descendants will use the same genetic code as their ancestors. These biological assumptions (which Crick summarizes in the phrase "frozen accident") entail that the universality of the code would be very surprising if the organisms now on Earth were not genealogically related (e.g., were products of 27 separate startups) but is precisely what one should expect if all life can be traced back to a single progenitor. Because of this difference, Crick concludes that the observed universality strongly favors one hypothesis over the other. Notice that Crick's argument compares the likelihood of two hypotheses:

> P (the code is now universal | all current life traces back to a single progenitor) $>$ P (the code is now universal | current life traces back to 27 original progenitors and no fewer).

Here 'likelihood' is used in the technical sense introduced by R. A. Fisher (1925): The likelihood of a hypothesis is the probability it confers on observations, not the probability of the hypothesis, given the observations. The likelihood of H is $P(O|H)$; its posterior probability is $P(H|O)$. According to the law of likelihood, the observations differentially support the hypothesis of higher likelihood (Edwards 1972; Hacking 1965; Royall 1997).

The hypothesis that life can be traced back to a single progenitor is simpler than the hypothesis that it has 27 separate startups (since $1 < 27$). Crick's argument thus provides an example in which the principle of parsimony has a justification in terms of likelihood. However, the connection of likelihood and parsimony in this instance depends on specifically biological assumptions about the genetic code—that it is arbitrary and that it is subject to stabilizing selection. If parsimony has a rationale based on likelihood in inferential problems that arise in other sciences, different empirical assumptions will be required to show that this is so. But more important, there seem to be problems in which parsimony cannot be justified in terms of likelihood; in these problems, likelihood and parsimony are actually at odds.

The inferential task of curve fitting provides an example. Consider the following experiment. A sealed pot is put on a stove. Attached to the pot are a thermometer and a device that measures how much pressure the gas inside exerts on the walls of the pot. The pot is heated to various temperatures, and the resulting pressures are observed. Each temperature reading with its associated pressure reading can be represented as a point in a coordinate system (see Figure 1). The problem is to use these observations to determine the general relationship between temperature and pressure for this system. Each hypothesis about this general relationship takes the form of a curve. Which curve is most plausible, given the observations?

One factor that scientists take into account is goodness of fit. A curve that comes close to the data fits them better than a curve that is more distant. If goodness of fit were the only relevant consideration, scientists would always choose curves that pass exactly through the data points. But they do not do this—and even if they did, the question would remain how to choose among the infinity of curves that fit the data perfectly. Another consideration apparently influences their decisions, and this is simplicity. Often, extremely bumpy curves are thought to be complex, whereas smoother curves are thought to be simpler. Scientists sometimes reject an extremely bumpy curve that fits the data perfectly in favor of a smoother curve that fits the data slightly less well. Scientists care about both goodness of fit and simplicity, which influence how they choose curves in the light of data. However, these two desiderata conflict: Increasing simplicity typically involves reducing goodness of fit.

A curve represents a deterministic relationship between temperature and pressure; it maps x-values onto unique y-values. However, a curve plus an error distribution represents a probabilistic relationship: Any x-value is associated with a distribution of possible y-values, each with its own probability

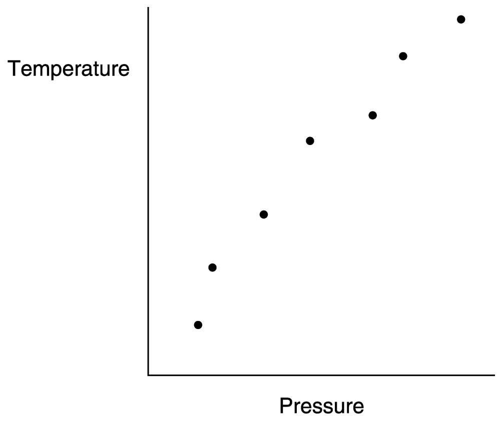


Fig. 1. Data gathered from an experiment in which a pot on a stove is raised to different temperatures and the pressure is recorded.

(density). In the example at hand (Figure 1), the concept of curve plus error distribution is more plausible, since the data are the joint product of the true underlying relationship of temperature and pressure and the measurement errors introduced by the imperfections of the thermometer and the pressure gauge. A standard model of error effects a connection between goodness of fit and likelihood: If one curve fits the data better than another, then the former confers a higher probability on the data. Given the data set depicted in Figure 1, a straight line will have a lower likelihood than a sufficiently complex curve that passes exactly through each data point. Thus, even if simplicity has a rationale based on likelihood in Crick's argument, simplicity and likelihood apparently conflict in the context of curve fitting.

Simplicity and Parsimony

It seems natural to say that curves differ in their simplicity. But what would it mean to say that they differ in parsimony? Parsimony involves paucity of postulation, but how does the idea of abstemiousness apply to curve fitting? Curves are visual representations of equations. For example, a straight line is a representation of a linear equation, which has the form

$$(\textsc{lin})y = a + bx$$

and a parabola is a representation of a quadratic equation, which has the form

$$(\textsc{par})y = a + bx + cx^2$$

where x and y are the independent and dependent variables, respectively, and a, b, and c are adjustable parameters. In such equations, the adjustable parameters represent existential quantifiers—for example, (LIN) says that there exist values for a and b such that $y = a + bx$. Therefore, (LIN), which makes two "existence claims," may seem more parsimonious than (PAR), which makes three. This point pertains to equations of the form (LIN) and (PAR), not to a specific straight line and a specific parabola (an important distinction, which will come up again). It is worth asking whether simplicity and parsimony in their vernacular meanings always come to the same thing. However, as noted at the outset, the present discussion follows the conventional practice of treating them as equivalent.

Bayesianism

Bayesianism is not the same as Bayes's theorem. The theorem says that the conditional probability $P(H|O)$—the probability of H, given O—is a function of three other quantities:

$$P(H \mid O) = P(O \mid H)P(H)/P(O).$$

This theorem is a consequence of the standard definition of conditional probability: $P(H \mid O) = P(H \wedge O)/P(O)$. Bayesianism is a philosophical position, not a mathematical truth; in its strongest form it asserts that the epistemic notion of plausibility can be understood in terms of the mathematical concept of probability and, furthermore, that all the epistemic concepts bearing on empirical inquiry can be understood in terms of the probabilistic relationships described by Bayes's theorem. A double application of this theorem yields the following comparative principle:

$$P(H_1 \mid O) > P(H_2 \mid O) \text{ if and only if } P(O \mid H_1)P(H_1) > P(O \mid H_2)P(H_2).$$

This biconditional makes it clear that Bayesianism can use exactly two ingredients in explaining how parsimony is able to render one hypothesis more plausible than another in light of a set of observations. If parsimony influences plausibility, it must do so through prior probabilities, likelihoods, or both. If the relevance of simplicity cannot be accommodated in one of these two ways, then either simplicity is epistemically irrelevant or (strong) Bayesianism is mistaken. As noted previously in connection with curve fitting, likelihood can be maximized by making one's hypothesis sufficiently complex; this seems to leave Bayesianism only one alternative: If simplicity in such cases influences the plausibility of a hypothesis, it must do so because simpler theories have higher prior probabilities. This led Jeffreys (1957) to introduce a "simplicity postulate," according to which the complexity of an equation is measured by summing its variables, exponents, and parameters. This simplicity ordering is then said to provide an ordering of the prior probabilities of the hypotheses.

Popper (1959) argued that this postulate is incompatible with the axioms of probability. It assigns (LIN) a higher prior probability than (PAR), but this is impossible, since (LIN) entails (PAR). Howson (1988) replied that the problem can be evaded by stipulating that the parameters in a model have nonzero values. Instead of comparing (LIN) and (PAR), we should compare (LIN*) and (PAR*), which stipulate that $a, b, c \neq 0$. These models are disjoint, not nested, so assigning (LIN*) a higher prior probability is consistent with the axioms.

This suggestion raises two new questions. First, why should the original problem—comparing (LIN) and (PAR)—be ignored? Should it be said that these two models are not in competition, because they are compatible? If so, scientific practice needs to change, since scientists often compare nested models.

Second, why should (LIN*) be assigned a higher prior probability than (PAR*)? Why think that $c = 0$ is more probable than $c \neq 0$? If probabilities are merely degrees of subjective belief, it is undeniable that someone might have greater confidence in the hypothesis that $c = 0$. However, it is puzzling why, in the absence of evidence, one should feel this way. If a sharp pin is dropped on a line a mile long, would you bet that the pin will land exactly at the beginning of the line or that it will land somewhere else? In the absence of information concerning how the pin is dropped, it is hard to see why you should bet on the first probability—yet this is precisely what Jeffreys's simplicity postulate recommends.

Another problem with this postulate has to do not with its correctness but with its completeness: It imposes an ordering of prior probabilities without providing specific values. This is important in inferential problems when the more complex hypothesis has a higher likelihood. If H_1 has the higher likelihood and H_2 has the higher prior probability, which has the higher posterior probability? Determining how simplicity trades off against likelihood requires more than a simplicity ordering.

Although Jeffreys held out no hope of getting likelihood and parsimony to coincide, later Bayesians saw a way to reopen the question. To grasp their idea, it is important to understand the difference between a model (which contains at least one adjustable parameter) and a specific hypothesis (which contains none). In this regard (LIN) is a model, but $y = 2 + 3x$ is not; it is a specific linear hypothesis. In effect, a model is a disjunction of specific hypotheses. When it was noted earlier that a sufficiently complex equation will fit the data better than a simpler equation, the point pertained to specific hypotheses. However, what would it mean to talk about the likelihood of a model? It is clear how $y = 2 + 3x$ "probabilifies" the data (once an error distribution is specified). But what probability does (LIN) confer on them? The answer is that the likelihood of (LIN) is the average likelihood of the set of straight lines ($i = 1, 2, \ldots$):

$$P(\text{data} \mid \text{LIN}) = \Sigma_i P(\text{data} \mid \text{straight line } i)\, P(\text{straight line } i \mid \text{LIN}).$$

The first term in this summation makes sense, but what should we make of the second? If the relation between temperature and pressure in the example of a pot on a stove is linear, what probabilities do the different specific linear hypotheses have? Schwarz (1978) approached this problem by thinking about the ratio of the average likelihoods of two models, assuming that there is a flat, uniform distribution over parameter values in each model. He derived the following result, which came to be known as the Bayesian information criterion (BIC):

$$\log[P(\text{data} \mid \text{model } M)] = \log\{P[\text{data} \mid L(M)]\} - (k/2)\log(N),$$

where $L(M)$ is the likeliest member of model M, N is the number of data, and k is the number of parameters in M. Notice that BIC includes a penalty term for complexity. If the best-fitting straight line and the best-fitting parabola fit the data in our example about equally well, (PAR) will have the lower estimated average likelihood because it is more complex. Complexity is relevant to estimating the average likelihoods of models, so Jeffreys's recourse to "priors" in his simplicity postulate is not, as it turns out, the only Bayesian approach to the problem.

One virtue of Schwarz's analysis is its avoidance of the criticism already noted—that it seems arbitrary and implausible, if not contradictory, to assign simpler models higher prior probabilities. (Nonetheless, questions can be raised about the assumed flat prior distribution of the values a parameter might have in a model.) Another virtue is that BIC specifies an exact quantitative rule for trading off simplicity and the likelihood of $L(M)$; it describes how much of a gain in one is required for a given loss in the other, if there is to be a net improvement in the estimated average likelihood of the model. However, there is a fly in the ointment. Schwarz's derivation uses improper priors (i.e., priors that do not sum to unity) in such a way that his derivation is not invariant under reparameterization (Forster and Sober 1994). Subsequent Bayesian work derives BIC so as to avoid this defect: The strategy is to use some of the data to transform the initial, improper priors into proper "posteriors"; thereafter, the rest of the data are taken into account to compute the final, average likelihood. (For further discussion, see Wasserman 2000.)

Popper and Falsifiability

Popper (1959) proposed a demarcation criterion that separates scientific from nonscientific statements:

The former are falsifiable. A falsifiable statement is one that is incompatible with a finite conjunction of observation statements. Falsifiable statements do not have to be false; rather, they have the nice property that observation can disprove them if, in fact, they are untrue.

Just as falsifiability separates science from nonscience, so degree of falsifiability distinguishes some scientific statements from others. The (LIN) model can be falsified by three data points, but not by any smaller number. A single data point, or any pair of data points, can be supplied with a straight line that passes through them exactly. On the other hand, (PAR) requires at least four data points to be falsified. This means that (LIN) is more falsifiable than (PAR).

Popper saw this as the key to understanding simplicity in science. Simpler theories are easier to falsify: If they are false, fewer data are required to show this. Popper turns Jeffreys's simplicity postulate on its head; whereas Jeffreys thinks that simpler theories are more probable, Popper thinks that simplicity goes with greater content: Simpler theories say more and hence are less probable.

It is clear that more falsifiable hypotheses have a pragmatic virtue: It is easier for us to prove them false, if they are false. The principal reservation philosophers have had regarding Popper's analysis is that it fails to account for the epistemic significance of parsimony. Why should predictions be based on simpler models rather than on more complicated models that fit the data equally well? It is here that Popper aligns himself with the skeptic and in opposition to the Bayesian. There is no assurance that our best hypotheses are true or even probably true. All that can be said is that they so far have evaded our best attempts to disprove them. Simplicity can provide no guarantee of truth or of probable truth, for the simple reason that nothing can.

There are further problems with Popper's account of simplicity. First, although it entails that (LIN) is simpler than (PAR), it does not have this consequence when a specific straight line and a specific parabola are compared. Each can be falsified by a single data point, so the two are equally falsifiable; this means that Popper must say they are equally simple. In addition, Popper's notion of degrees of falsifiability is restricted to hypotheses that have deductive consequences (perhaps in conjunction with auxiliary assumptions) about observations. If the hypotheses in question confer only probabilities on the data, they are not falsifiable. Since observation is virtually always subject to error, this is a large gap in Popper's theory.

Akaike and Selecting Models

The Bayesian approach to selecting models is not the only game in town. Before Schwarz (1978) proved his result, Akaike (1973) provided an alternative treatment (see also Sakamoto, Ishiguro, and Kitagawa 1986; Burnham and Anderson 1998). In fact, Akaike's contribution was twofold: He described a goal for selecting models—predictive accuracy—and he proved a theorem concerning how the predictive accuracy of a model can be estimated (Forster and Sober 1994).

How might a model like (LIN) be used to make a prediction about the pressure in our pot if the pot is brought to a certain temperature? A specific linear hypothesis, such as $y = 2 + 3x$, makes a prediction about the y-values that will be associated with newly observed x-values, but what does (LIN) tell us to expect? The answer is that (LIN) makes predictions by a two-step process. First, one uses old data to estimate the maximum likelihood values of the parameters in (LIN); then one uses this fitted model to predict new data. Thus, from the old data and (LIN) one obtains L(LIN), the likeliest member of (LIN); it is L(LIN) that makes a definite prediction about new data.

How well will L(LIN) predict new data? That depends, of course, on the true underlying relationship between temperature and pressure. In addition, since different data sets drawn from the same underlying distribution may differ, L(LIN) may make fairly accurate predictions about some data sets and rather inaccurate predictions about others. Because data sets may vary, it makes sense to define the predictive accuracy of a model as its average performance across multiple data sets.

If maximizing predictive accuracy is the goal, how is this goal to be achieved? How can we tell whether a model will make accurate predictions about new data, given just the single data set that we have at hand? If we opt for the model that best fits the data, we will usually select a fairly complex model. Working scientists know from practical experience that a complex model fitted to old data is often a poor predictor of new data; in such cases, the model is said to "overfit" the data. Sometimes a simpler model, although it does not fit the old data as well, will be a better predictor of new data. A mathematical explanation of this familiar fact is provided by Akaike's (1973) theorem:

$$\text{An unbiased estimate of the predictive accuracy of model } M = \log P\,[\text{data} \mid L(M)] - k,$$

where k is the number of adjustable parameters in the model. We obtain the log-likelihood of the best-fitting member of the model and then subtract k, which is a penalty for complexity. This estimate is called the Akaike information criterion (AIC) score of the model. Forster and Sober (1994) recommend representing the estimate per datum—that is, multiplying the right-hand side by $1/N$, where N is the number of data; this helps defuse the criticism that AIC is statistically inconsistent (Forster 2002). Although it is intuitive to think about Akaike's framework in the context of curve fitting, it and other criteria for selecting models apply to a far larger range of inference problems, including those that arise in causal modeling (Forster and Sober 1994).

Akaike's theorem, as such, must be considered for the assumptions that go into its proof. First, there is an assumption about the uniformity of nature, which has two parts: (1) It says that the old and new data sets described in the definition of predictive accuracy are drawn from the same underlying distribution, and (2) it assumes that the x-values sampled in different data sets are drawn from a single distribution. For this reason, Forster (2000) describes AIC as addressing the problem of interpolation; the model-selection criterion that would be appropriate for extrapolation is not described by Akaike's theorem, whose proof also requires an assumption about normality; roughly, this says that repeated estimates of a parameter in a model form a normal distribution.

What does it mean to say that AIC is unbiased? If your bathroom scale is unbiased, it may give different readings of what you weigh, but the average of these must be your true weight. If the scale is unbiased, so is the procedure of adding or subtracting 50 percent of what it says, depending on the result of tossing a fair coin. This second estimation procedure also is centered on the true value, but it has higher variance than the procedure that just takes the scale's reading at face value. Similarly, the fact that AIC provides an unbiased estimate of a model's predictive accuracy leaves open whether its estimates have minimum variance. Furthermore, it is not clear that lack of bias should be regarded as a necessary condition for an acceptable estimator. Suppose that a scale has very low variance but is slightly biased; on average, it reads a little too high or a little too low (it is not clear which). Would one decline to use this scale if the alternative is to use a scale that is unbiased but has enormous variance?

In the literature on selecting models, AIC and BIC are often treated as competitors. This is odd, since the two criteria were derived as solutions for different problems. BIC estimates average likelihood; AIC estimates predictive accuracy. This does not mean that they cannot be considered as possible solutions to the same problem; however, to do so involves wrenching one of them from its natural conceptual home. Forster (2002) describes a set of simulations in which AIC is better at estimating predictive accuracy in some circumstances, while BIC is better in others. If we knew in advance where the problem we want to solve is located in parameter space, such simulations might indicate which model-selection criterion to use. However, the sad fact of the matter is that we often do not know enough about the factual setting of a problem for this to be possible.

Akaike's framework and criterion have important implications for the debate concerning realism, empiricism, and instrumentalism. It often turns out that a model known to be false has a higher AIC score than a model known to be true. This means that the goal of finding predictively accurate models differs from the goal of finding true models. If realists maintain that the goal of science is to find true theories, and empiricists maintain that the goal of science is to find empirically adequate theories (van Fraassen 1980), then Akaike's framework and theorem open the door to a third possibility. Instrumentalism, shorn of the faulty philosophy of language, which led it to deny that theories have truth values, becomes an option worth exploring (Sober 2002).

Elliott Sober

The author thanks Malcolm Forster, Steven Nadler, and Kyle Stanford for helpful discussion.

References

Akaike, H. (1973), "*Information Theory as an Extension of the Maximum Likelihood Principle*," in B. Petrov and F. Csaki (eds.), *Second International Symposium on Information Theory* Budapest: Akademiai Kiado, 267–281.

Burnham, K., and D. Anderson (1998), *Model Selection and Inference: A Practical Information-Theoretic Approach*. New York: Springer.

Crick, F. (1968), "*The Origin of the Genetic Code*," *Journal of Molecular Biology* 38: 367–379.

Edwards, A. (1972), *Likelihood*. Cambridge: Cambridge University Press.

Fisher, R. (1925), *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Forster, M. R. (2000), "*Key Concepts in Model Selection: Performance and Generalizability*," *Journal of Mathematical Psychology* 44: 205–231.

——— (2002), "*The New Science of Simplicity*," in A. Zellner, H. Keuzenkamp, and M. McAleer (eds.), *Simplicity, Inference, and Modeling*. Cambridge: Cambridge University Press, 83–119.

Forster, M., and E. Sober (1994), "*How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will*

Provide More Accurate Predictions," *British Journal for the Philosophy of Science* 45: 1–36.
Hacking, I. (1965), *The Logic of Statistical Inference*. Cambridge: Cambridge University Press.
Howson, C. (1988), "*On the Consistency of Jeffreys's Simplicity Postulate and Its Role in Bayesian Inference*," *Philosophical Quarterly* 38: 68–83.
Jeffreys, H. (1957), *Scientific Inference*, 2nd ed. Cambridge: Cambridge University Press.
Kuhn, T. (1957), *The Copernican Revolution*. Cambridge, MA: Harvard University Press.
Leibniz, G. ([1686] 1973), *Discourse on Metaphysics*. LaSalle, IL: Open Court. (First published in 1840.)
Newton, I. ([1686] 1953), "*Rules of Reasoning in Philosophy*," from *Philosophiae naturalis principia mathematica*, in H. Thayer (ed.), *Newton's Philosophy of Nature*. New York: Hafner.
Popper, K. (1959), *Logic of Scientific Discovery*. London: Hutchinson.
Royall, R. (1997), *Statistical Evidence: A Likelihood Paradigm*. Boca Raton, FL: Chapman and Hall.
Sakamoto, Y., M. Ishiguro, and G. Kitagawa (1986), *Akaike Information Criterion Statistics*. New York: Springer.
Schwarz, G. (1978), "*Estimating the Dimension of a Model*," *Annals of Statistics* 6: 461–465.
Sober, E. (1988), *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, MA: MIT Press.
——— (1990), "*Let's Razor Ockham's Razor*," in D. Knowles (ed.), *Explanation and Its Limits*. Cambridge: Cambridge University Press, 73–94.
——— (2002), "*Instrumentalism, Parsimony, and the Akaike Framework*," *Philosophy of Science* 69: S112–S123.
van Fraassen, B. C. (1980), *The Scientific Image*. New York: Oxford University Press.
Wasserman, L. (2000), "*Bayesian Model Selection and Model Averaging*," *Journal of Mathematical Psychology* 44: 92–107.
Wood, R. (1996), *Ockham on the Virtues*. West Lafayette, IN: Purdue University Press.

See also **Bayesianism; Popper, Karl**

PARTICLE PHYSICS

Because its subject matter is the structure and behavior of the fundamental constituents of the physical world, particle physics involves a unique blend of experimental, theoretical, and philosophical issues. In the 1930s, early particle physics looked at cosmic ray traces in Wilson cloud chambers and results from particle accelerators probing distances of the order of the size of an atomic nucleus (energies of about 100 million electron-volts [MeV]) (Brown and Hoddeson 1983). Particle physics made use of a patchwork of models that included quantum electrodynamics, Enrico Fermi's theory of radioactive β-decay, and Hideki Yukawa's meson theory. Today, the Tevatron at Fermilab collides protons and antiprotons accelerated to 980,000 MeV to create charged jets of strongly interacting particles that have probably not existed in the universe since the big bang. The standard model in particle physics, with its unified theory of three of the four fundamental forces in nature, is the current dominant paradigm. A central conceptual problem of particle physics since its inception has been to define and establish an appropriately tight connection between an account of the basic building blocks of matter (fundamental theory) and experimental predictions (phenomenological models).

Early Particle Physics

Particle physics emerged in the 1930s as a confluence of three distinct fields of physics: nuclear physics, cosmic ray physics, and quantum field theory (QFT). The consensus at the time was that matter was composed entirely of two fundamental particles—negatively charged electrons and positively charged protons—and that there were two fundamental forces of nature: electromagnetism, mediated by the photon, and gravitation. Atoms were thought to be composed of electrons orbiting a nucleus composed of protons and electrons bound tightly together. Early observations of fair-weather "atmospheric electricity" at the turn of the twentieth century had expanded by late 1920s into a well-developed program of research studying cosmic rays, that is, ionizing radiation from outer space. This radiation was thought to consist of high-energy photons.

On the theoretical side, physics by the mid-1920s appeared equally successful and complete. Einstein's theories of special and general relativity gave a framework for understanding gravitation and the macroscopic structure of the universe (see Space-Time). The theory of quantum mechanics, established in 1925 and 1926, offered a successful account of the dynamics of atomic particles (see Quantum Mechanics). Modulo a smattering of puzzles and anomalies, relativity and quantum mechanics appeared to cover all the fundamental laws of nature. Early particle physics (1930–1947) is largely the story of how this picture broke down.

The main theoretical challenges were to develop a relativistic account of the electron and a quantum theory of fields (see Quantum Field Theory). Paul Dirac's 1927 theory of quantum electrodynamics (QED), an early QFT, allowed one to calculate the probability distributions for various observable properties of an electron in the relativistic, high-energy domain. As Werner Heisenberg pointed out, however, Dirac's theory had a troubling consequence. In classical relativistic mechanics, particle transitions from positive to negative energy states are impossible. Dirac's theory, by contrast, predicted that positive-energy, negatively charged electrons should fall into negative-energy, positively charged states, rendering all matter highly unstable (Brown and Hoddeson 1983). In early 1930, Robert Oppenheimer suggested a possible solution to this problem: the transitions to negative-energy states do not occur because all these states are already filled. The only positively charged particle known at that time was the proton, but its mass (4,000 times that of the electron) made it difficult to fit into Dirac's theory. It took a year for Dirac publicly to speculate that his theory predicted the existence of a new particle, with the same mass but opposite charge of the electron. It took another year for Carl Anderson to publish cosmic ray experiments showing cloud chamber tracks of a new particle, dubbed the positron, fitting Dirac's description.

Anderson's discovery of the positron had a large impact on early particle physics, making it possible for experimentalists and theortists to extend their ontology beyond the two-particle (electron and proton), two-force (electromagnetism and gravity) picture. Two new nuclear forces were introduced, along with an array of new particles. The subsequent discovery of the neutron solved fundamental problems of nuclear physics. The weak nuclear force was introduced to explain the radioactive decay of nuclei and, in particular, the decay of the neutron into a proton, electron, and antineutrino. Finally, Hideki Yukawa developed a revolutionary model introducing the strong nuclear force. Yukawa knew that the force binding protons and neutrons in the atomic nucleus was much stronger than the electromagnetic force. He visualized this strong nuclear force as due to an exchange of some particle. However, no known particle fit the bill: The particle had to have a mass somewhere between 200 times more than the electron and 10 times less than the proton. This speculation, published in November of 1934 (Yukawa 1934; Brown 1981), was extravagant: there must exist a new quantum field requiring a novel kind of particle called a meson, a particle as massive as the electron yet behaving statistically like the photon. A series of cosmic ray experiments in 1937 revealed this new meson, and by the end of the 1930s Yukawa's meson field theory had become universally accepted. (The first nuclear bombs were constructed using Yukawa's meson field theory.) Meson theory became (and remains today) a paradigm for particle physics, in which the fundamental structure of matter involves the creation and annihilation of esoteric and ephemeral particles that have no analog or basis in classical physics and in which abstract, highly mathematical theories are necessary for an acceptable interpretation.

By the early 1940s, particle physics had become an established field of physics, the primary goal of which was the search for fundamental particles and fields. Particle physics could account for three of the four forces of nature (gravity is too weak to be effective in the subatomic domain) and offered a satisfying classification of matter at a fundamental level reminiscent of Mendeleev's periodic table. Particles that underwent strong interactions, such as the proton, neutron, and Yukawa's meson, were classified as hadrons. Particles that underwent only weak or electromagnetic interactions were classified as leptons, and these included the electron, photon, and neutrino. But perhaps the most important result from this period is the classification of particles based on their spin properties, and the profound (and to this day mysterious) connection between spin and the statistical behavior of particles first investigated by Pauli (1940).

Elementary particles, such as electrons, are identical. This means that given a system of *n* electrons, if one state of that system is allowed by the dynamical laws of nature, so will any other state that can be obtained by permuting (switching) the particles. How does nature know which of these states is the correct one? Or are they all correct? Nature, it turns out, divides particles into two exhaustive and mutually exclusive groups with distinct symmetry (permutation) properties: particles of integer spin (0, 1, 2, ...), called bosons, and particles of half-integer spin ($\frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, ...), called fermions. These symmetry properties have profound consequences for the structure of the universe: they enable fermions (e.g., electrons, protons, neutrons) to form substantial matter, and bosons (e.g., photons, mesons) to provide the glue that holds this matter together at the atomic and subatomic levels. This connection between spin and statistical properties of particles was established by Pauli based on very general principles such as Lorentz invariance, microscopic causality, and the conservation of probability (unitarity) (see Quantum Field Theory).

A Proliferation of Symmetries and Particles

From the end of World War II until the mid-1960s, particle physics was driven by a flood of experimental results that completely swamped efforts to develop a unified understanding of the subatomic domain. Particle physics no longer had to rely on cosmic ray observations, as a new generation of powerful particle accelerators was able to produce and detect high-energy particles in a more systematic and efficient way. This resulted in the creation and detection of dozens of new particles, from the *k*-mesons (kaons) of the early 1950s to the ultra-heavy omega-minus (Ω^-) hadron in 1964. It also resulted in astonishing discoveries, such as that of parity violation, which upset deeply held beliefs about the fundamental structure of matter. One common intuition about space, famously articulated by Kant, is that there can be pairs of objects, such as a left hand and a right hand, such that one object is the mirror image of the other, but no rigid motion (translation or rotation) can transform the one into the other. While familiar objects in the world commonly exhibit this kind of handedness, Kant assumed (as have all philosophers and physicists since) that the fundamental structure of nature does not.

Parity is the idea that for any physically possible state, the mirror image state (obtained by inverting one spatial coordinate) is also physically possible, and this obvious and basic spatial symmetry was implicitly assumed in theories in particle physics. Testing for parity is difficult, however, since expectation values of most quantum observables are equal for physical processes and their mirror images. Important exceptions are pseudoscalar processes, those that are the scalar product of vector quantities (which transform $\mathbf{v} \to -\mathbf{v}$ under reflection) and axial-vector quantities (which transform $\mathbf{a} \to +\mathbf{a}$ under reflection). Nonzero pseudoscalar processes were observed in an experiment by Wu on the β-decay of cobalt 60 in 1957 (Wu et al. 1957). This violation of parity at a fundamental level came as a severe shock to the particle physics community, and the question posed by Pauli still has not been answered: what physical reason is there for parity to be violated in weak interactions but conserved in strong and electromagnetic ones?

Other symmetries, such as *strangeness*, were invented to try to make sense of particular experimental results. Certain kinds of particle interactions that should be physically possible were not observed, or observed only rarely. The negatively charged sigma particle ($\sum^-$), for instance, was readily produced by the strong interaction $\pi^- p \to K^+ \sum^-$ (where the pion [π^-], proton [p], and kaon [K^+] are all particles), but it had an anomalously long lifetime and was observed to decay only weakly via $\sum^- \to n\pi^-$ (where n is the neutron). An economical and powerful way of accounting for these facts is to posit a new symmetry (an additive quantum number), that strangeness, and to stipulate that in strong and electromagnetic interactions strangeness is conserved (these interactions are symmetrical with respect to strangeness), while strangeness is violated in weak interactions. The absence of the strong decay $\sum^- \to n\pi^-$ is explained by the fact that this decay fails to conserve strangeness. In this way, strangeness is used to explain why certain particles undergo strong or electromagnetic interactions in some situations but not in others. The strangeness symmetry scheme, and many others like it, were clearly stopgap measures in the face of huge amounts of new data and the general failure of quantum field theoretic approaches to predict, retrodict, or explain these results.

Quantum field theoretic approaches also faced a serious problem having to do with the presence of infinities in the theory. For any QFT describing interactions, predictions cannot be derived analytically, but only by means of perturbation techniques. In a perturbation expansion, these QFTs give rise to divergent integrals that yield values that differ, infinitely, from experimental results. Renormalization techniques eliminate these divergent integrals and render the corresponding phenomenological models predictively accurate (see Quantum Field Theory). A guiding principle in particle physics almost since its inception has been that only renormalizable QFTs are even candidates for true fundamental theories, because only renormalizable quantum field theories exhibit the appropriately tight connection between an account of the basic building blocks of matter and experimental predictions. For QED the problem turned out to be soluble by means of renormalization techniques developed by Richard Feynman and others in the late 1940s (Schweber 1994). QFTs of the strong and weak nuclear forces, by contrast, were not renormalizable and did not work very well computationally.

By the early 1950s, the ability of a theory to produce the right numbers was taken to be its primary, if not exclusive, virtue. A new generation of experiments in particle physics was producing unprecedented amounts of data, enabling a great increase in the accuracy of experimental measurements. There was a widespread conviction among physicists that it would be many years before field-theoretic approaches would yield a theory of the strong or weak interactions that gave the right numbers. This feeling persisted right to the end of the 1960s, and the developments described below—the rapid rise of QFT in the early 1970s—thus came as a surprise.

But which elements of the field-theoretic approach should be abstracted and incorporated into the new, more successful theories? Work in theoretical high-energy physics in the 1950s and 1960s was guided by two main answers to this question. One approach attempted to salvage much of the QFT model, but without the notion of a physical field or the claim to be describing nature at a fundamental level. What replaced field theories were what physicists called *phenomenological theories:* computational schemes of limited scope and accuracy based closely on experimental results and heuristic considerations. The two most well known examples of this kind of approach are the vector/axial-vector (V-A) model of weak interactions and the current-algebra approach to strong interactions (which involves abstracting an interaction current from a hypothetical field and exploring its algebraic properties) (Brown, Dresden, and Hoddeson 1989). Both of these approaches use a field-theoretic apparatus but without any notion of a physical field or particle mediating the interactions. Murray Gell-Mann's oft-quoted culinary metaphor is apt to describe how field-theoretic approaches were used:

> In order to obtain such relations that we conjecture to be true, we use the method of abstraction from a Lagrangian field theory model. In other words, we construct a mathematical [model] . . . , which may or may not have anything to do with reality, find suitable algebraic relations that hold in the model, postulate their validity, and then throw away the [field-theoretic] model. We may compare this process to a method sometimes employed in French cuisine: a piece of pheasant meat is cooked between two slices of veal, which are then discarded. (Gell-Mann 1964, 73)

The S-Matrix Program and Particle Democracy

A more radical approach eschewed the field-theoretic paradigm completely. In the mid-1950s, it was suggested that several constraints of a general sort (such as Lorentz invariance, analyticity, unitarity, and simple boundary conditions) might be enough to specify both strong-interaction dispersion relations and the scattering matrix (S-matrix), two important calculational results that had traditionally been derived using field-theoretic methods (Chew 1962; Cushing 1990). These results describe virtually all experimentally observed quantities and enable one to make empirical predictions. By 1961, an S-matrix research program was well established, which would dominate strong-interaction physics for the next decade. As the leading proponent of the S-matrix research program, Geoffrey Chew (1961), put it:

> So that there can be no misunderstanding of the position I am espousing, let me say at once that I believe the conventional association of fields with strongly interacting particles to be empty. . . . I am convinced that future development of an understanding of strong interactions will be expedited if we eliminate from our thinking such field-theoretical notions as Lagrangians, "bare" masses, "bare" coupling constants, and even the notion of "elementary particles." (3–4)

The most radical philosophical aspect of the S-matrix program is its proposal of "particle democracy," the idea that the concept of an elementary particle, as it is commonly used in field-theoretic approaches, is incoherent. QFTs proceed by writing down a Lagrangian (field equation) based on information about elementary particles and their corresponding fields, from which eventually an S-matrix is derived. This entails a distinction between elementary particles—those that correspond to quantum fields—and composite particles such as bound states; this distinction is important, since by definition elementary particles and their corresponding fields are those that are sufficient to determine the S-matrix. But this definition begs the question. Only once a complete dynamical theory has been developed can one say which particles are elementary and which particles are not—but in order to solve the dynamical problem, according to QFT, one must know which particles are elementary. Thus, the notion of an elementary particle is not well defined in QFT.

The S-matrix program avoids this problem by offering a dynamical theory that does not require the notion of an elementary particle: all particles, bound states, and resonances are ontologically equal. Of course, the S-matrix program offers a different sort of conception of dynamics than that offered by field theory, and in two major ways. First, S-matrix theory does not offer a picture of the unobservable elements of strong interaction dynamics, that is, what happens between the measured initial and final states. QFT, on the other hand, appears to offer a picture of particles constrained by potential fields. According to its proponents, S-matrix theory was superior to field theory in this regard, because the picture of the unobservable dynamics offered by QFT is at best unnecessary and at worst extremely problematic (given the above-mentioned problem with the notions of elementary particle and field). Second, field theory is foundational in the sense that it contains a small set of fundamental elements from which all the dynamics can be calculated. S-matrix theory, by contrast, is anti-foundational. Without the elementary-particle concept to focus attention on particular elements of the S-matrix, the question immediately arises: Where does one begin a dynamical calculation? The answer is that it does not matter; one may begin anywhere, taking an arbitrary "particle" (bound state, resonance, etc.) as a starting point and attempting to reach as much of the S-matrix from this point as computational ability allows.

Quantum Field Theory Redux: The Standard Model

The establishment of the standard model in the early 1970s brought with it a renaissance of QFT as the central theoretical tool in particle physics (Hoddeson et al. 1997). In fact the standard model may be described somewhat loosely as two applications of QFT to the real subatomic world, namely quantum electroweak dynamics (QEWD), which covers electromagnetic and weak interactions, and quantum chromodynamics (QCD), which covers strong interactions.

QEWD, first proposed in 1967, extends QED to cover weak interactions by means of a mathematical technique called *spontaneous symmetry breaking* (Weinberg 1967; Salam 1968). This technique allows QEWD to have symmetries that are broken, or not present, in nature. An example of spontaneously broken symmetry in nature is a metal rod, balanced vertically on a flat surface, where a downward force is applied to the top end. The complete theoretical description of the vertical rod and the vertical downward force is rotationally symmetrical around the vertical axis. But if the force is large enough, the rod will buckle, breaking the rotational symmetry of the system. Spontaneous symmetry breaking was the key to a unified theoretical description of electromagnetic and weak forces, and it also provided a theoretical mechanism for giving rise to masses of the leptons and quarks that make up matter and some of the interacting bosons. QEWD was extended to cover electromagnetic and weak interactions of hadrons in 1970 (Glashow, Iliopoulos, and Maiani 1970), was shown to be renormalizable in 1972 (t'Hooft and Veltman 1972), and received a key experimental confirmation with the detection of weak neutral currents in 1974 (Galison 1987 and 1997).

QCD faced an additional conceptual hurdle. In the 1960s, several physicists proposed that hadrons, such as protons and neutrons, were not fundamental entities but were composed of point particles with fractional electric charges, called quarks. Experiments seemed to show that these quarks floated freely inside hadrons. So, why did the hadrons not just fly apart? And why had no single quarks (easily identifiable by their fractional charge) ever been observed? David Gross, the founder of QCD, reasoned that the strong interaction had to be strong at large distances, to keep the quarks together inside hadrons as well as to account for observed strong-interaction phenomena, but very weak at short distances, to account for their behavior as free particles inside hadrons and thus for the experimental results (precisely the opposite of the gravity, electromagnetism, and the weak force!) (Gross and Wilczek 1973). Unlike electrical charge, which is bivalent (positive and negative), the strong charge in QCD is trivalent. Its three values are conventionally called 'colors'—red, green, and blue—and there are eight bosons, called gluons, mediating the strong color force. There remains, however, the problem of confinement: No free quarks have ever been observed, no free gluons have ever been observed, and no color charge has ever been observed (all observed particles are color neutral). On a theoretical level, QCD does not provide a dynamical mechanism that explains confinement. On a conceptual level, the question remains whether and in what sense one can have a fundamental QFT in which the elementary entities literally do not exist in free space.

QEWD and QCD, which collectively make up the standard model, are characterized by a Lagrangian whose form is determined in part by the constraints of gauge invariance and renormalizability (see Quantum Field Theory). Connection with experimental results is achieved through phenomenological models. In contrast to S-matrix, V-A, and current-algebra theories, the standard model was taken to offer a *fundamental* theory of the subatomic domain, and this was understood as the source of many of its theoretical virtues. The standard model admits a realistic interpretation, in the sense that the matter and interaction fields occurring in the basic Lagrangians can be taken to describe physical quantum fields that produce and explain observed scattering cross-sections and other experimental results. The realist interpretation, moreover, presents a tidy ontological picture in which there are only a small number of fundamental entities: 24 or so basic fields and their associated particles. This realistic picture can be well confirmed by experimental results, in a way that purely phenomenological theories cannot, because confirmation of each phenomenological model accrues also to the fundamental theory from which the model is derived. The point most emphasized by physicists, however, is the unity of the standard model at a fundamental level and the understanding of nature that follows: the whole subatomic realm is composed of a small number of elementary particles and fields, and all dynamics follow from a single equation of motion (Weinberg 1980; Wayne 1996).

The standard model's status as a fundamental theory, and the theoretical virtues that follow, depend on an appropriate connection between QFTs and phenomenological models. Roughly, what is required is that a phenomenological model *is* what the fundamental theory says in the particular circumstances covered by the model. In a semantic approach to scientific theories, this is usually taken to mean that the phenomenological model is a submodel of the fundamental theory (see Theories). In a syntactic approach, the relation is typically taken to be one of deductive entailment, where the phenomenological theory is deduced from the fundamental theory plus a description of the specific conditions in which the phenomenological theory applies. In the standard model, as in earlier QED, the connection is secured by renormalization. Standard-model physics, like early particle physics, has as its central task the search for a fundamental theory. Thus the requirement that standard-model QFTs be renormalizable has functioned as a regulative principle in theory construction and is supposed to yield QFTs that are candidates for true descriptions of the world at all scales, from the ultra-high-energy Planck length to the macroscopic world. (The Planck length, about 10^{-35} meters, is the distance at which gravitational effects become as strong as those due to the other forces of nature.)

Since the standard model was established in the mid-1970s there have been a variety of challenges to its status *qua* fundamental theory, and for the most part, these challenges hinge on the issue of renormalization. A philosophical challenge stems from the fact that renormalization techniques involve steps in which additional information is introduced. This happens first when the QFT is regularized: divergent integrals are replaced with finite ones, via the introduction of momentum cutoffs, dimensional regularization, or some other regularization procedure. Thus, phenomenological theories are not submodels of, or deductive consequences of, the fundamental QFTs (Huggett 2002). There are also a number of technical worries about the mathematical stability, coherence, and domain of applicability of perturbative renormalization techniques (Cao and Schweber 1993). Finally, attempts to extend the standard model to cover the fourth force of nature, gravity, have been unsuccessful. These attempts introduced a new quantum field and associated particle, the graviton, with mass zero and spin 2, which propagates in a flat Minkowskian space-time. It was found, however, that perturbative quantum gravity was not renormalizable and hence not a candidate for fundamental theory (Isham, Penrose, and Sciama 1981). The development of particle physics since the 1970s has been framed by the resulting dilemma: One can retain standard-model QFTs yet abandon the claim that they represent fundamental theory, or one can abandon standard-model QFTs and focus on alternative routes to fundamental theory.

The effective field theory (EFT) program, first developed by Weinberg (1979) in the late 1970s, takes the first horn. The EFT program begins with the assumptions that QFTs at different energy scales take different forms and that no one QFT is applicable across all energy scales. Fermi's nonrenormalizable β-decay model of the weak nuclear force is a prime example of an EFT, valid at an energy scale (up to 300 GeV) set by the masses of the heavy W and Z bosons mediating weak interactions. In this case, there is a renormalizable QFT—electro-weak theory—of which

Fermi's model is a low-energy approximation. In general, however, EFTs are constructed without knowledge of whether there exists any renormalizable theory at a higher energy scale. In the EFT approach, the regulative principle of renormalizability is replaced with a principle that nonrenormalizable EFTs are to be pursued subject to a constraint, which can be formulated precisely, concerning the partial decoupling of the physics in the domain of the EFT from the physics in higher-energy domains. In short, the EFT approach consists of phenomenological models all the way down.

The superstring research program, which began in the mid-1980s, grasps the second horn of the dilemma (Callender and Huggett 2001). The superstring program retains renormalizability as a guiding principle in theory construction and aims at a unified description of the fundamental forces of nature. Superstrings are one-dimensional closed strings that propagate in a space-time of dimension greater than 4, and quantization results in a theory that is perturbatively renormalizable. Superstring theory fits squarely within the fundamentalist tradition in particle physics, in which a Lagrangian field theory is combined with perturbative techniques and renormalization to yield phenomenological models. Work since the 1990s on the foundations of superstring theory is beginning to suggest ultimate limits on the applicability of this sort of approach, stemming in part from the result that the very concept of a space-time manifold is not applicable at the Planck length. Whatever theoretical form it takes next, the desire within particle physics for an account of the fundamental structure of the physical world remains strong.

Andrew Wayne

References

Brown, L. M. (1981), "*Yukawa's Prediction of the Meson*," *Centaurus* 25: 71–132.

Brown, L. M., M. Dresden, and L. Hoddeson (1989), *Pions to Quarks: Particle Physics in the 1950s: Based on a Fermilab Symposium*. Cambridge and New York: Cambridge University Press.

Brown, L. M., and L. Hoddeson (1983), *The Birth of Particle Physics*. Cambridge and New York: Cambridge University Press.

Callender, C., and N. Huggett (2001), *Physics Meets Philosophy at the Planck Scale: Contemporary Theories in Quantum Gravity*. Cambridge and New York: Cambridge University Press.

Cao, T. Y., and S. S. Schweber (1993), "*The Conceptual Foundations and the Philosophical Aspects of Renormalization Theory*," *Synthese* 97: 33–108.

Chew, G. (1961), "*The S-Matrix Theory of Strong Interactions*," *Lawrence Radiation Laboratory Preprint, UCRL-9701, May 15, 1961*. New York: W. A. Benjamin and Company.

——— (1962), "*S-Matrix Theory of Strong Interactions without Elementary Particles*," *Reviews of Modern Physics* 34: 394–401.

Cushing, J. T. (1990), *Theory Construction and Selection in Modern Physics: The S-Matrix*. Cambridge: Cambridge University Press.

Galison, P. (1987), *How Experiments End*. Chicago: University of Chicago Press.

——— (1997), *Image and Logic: A Material Cultural of Microphysics*. Chicago: University of Chicago Press.

Gell-Mann, M. (1964), "*The Symmetry Group of Vector and Axial Vector Currents*," *Physics* 1: 63–75.

Glashow, S., J. Iliopoulos, and L. Maiani (1970), "*Weak Interactions with Lepton-Hadron Symmetry*," *Physical Review* D2: 1285–1292.

Gross, D., and F. Wilczek (1973), "*Asymptotically Free Gauge Theories I*," *Physical Review* D8: 3633–3652.

Hoddeson, L., M. Riordan, M. Dresden, and L. Brown (eds.) (1997), *The Rise of the Standard Model: Particle Physics in the 1960s and 1970s*. Cambridge: Cambridge University Press.

Huggett, N. (2002), "*Renormalization and the Disunity of Science*," in M. Kuhlmann, H. Lyre, and A. Wayne (eds.), *Ontological Aspects of Quantum Field Theory*. Singapore: World Scientific, 255–280.

Isham, C. J., R. Penrose, and D. W. Sciama (eds.) (1981), *Quantum Gravity: A Second Oxford Symposium*. Oxford: Clarendon Press.

Pauli, W. (1940), "*The Connection Between Spin and Statistics*," *Physical Review* 58: 716.

Salam, A. (1968), "*Weak and Electromagnetic Interactions*," in N. Svartholm (ed.), *Proceedings of the 8th Nobel Symposium*. Stockholm: Almqvist and Wiksell, 367–377.

Schweber, S. S. (1994), *QED and the Men Who Made It: Dyson, Feynman, Schwinger, and Tomonaga*. Princeton, NJ: Princeton University Press.

t'Hooft, G., and M. Veltman (1972), "*Regularization and Renormalization of Gauge Fields*," *Nuclear Physics* B44: 189–213.

Wayne, A. (1996), "*Theoretical Unity: The Case of the Standard Model*," *Perspectives on Science* 4: 391–407.

Weinberg, S. (1967), "*A Model of Leptons*," *Physical Review Letters* 19: 1264–1266.

——— (1979), "*Phenomenological Lagrangians*," *Physica* 96A: 327–340.

——— (1980), "*Conceptual Foundations of the Unified Theory of Weak and Electromagnetic Interactions*," *Reviews of Modern Physics* 52: 512–523.

Wu, C.-S., E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson (1957), "*Experimental Test of Parity Conservation in Beta Decay*," *Physical Review* 105: 1413–1415.

Yukawa, H. (1934), "*On the Interaction of Elementary Particles I*," *Proceedings of the Physico-Mathematical Society of Japan* 17: 48–57.

See also **Quantum Field Theory; Quantum Mechanics; Space-Time**

PERCEPTION

Perception is a traditional focus of philosophical investigation. It has frequently been an area of collaboration between science and philosophy. Plato, Aristotle, Alkindi, Alhazen, Descartes, and Berkeley are philosophers who count as pivotal figures in the history of perceptual science (Lindberg 1976). The study of perception is an area in which sharp lines between science and philosophy get erased. This is one reason why the theory of perception is of particular interest to the philosophy of science. Another reason is that the theory of perception is one of the best-developed areas of scientific study of the mental. Empirical theories of perception stand as object lessons in what a science of the mind might hope to accomplish.

This article reviews some important themes in the theory of perception, with an eye to the dialogue between science and philosophy. It begins, however, with some general comments on the philosophy of perception.

The Philosophy of Perception: An Overview

Philosophical puzzlement about perception stems from two basic facts:

1. Perceptual states (experiences, percepts) are subjective states of consciousness. It is natural to think that perceptual states happen in the perceiver (in the mind, say, or in the nervous system). Perceivers know such states from the inside. Perceptual experiences have a definite felt character. Visual experiences are different from auditory ones, and smelling is different from touching. Whatever else they are, these are differences in their felt character.
2. Perceptual experiences are world referring. They always present things as being some way or other. One might feel that the temperature has gone down in the room, and then look up to see that the fire is out. In this way, perceptual experience raises the question, at least implicitly, of whether things are as they are experienced as being. Perceptual experiences are always either veridical or nonveridical. Many philosophers assume that perceptual experiences are world referring because they are *intentional*, in the philosopher's technical sense: Like thoughts or sentences, perceptual experiences are *about* the world (see Intentionality).

The central challenge faced by any theory of perception is to explain each of these features and how they can both be true.

Many writers have assumed that insofar as perceptual states are subjective, they are a special kind of bodily sensation. For example, there is the sensation of a pin prick, and there is the sensation of red. Bodily sensations, however, are not world referring. They do not present the world as being a certain way. A headache, for example, may hurt (and perhaps suggest overindulgence), and the experience of the headache may likewise inform one of an event taking place in a particular part of one's body (e.g., the left temple), but the headache itself, a mere sensation, does not refer beyond itself; it is not intrinsically *as of* a state of affairs. What is it, then, about *perceptual* sensations that differentiates them from mere bodily sensations and allows them to present the world? One influential strategy for addressing this issue has been to view sensations as, as it were, "natural signs" of that which causes them. This is the guiding metaphor of much empiricist work on perception (in the tradition of Locke, Berkeley, Hume, and Reid). Sensations (or ideas or impressions) are *effects* in one produced by things in the world. Knowledge of what is going on in the world—the very ability of one's perceptions to present states of affairs in the world—derives from one's experience of such patterns as similarity, proximity, and conjunction in simple sensory effects.

This approach has two clear virtues: (a) It is naturalistic, conforming to a plausible model of perceivers as animals on whom the environment impinges through different patterns of physical stimulation, and (b) it nicely addresses the first fact mentioned above, that perceptual states are states of consciousness. In the empiricist view, perceptual states are sensations (see Empiricism). A basic problem with the empiricist strategy, however, is that it falls short of providing an adequate account of the world-referring character of perceptual experience. As a result, it actually fails to do justice to the character of perceptual consciousness.

Consider the visual experience of deer grazing in a meadow. This is an experience of deer, or at least of what visually appear to be deer. It is not an experience of sensations on the basis of which one infers or concludes that it is the experience of deer. The empiricist confuses the question of veridicality and that of intentionality. It is surely true that there is nothing in the visual experience of deer in the meadow that, as it were, self-certifies the experience as veridical. In that sense, perceptual *judgment* goes beyond what is given in experience. The mere fact that it looks to one as if there are deer is not enough to settle the question of whether there are. However, from this it does not follow that there is also an open question about *how* the experience presents things as being. One may be mistaken that there are deer; but one is not likely to be mistaken about the fact that it looks as if there are. The intentional content of an experience—how it presents things as being—is given *immediately* with the experience. The empiricist fails to account for the intrinsic intentionality of perceptual states and, because of this, misdescribes what sort of conscious states perceptual experiences are. The upshot of these considerations is that perceptual states are not composed of bodily sensations. One needs to look elsewhere for an account of their qualitative character.

Intentionality may provide part of the solution. After all, surely a good deal of what determines the qualitative character of a given perceptual experience is the fact that it is, say, the visual experience *of deer grazing in a meadow*. Some philosophers believe, however, that this cannot be the whole story (e.g., Peacocke 1983). There are aspects of what it is like to have a perceptual experience that are not fixed by the intentional content as features of the way the experience presents the world as being. Philosophers refer to these additional features as qualia (singular: quale), or phenomenal or sensational properties of experience.

In addition, many theorists now believe that not all perceptual states with intentional content are phenomenally conscious. As an example of this, consider "blindsight." Perceivers with damage in the visual cortex have scotomas, or blind fields, where they cannot see. If asked to identify objects presented in the blind field, such patients respond that they cannot see them. If asked to guess as to the identity of presented objects, however, such patients may answer correctly significantly higher than would be expected from chance alone. In cases such as this, it is tempting to say that there is an absence of phenomenal consciousnesss but the presence of intentional content.

Phenomena such as blindsight, if taken at face value, would seem to show that perceptual phenomenology cannot be just a matter of perceptual intentionality. Further support for this conclusion is provided by a consideration of *visual agnosia*, which may exemplify perceptual awareness without intentionality. Some perceivers with ventral stream damage are unable, on the basis of vision, to form a clear judgment of the properties of the scene before them, although they are able to use what they see to guide action. For example, they might be unable to judge whether a slot is vertical or horizontal, but they would be able to orient an envelope appropriately so as to guide it through the slot. Such perceivers see, but it is unclear to what extent their perceptual states are intentional. (It is also unclear to what extent their action-guiding perceptual states are consciousness.)

Perceptual consciousness and the intentionality of perceptual states are central areas of research in contemporary perceptual theory, whether in science or in philosophy.

Two Traditional Puzzles about Perception

Science Versus Common Sense

The world perceptually seems to be noisy, colorful, and full of odor and flavor. Science would seem to teach, however, that color, sound, odor, and flavor, like the sensations of touch, are sensory effects in perceivers brought about by their contact with the world. From this standpoint, one is no more entitled to believe that redness inheres in the tomato than that *paininess* inheres in the pin that pricks us. This is what Galileo, Boyle, Locke, and Newton believed.

If perception informs one not of how things are in themselves, but merely of one's subjective alterations in the face of things, then how can perception ever be a source of knowledge? Some philosophers, such as Descartes, bite the bullet and seem willing to deny that perception can on its own be a source of knowledge. Science proceeds *despite* the misleading influence of perceptual experience. Other philosophers have accepted the scientific starting point but have sought to unsettle the apparent consequence that perceptual knowledge is impossible. Locke seems to have thought that there is enough similarity between interior representations of things and the way they really are to allow perceptual states to be a source of knowledge. It is difficult to see how such a position could be coherent, however. After all, one is never in a position to test whether things are the way they are represented in experience. To do that, one would need, impossibly, to compare experience and reality from a neutral perspective.

Phenomenalism is another strategy. Hume, for example, granted that what one knows are the subjective contents of the mind, but he asserted that that is all one needs to know, for objects and states of affairs are themselves just patterns of organization in subjective states of consciousness. One problem with this view is that it seems phenomenologically far-fetched. Perceivers take themselves in experience to be aware of things whose nature is independent of their sensory effects, such as shoes, sunsets, and the like (see Phenomenalism).

Kant argued, against Hume and Locke, that if the character of experience were determined, exhaustively, by sensory effects set up in perceivers by their interaction with the world, then experience could not even *seem* to present the world as being this way or that. In order for mere sensory effects to rise to the level of glimpses of a mind-independent world, one must already take oneself to have access to that world, and one must already have concepts of objects and other phenomena to apply to mere sensory stimulation. Only then can one have experience with genuine, world-presenting content. The debate between Kant and empiricism still shapes thinking in the philosophy of perception (see e.g., McDowell 1994).

Sense Data and the Argument from Illusion

Many philosophers have thought that one does not perceive what one thinks one perceives; what one is really aware of in perception are mental intermediaries, or sense data. To establish this, it is not necessary to refer to the discrepancy between how things perceptually seem and what science demonstrates to be the case. Rather, it suffices merely to consider that one can be in the subjective state of seeming to see a tomato, for instance, even though there is no tomato in front of one. Perhaps one is dreaming or is the subject of a drug-induced hallucination. If it is possible for one to have non-veridical perceptual experiences that are qualitatively identical to veridical experiences, then the objects that one sees when the experiences are veridical cannot themselves be constitutive of the experience. If they did play such a constitutive role—if one were aware of them—then surely their absence would alter the experience.

This is known as the argument from illusion, which has been used by philosophers to defend the thesis that the real objects of perception (what one *really* perceives) are mental items (or sense data). This view is closely allied to the so-called *causal theory of perception*, acccording to which when one sees a tomato, one has an experience of a certain class of tomato-like sense data that depends causally in the right sort of way on the presence of an actual tomato.

The problem with the argument from illusion is that it takes the fact that one might be unable to tell whether one is hallucinating to be grounds for believing that hallucinatory and veridical perceptual experiences must be qualitatively identical. One may not realize, when one is in a dream, that one is in a dream. This does not entail, however, that there is not in fact a significant qualitative difference between a dream perceptual experience and the corresponding perceptual experience. Without this entailment, the argument from illusion fails.

David Marr and the Computer Model of Mind

A landmark in the recent scientific study of vision is the work of Marr (1982). Vision, according to Marr, is a process of producing a description of the environment on the basis of information contained in the retinal image. The importance of the retinal image is not that it is a *picture,* that is, an object to be contemplated in the mind's eye. Rather, its importance is that it contains information about the environment. The retinal image can be thought of as an array of points, each of which represents the intensity of light at a point in the scene. From information about the distributions of light intensities, together with a cast of assumptions—for example, that sharp discontinuities in the intensity of light are likely to be edges or object boundaries—it is possible to compute a description of the scene before the eyes.

This analysis has important consequences. First, according to it, vision is a computational rather than a biological process; it can be realized, in principle, in a variety of systems, both biological and artificial (see Cognitive Science). Second, once the computational problem of vision has been spelled out, the central business of the theory of vision is the investigation of algorithms (mechanical procedures) for generating the sought-after descriptions on the basis of the available information about point-light intensity arrays. Crucially, this task is autonomous with respect to implementation-level details. Implementation may be important when one turns to the question of whether a theory is psychologically or biologically real, or whether, biologically speaking, vision could have evolved in such a way. Attention to the level of implementation provides, at most, a constraint on the study of vision (at the proper algorithmic level).

The power of this approach is evident: (1) It provides a clear program for research; (2) it ensures a domain of investigation that does not reduce to neuroscience (vision science, like cognitive psychology more generally, is an autonomous special science; see Cognitive Science); (3) it addresses the worry (first articulated by Descartes) that explanation of mental powers by appeal to the mental powers of agencies within the animal is circular—the theory of computation (and the existence of the digital computer) demonstrates that it is possible for a system to perform computations without appeal to an internal homunculus.

Marr's theory directly applies the computational model of mind (or functionalism) to the domain of vision. According to functionalism, *mind* stands to *brain* as a *program* stands to the *hardware* on which it is implemented. Mental states are not identical to the physical states in which they are realized; rather, they are functional or computational roles those states perform.

The general approach to vision pioneered by Marr remains influential. Nevertheless, it has been criticized in philosophically interesting ways.

Criticism from Perceptual Psychology

Marr's theory assumes that vision is the process whereby a detailed internal representation of the environment is produced on the basis of information made available to the system. The content of what one sees is given by these representations. What makes the phenomenon of vision difficult to understand is the fact that the information in the retinal image does not uniquely determine a description of the environment. Consider, for example, that a small object nearby and a large object farther away might project the same pattern of stimulation to the eye. Moreover, the retinal image is highly defective: The eyes are in nearly constant motion, the resolving power of the retina is nonuniform, and there is a gap in retinal photoreceptors (the so-called blind spot). Vision, therefore, must be a process whereby the system compensates for these supposed defects. However, a great deal of work in psychology (especially recent work on change and inattentional blindness) suggests that the content of perceptions may be far less rich than has been supposed. If so, then there is much less computational work for the system to perform.

Given this, it may be that the sense of the presence of a richly detailed scene is an illusion; alternatively, it may be that it is not the case that perceivers really take themselves to have all of the detailed scene in visual consciousness at once. On either possibility, visual theory need not concern itself with the processes whereby a detailed internal representation of the scene is produced, because no such model is produced. Two theoretical commitments have tended to occlude this possibility. First, many scientists have just assumed that perceivers take themselves, when they see, to enjoy a richly detailed, picture-like consciousness of the scene. On the assumption that this perceptual consciousness is not a confabulation, the theory of vision must explain how the experience is generated on the basis of the impoverished stimulus. Second, it is widely believed that vision must give rise to detailed models in the head, for how else can vision play the role it does in guiding action?

Several lines of thought have conspired to shake confidence in each of these propositions. First, as already mentioned, work on change and inattentional blindness calls into question whether vision is really like a snapshot (putting to one side the question of whether ordinary, theoretically innocent perceivers *think* their experiences are snapshot-like) (e.g., Rensink, O'Regan, and Clark 1997). Second, in the last few years a number of authors have suggested that given the sparseness of actual visual content, there is no need to produce a detailed model; it suffices if the perceiver has quick and effective access to environmental information when it is needed (e.g., O'Regan 1992). This strategy has been proposed by philosophers, psychologists, and researchers in artificial intelligence and robotics. Third, some perceptual neuroscientists now believe that there are two functionally and anatomically distinct visual systems (Milner and Goodale 1995). One is responsible for experiences of seeing, whereas the other is responsible for visually guided action. A good deal of visually guided action, according to this line of thought, is independent of perceptual consciousness.

The Criticism from Neuroscience

In the last few years there has been an increase in knowledge of the brain and nervous system. Until recently, the study of the behavior of neurons was confined to invasive single-cell research in animals. Large-scale information about brain structure depended on postmortem examination. In the last few years new technologies have become available that allow for the imaging of neural activity in healthy and intact animals (including humans). New possibilities for learning about the neural basis of cognition and perception have been opened up by positron-emission tomography, functional magnetic resonance imaging, magnetoencephalography, computerized axial tomography, and electroencephalography.

From the standpoint of the neuroscience of perception and cognition, functionalism is viewed with suspicion. Two lines of argument by philosophers have been particularly influential in this area. First, Churchland (1986) has argued, convincingly, that there is no biological reality to the distinction between software and hardware. The brain is a complicated system, to be sure, but *it* is the proper object of investigation for a theory of mind. Second, Searle (1980) has mounted an attack on the idea that minds are or could be suitably programmed computers. Computers perform *syntactic* operations. But mind is an intrinsically *semantic* phenomenon.

Churchland's and Searle's criticisms are important, but they by no means settle the issues. Churchland may be right that the distinction between software and hardware does not correspond to any distinction found in biology. It remains true, nevertheless, that most work in contemporary cognitive and perceptual neuroscience investigates neural processes insofar as they can be understood to realize cognitive function. Actual scientific practice remains very close to what Marr advocated. As for Searle, he is certainly correct that computers are, at best, syntactic machines. It is doubtful, however, that functionalists would be willing to accept that it ought to be possible to derive semantics from syntax. There is no reason for a functionalist to assume that the semantic significance of a computer's (or a brain's) states are determined *only* by internal states of the computer. Semantic content probably depends, as well, on causal history and ongoing interactions with the environment.

The Problem of Consciousness

A more general worry about functionalism is that, like behaviorism, it is unable to account for the subjective, qualitative character of experience (see Behaviorism), which, it is argued, is intrinsic to the experience; but functionalism identifies perceptual quality with extrinsic, functional relations. One of the best-known lines of argument in this area is the so-called inverted spectrum hypothesis. In this view, there is nothing incoherent in the supposition that functionally identical individuals might nevertheless differ in the qualitative character of their color experiences. Perhaps the quality they both call 'red' is experienced by *A* the way the quality they both call 'green' is experienced by *B*. This is a fascinating possibility, which has attracted a great deal of attention. It goes beyond the space allowed here to discuss it fully. If there could be creatures whose color experiences were inverted in this way, then, it would seem, the relevant differences in their color experience would not be accounted for by differences in their functional states. In recent years, not only philosophers, but scientists have explored this question.

Importantly, this criticism of functionalism is frequently advanced by physicalists, who identify qualitative states with brain states or processes. According to physicalism, the qualitative character of experience tracks neural processes, not functional roles. Physicalism has drawbacks, however. First, as of now, no one has even the roughest idea how neural states determine qualitative states. As Crick and Koch (2003) have stated, as of now no one has any idea how the redness of red is produced by action in the brain. This gap in understanding is sometimes called the explanatory gap, in light of which it is hard to see whether functionalism is any worse off than any other approach. Second, the empirical literature may not support physicalism. There is a significant literature on neural plasticity in which neural activity in a given region *changes* its qualitative character, apparently in order to subserve the demands of the larger functional role that neural activity is playing.

Gibson's Attack on Computationalis

Gibson (1979) has criticized Marr's approach on the grounds that it mischaracterizes vision at what Marr calls the computational level. It is just not the case that vision is a process of computing a representation of the scene on the basis of an array of point intensities of light. Neither the retina nor the brain, argues Gibson, is the subject in perception. The subject is the active animal. Moreover, the active animal has access to a *lot* more information than is available on the retina. Gibson suggests that it is a mistake to think of vision as something that unfolds inside the animal. Rather, vision is a kind of activity that the animal engages in. Gibson's "ecological" approach has been very influential, although it is certainly a minority point of view.

Gibson's view has been characterized as a theory of perception at the *personal* (or *animal*) level, whereas Marr's approach is pitched at the *subpersonal* level. Indeed, some writers have suggested that insofar as the views operate at different explanatory levels, they may be compatible. However, this is unlikely. If Gibson is right, then Marr's analysis of vision as the process whereby a description of the environment is produced on the basis of retinal inputs cannot be right.

Gibson's fundamental point—that perception is not a neural phenomenon—has led a number of writers to investigate the significance of a more embodied, dynamic approach to perception and perceptual consciousness.

Some Outstanding Problems

The theory of perception is a fertile area of scientific and philosophical research. Among areas where further work is needed, five are mentioned briefly here.

1. *What are sensory modalities?* There is no accepted account of the individuation of modalities. Are there really only five senses? Could there be others (e.g., artificial systems)? Are perceptual experiences genuinely unimodal, or are they intrinsically multimodal?
2. *Sound and speech.* For those with sight, vision is the dominant sense. Vision shapes perception in other modalities. For example, when one watches television, one *hears* the voice coming from the lips whose movements one sees, even though it does not really. Speech, and the perception of sound more generally, raises a host of important questions, the central of which is, *what does one hear?* The phenomenology and philosophy in this domain is underdeveloped.
3. *Neural correlates of consciousness.* Are neural systems alone sufficient for experience? Most philosophers and scientists are inclined to think so. This assumption has led to a great increase in knowledge of the neural substrates of experience, but to negligible progress on the explanatory gap. Perhaps neural systems are *not* sufficient for experience. Perhaps neural systems are sufficient for experience only insofar as they are properly embodied in the properly situated, active animal. This is an important area for philosophical and empirical collaboration.
4. *Comparative perception.* How does perceptual consciousness vary from species to species? This is an important topic, but one that too many philosophers have tended to neglect.
5. *Is perceptual experience conceptual?* Most philosophers grant that one needs certain concepts to have some experiences. For example, something could not look like a television to an observer unless the observer knew what televisions were. Many philosophers also believe that the content of experience cannot be entirely conceptual. After all, nonhuman animals and infants have perceptual experience but do not have concepts (or so it is supposed). Many cognitive psychologists believe that perception is cognitively impenetrable. For a broad range of cases, how things look is unaffected by belief and desire. To get clear about these matters is a central problem for contemporary work.

Alva Noë

References

Churchland, P. S. (1986), *Neurophilosophy*. Cambridge, MA: MIT Press.

Crick, F., and C. Koch (2003), "*A Framework for Consciousness*," *Nature Neuroscience* 6: 119–126.

Gibson, J. J. (1979), *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lindberg, D. C. (1976), *Theories of Vision from Al-Kindi to Kepler*. Chicago: University of Chicago Press.

Marr, D. (1982), *Vision*. New York: W. H. Freeman and Sons.

McDowell, J. (1994), *Mind and World*. Cambridge, MA: Harvard University Press.

Milner, A. D., and M. A. Goodale (1995), *The Visual Brain in Action*. Oxford: Oxford University Press.

O'Regan, J. K. (1992), "*Solving the 'Real' Mysteries of Visual Perception: The World as an Outside Memory*," *Canadian Journal of Psychology* 46: 461–488.

Peacocke, C. (1983), *Sense and Content*. Oxford: Oxford University Press.

Rensink, R. A., J. K. O'Regan, and J. J. Clark (1997), "*To See or Not to See: The Need for Attention to Perceiver Changes in Sciences*," *Psychological Science* 8: 368–373.

Searle, J. (1980), "*Minds, Brains and Programs*," *Behavioral and Brain Sciences* 3: 417–424.

See also **Behaviorism; Cognitive Science; Consciousness; Empiricism; Phenomenalism; Physicalism; Psychology, Philosophy of; Verificationism**

PHENOMENALISM

In certain disputes that have persisted since Descartes, various philosophers have taken the objects, properties, relations, and facts that are experienced directly, and hence know by acquaintance, to be phenomenal entities (Broad 1951; Russell 1956, 7–26). The sense data of Moore, Russell, and Broad, as well as the mental acts (of perception, thought, etc.) of Brentano, Moore, and Husserl (Broad's modes of cognition and states of mind), were taken to exist. Some philosophers, recognizing phenomena as objects of direct experience, and influenced by Kant's distinction between noumena and phenomena and his notion of a "thing in itself" behind phenomena, thought of physical objects as theoretical entities that were postulated to explain the occurrence as well as the order and coherence of phenomena. Others took physical objects as complexes of phenomena.

Thus, "phenomenalism 1," the view that phenomenal entities, not physical objects, are directly experienced in perceptual situations, led some thinkers to "phenomenalism 2," a form of idealism that construed physical objects as logical constructions of mental entities. Phenomenalism 1 also influenced the development by Ernest Mach (1914) of an antimetaphysical empiricism that was a form of "neutral monism" construing the physical and the psychical as constructions out of simpler neutral sense data, and thus emphasizing the "unity" of science. Carnap (1967) is often read as adopting a linguistic variant of either phenomenalism 1 or phenomenalism 2, since he held that a linguistic schema with an "autopsychological" basis, in which statements about physical objects are transcribed into a language of sense data, was "epistemologically primary" (94). This reflected links among logical empiricism, classical empiricism, and Mach's empiricism.

In the vein of Mach's empiricism, Moore and Russell introduced "unsensed sensa" as (some) sense data held to be objects of direct acquaintance that existed whether or not they were apprehended and hence were neither mental nor dependent on the mind. At times, this led Russell to "phenomenalism 3," a form of physicalistic phenomalism with sense data independent of the mind, in physical space, constituting ordinary physical objects.

The introduction of unsensed sensa was implicit in one early argument by Moore (1903) against phenomenalistic idealism. Hume distinguished such processes as perceiving, remembering, and imagining in terms of characteristics of the experienced datum—force and vivacity—thereby not separating mental acts from objects of such acts. This led the idealists, or phenomenalists 2, to exploit a fourfold ambiguity in terms like 'perception' and 'sensation.' Moore distinguished the following:

- A *mental act* of a certain kind (sensing, imagining, etc.)
- An *object* of such an act, such as the color (shade of) blue or a blue patch
- The generic relation, *consciousness of*, between acts and such objects
- The *fact* that an act stands in this relation to an object

Phrases like 'sensation of blue' tend to confuse these distinctions and lead to the idealist formula *esse est percipi* ("to be is to be perceived"). Preserving the distinctions, however, implies that an object of an act can, logically, exist without being so related to a mental act, and thus rebuts idealism. But this analysis of the perceptual situation led Moore to unsensed sensa, which later troubled him, given sense data such as pain.

Phenomenalism and phenomenal entities have been consistently attacked by contemporary materialists, who reject both the phenomenal objects of direct apprehension and the mental acts of apprehension, and by others who simply reject the notion of direct acquaintance. Sellars (1963) attacked the "myth of the given" and, along with Chisholm (1966), set forth an "adverbial" account of perception, involving states of the perceiving subject ("sensing bluely"), in order to dismiss phenomenal objects. Quine (1953a) argued that sensa are not data of experience but hypothetical entities—"myths, like the gods of Homer"—as are ordinary physical objects and the theoretical objects of physics (44). Such myths are justified not by direct access to them (which subjects do not have) but by the success of theories that invoke them. Yet Quine (1953b) speaks problematically of the "linguistic material" being "tied here and there to experience" (198; Hochberg 1959, 198). Moore claimed that such direct access was unproblematic, since there are mental acts of direct apprehension, as did Brentano, Husserl, Broad, and Sartre, in one form or another (Brentano 1981, 4). Others have defended the recognition of phenomena and mental acts by arguing that they are required to fit the fact that the knowledge of one's own mental states and experiences is quite different from knowledge of the mental states and experiences of others. The materialists' attempts at physicalistic analyses of mental states thus fail in principle. By contrast, Quine's rejection of purported translations of statements about physical objects into a language of sense data—as too complex to be realistic—does not point to a failure in principle and ignores the similar problem faced by his own physicalistic elimination of phenomena (Quine 1953b; Hochberg 1959).

Although Russell sometimes construed physical objects as logical constructions out of phenomena independent of mind, at other times he took the radically different view that macrophysical objects were like the theoretical entities of physics: unknown hypothetical entities providing a causal explanation of known phenomenal data. His hypothetical realism embraced phenomenalism 1 while rejecting phenomenalistic constructions of physical objects. He took physical objects, properties, and relations to be hypothetical causal correlates of directly experienced phenomenal particulars, qualities, and relations, comparing his view to Kant's distinction between noumena and phenomena. According to Russell (1956, 30–34, 47, 86), a physical object was what Kant called a thing in itself, as the cause of sensations, which could be known only by descriptions such as "the physical object which *causes* such-and-such sense-data."

Whitehead's influential technique of "extensive abstraction" led Russell (1914) to shift to construing physical objects as logical constructions—and therefore logical fictions—in *Our Knowledge of the External World,* which, along with Russell and Whitehead's *Principia Mathematica*, greatly influenced Carnap's (1967) constructivism espoused in *Der logische Aufbau der Welt* of 1922–1928. But Russell (1919) returned to speaking of physical objects as the unknown inferred causes of directly experienced sense data, standing in hypothetical relations with the same logical "structure" as the experienced relations of sense data (61). Such hypothetical entities provided a basis for inferences about the structure but not the content (objects, properties, and relations) of the physical world. Thus, given phenomenal entities instantiating phenomenally given spatial relations (e.g., in a visual field), one could hypothesize that there were unknown physical particulars, properties, and relations among them causally correlated with the phenomenal particulars, properties, and relations, such that the physical relations shared logical properties of their phenomenal counterparts: The correlates of "red" and "yellow" stood in a relation that was the correlate of "darker than" and, like the correlate of the phenomenal "left of," shared logical properties (transitivity, asymmetry, etc.) with their correlates—hence the emphasis on structure and the notion of structural realism. Russell (1927) later argued in greater detail for such a view.

Moore's (1953) commonsense realism, as set out in lectures in 1910–1911, contrasted sharply with Russell's concept and with Broad's variant of phenomenalism 1. From the time when Moore (1903) rebutted idealism, he had consistently rejected phenomenalistic reconstructions of physical objects; but in 1910–1911 he accepted a theme of phenomenalism 1: that subjects do not directly apprehend physical objects and therefore cannot directly refer to them. Phenomenal entities are what one directly experiences and directly refers to when one makes assertions like "This is a hand." Moore sought to analyze such a claim so as to acknowledge that subjects do not directly apprehend the hand, while nevertheless preserving certain commonsense truths: that in speaking about a physical object one knows that it exists.

Idealism and Realism

Russell's theory of descriptions provided the key for Moore's analysis. One indirectly refers to a physical object as the x, such that x is a physical object, and sCx, where x is, say, a hand; s is a directly apprehended datum; and the relation C is sometimes taken as "is a manifestation of." One also indirectly apprehends x by directly apprehending s and immediately knowing that there is a physical object.

Broad, soon afterward, in rejecting phenomenalism, took a similar view of direct and indirect apprehension and a relation like C, "is an appearance of," but without claiming to know immediately that the physical object existed. Moore's claiming to know the existence of the indirectly apprehended physical object immediately was crucial to his undercutting of the basic and familiar arguments of the empiricist tradition (from illusion, perspective, conditions of sense organs, etc.) leading to diverse forms of phenomenalism. The theory of descriptions provided a way of denoting objects that were not of direct acquaintance for Moore and Russell; and in Moore's case it also provided a way of claiming that subjects have immediate, not inferential, knowledge of the existence of physical objects. For although a physical object is "indirectly apprehended," one (1) "directly apprehends" the proposition that the physical object standing in *C* to the datum *s* exists and (2) "immediately knows" it to be true. The fact that the physical object is only indirectly apprehended (and thus known by description) does not mean that its existence must be inferred from that of the sense datum. However, as the relation *C* was not directly apprehended, Moore was ultimately forced to denote it also by a definite description, as the relation (satisfying certain conditions) obtaining between *s* and the indirectly apprehended physical object.

The obvious vicious circle may, in part, have led Russell to his hypothetical realism. But Moore, perhaps aware of the problem, sought to buttress his view with two additional, remarkably simple arguments for rejecting phenomenalist-inspired reconstructions of physical objects. These would set the pattern for his defense of commonsense realism.

First, all forms of phenomenalism, as well as Russell's hypothetical realism (inspired by Kant), entailed that Moore did not know that his hand was a physical object. Thus either the arguments for such views were mistaken or Moore's belief about his hand was. But to him it was obviously more likely that the arguments were mistaken than that he did not know his belief to be true. In fact he knew it was a hand, and hence a physical object. Second, to take a physical object to be a bundle of sense data was absurd. No one could seriously believe that his hand was a bundle of sense data.

Analyzing causation, Hume (1967) held that one cannot meaningfully speak of unexperienced material objects causing the occurrence of perceptual objects (104–105; Price 1948). This line of thought also applies to the concept of existence itself. Supposedly, one cannot meaningfully claim that material objects exist because one can make meaningful existential claims only about objects, qualities, and relations that are or can be experienced or can be described in terms of experienced qualities and relations. Moore also attacked this central theme of phenomenalism. He argued that the concept of existence can be sensibly asserted of nonexperienced and even nonexperienceable objects. The concept of being experienced is not and cannot be involved in the analysis of existence, since existence is a simple unanalyzable concept. Thus one may make existential claims about objects that are neither experienced nor experienceable.

In the version of phenomenalism 2 that Moore is attacking here, the realist about physical objects is assumed to make a meaningless or self-contradictory claim. It is assumed that the concept of a material object can be sensibly explicated only in terms of phenomenal concepts, by taking a material object as a coherent bundle of phenomenal entities. Such a phenomenalist thus takes this view to be logically true, in that the claim that there are material objects related to, but not composed of, phenomena is either inconsistent or incoherent, since it either involves a meaningless concept or is self-contradictory. To Moore, such a phenomenalist offers no argument but merely mistakenly assumes that one's having obtained concepts from experience implies that one's application of them is limited to objects that are possible from direct experience. But it was obvious to Moore that existence, spatial relations, and causal connections can be sensibly and truly ascribed to what is not and cannot possibly be directly experienced. Thus one can speak meaningfully of the existence of physical objects, their properties, and their relations without taking such objects, properties, and relations to be analyzable in terms of, or reducible to, phenomenal objects, properties, and relations. There is no reason to deny that one can project concepts originating in direct experience onto objects that one does not and cannot directly experience. Moore proceeded to use some such concepts, along with logical concepts, to explain what he meant by a "material object," which he characterized, in part, as something that was neither an object of direct acquaintance nor a mental act but was extended in a space that (unlike the visual field) was not that of direct experience.

Their divergent forms of realism led Russell and Moore to interpret phenomena and knowledge of physical objects in different ways. Russell was motivated by a desire to make as few hypotheses as possible, in order to lessen the probability of being wrong, and so his realism involved hypotheses about, or logical constructions of, physical objects but not claims of immediate knowledge of their existence. Unlike Moore, Russell did not believe he could refute a Humean solipsist, a phenomenalist in yet another, and extreme, sense. However, Russell (1956, 22) held that there was no reason to believe Humean solipsism and, like Hume himself (1967, 110), that no person could seriously adopt it.

Herbert Hochberg

References

Brentano, Franz (1981), *Sensory and Noetic Consciousness: Psychology from an Empirical Standpoint III*. Edited by Oskar Kraus, translated by Margarete Schättle and Linda McAlister. London: Routledge and Kegan Paul.

Broad, Charlie Dunbar (1951), *The Mind and Its Place in Nature*. London: Routledge and Kegan Paul, 150–159.

Carnap, Rudolf (1967), *The Logical Construction of the World and Pseudoproblems in Philosophy*. Translated by Rolf George. Berkeley and Los Angeles: University of California Press.

Chisholm, Roderick (1966), *Theory of Knowledge*. Englewood Cliffs, NJ: Prentice Hall.

Hochberg, Herbert (1959), "*Of Mind and Myth*," *Methodos* 42: 1–23.

Hume, David (1967), *An Enquiry Concerning Human Understanding*. Cambridge: Hackett.

Mach, Ernest (1914), *The Analysis of Sensations and the Relation of the Physical to the Psychical*. Translated by C. M. Williams. Chicago, IL: Open Court.

Moore, George Edward (1903), "*The Refutation of Idealism*," *Mind* 12: 433–453.

——— (1953), *Some Main Problems of Philosophy*. London: Allen and Unwin.

Price, Henry Habberley (1948), *Hume's Theory of the External World*. Oxford: Clarendon.

Quine, Willard Van (1953a), *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

——— (1953b), "*On Mental Entities*," *Proceedings of the American Academy of Arts and Sciences* 80: 197–204.

Russell, Bertrand Arthur William (1914), *Our Knowledge of the External World*. Chicago, IL: Open Court.

——— (1919), *Introduction to Mathematical Philosophy*. London: Allen and Unwin.

——— (1927), *The Analysis of Matter*. London: Kegan Paul, Trench, Trubner.

——— (1956), *The Problems of Philosophy*. London: Oxford University Press.

Sellars, Wilfrid (1963), *Science, Perception, and Reality*. London: Routledge and Kegan Paul, 140–170, 191.

See also **Carnap, Rudolf; Logical Empiricism; Mach, Ernest; Realism; Russell, Bertrand**

PHILOSOPHY OF PHYSICAL SCIENCES

The physical sciences abound with provocations to philosophy. Here are just a few:

- Einstein's general theory of relativity admits solutions in which observers can, traveling steadily into their own futures, reach their own pasts (see Malament 1984).
- If measurement processes obey the laws of quantum dynamics, the superposition rampant in the micro realm infects the measuring apparatus: Schrödinger's "diabolical device" leaves his cat superposed between life and death. On the textbook understanding of superposition, this creature is neither alive nor dead. Measurements involving more humane apparatus also lack outcomes (see Quantum Mechanics).
- Quantum theories countenance states in which spatially separated systems are correlated in a way that no model of causes acting locally can explain (see van Fraassen 1980). Disturbingly, the larger the domain of the quantum theory in question, the more "typical" such entangled states—which are central to protocols for quantum cryptography and quantum computation become (see Causality; Locality).
- The second law of classical thermodynamics, dictating that entropy always increases, reflects the immediate datum of consciousness, that time is directed. And yet classical thermodynamics is supposed to reduce to statistical mechanics, whose fundamental dynamical laws are time-symmetric (see Irreversibility; Kinetic Theory).

Many entries in this volume (given in the "See Also" list at the end of this essay) chronicle philosophical work on specific physical sciences. This article will offer a necessarily incomplete sketch of the forms of, and reasons for, philosophical engagement with the physical sciences.

The Foundations and Interpretation of Physical Theories

Physicists and mathematicians, as well as philosophers, engage in research on the foundations of physical theories. Foundations research aims to unmask puzzling or noteworthy features of theories (e.g., the failure of certain joint probabilities to be well defined in standard quantum mechanics [Fine 1982]), to identify and to analyze concepts and structures basic to these theories (e.g., the notion of equilibrium in statistical mechanics [Sklar 1993]) and to articulate relations between theories (for example, the ostensive reduction of chemistry to quantum mechanics). Perspicuous formulations of the theories in question aid foundational investigations; disambiguations of vague notions of philosophical currency, which enable the application of these notions to particular physical theories, abet them. The collaboration of philosophical and technical acumen in foundations research can bear fruit: The hierarchy of causal structures in general relativistic space-times (Wald 1984, Ch. 8), once characterized, becomes a powerful tool for advancing questions about determinism and prediction in those settings, once those questions are formulated (see, e.g., Geroch 1977). And relationalism, having been framed as a thesis about particular dynamical theories to the effect that those theories can be fully formulated in phase spaces of specifiable sorts, becomes a target for creative physics (see Belot 2000).

A related aim of foundations research is to develop, catalog, and evaluate interpretations of physical theories. To interpret a physical theory is to characterize the set of worlds possible according to it, to identify and describe the physical systems of which that theory is true. Thus an interpreter of general relativity must decide whether solutions of Einstein's field equations that are identical up to diffeormorphism correspond to one possible world or many (the second option sets the stage for a general relativistic reprisal of the Liebniz shift argument against substantival space). And an interpreter of quantum mechanics must decide whether or not to take that theory to encapsulate statistical generalizations about ensembles of worlds individually and precisely described by the hidden variables of Bohmian mechanics (an interpretive option made available by an effort of creative physics).

Foundational work galvanizes interpretive attention: An interpreter, piqued by foundational investigations, can ask, Of what manner of world is a theory with singularities (without joint probabilities/with dynamical symmetries/with entanglement/with closed timelike loops/with *both* time symmetric fundamental laws *and* steady entropic increase) true? Interpretations of physical theories can reflect epistemological and metaphysical prejudices. But it is unjust to conceive the project of interpretation as one of plastering such prejudices on a solid frame of self-sufficient science, for the impetus to interpret a theory often lies within that theory, in the form of an unresolved problem, an incomplete concept, or an apparent incompatibility with other well-established theories or with common sense.

Although an interpretation of a theory is what a realist believes about it, interpretation holds interest for those who are not confirmed realists. An interpretation enables the constructive empiricist to see the position from which he is virtuously withholding belief (see Empiricism; Scientific Realism). Interpretation helps those contemplating realism about several theories at once to assess whether and on what terms those theories can be true together. Constituting the scientific image, interpretation forms part of the principal task of philosophy, according to Sellars (1963): to combine that image "stereoscopically" with the "manifest image" presented to us by educated ordinary experience. Even working physicists can have a stake in interpretive questions, insofar as their answers suggest directions in which to develop new theories: Some forms of the project of quantizing gravity presuppose a decision about which observables of the classical theory are genuine (see Belot and Earman 1999). Interpretation serves a variety of purposes; the grounds and significance of commitment to particular interpretations vary concordantly and are themselves worthy topics of philosophical investigation.

General Philosophy of Science

Philosophers of science have traditionally sought general accounts of the structure of scientific theories and the nature of scientific explanation, confirmation, intertheoretic reduction, and the like. For such accounts, the history and present practice of the physical sciences form a wide and varied proving ground. In some cases, examples from the physical sciences serve to enliven points discernible in abstraction—the problem of old evidence for Bayesian confirmation theory (see Confirmation Theory) is latent in the statement of Bayes' theorem; it becomes violent when the old evidence in question is the anomalous precession of the perihelion of Mercury. In other cases, bringing general accounts to bear on particular physical theories is revelatory. Consider Reichenbach's principle of the common cause (Reichenbach 1971) as it might be incorporated into a causal-mechanical account of explanation. On the Reichenbachian account, correlated events require explanation in terms of a cause acting in their common past (see Causality). Made precise to apply to certain quantum correlations, the account implies that those correlations have no explanation, because any theory fostering a common-causal explanation of those correlations makes predictions different from the empirically validated predictions of quantum theory (see van Fraassen 1980, 25–31). By contrast, the correlations in question have an explanation in the deductive-nomological sense (see Scientific Explanation): They are a direct consequence of applying the Born rule to an entangled quantum state. Close attention to quantum theory reveals not only (what was evident at the outset) that Hempel's account of explanation and this Reichenbachian one are rivals, but also that the choice between them involves at least taking a stand on whether quantum correlations are explicable. (This same attention exposes as untenable any interpretation of quantum mechanics that asserts the theory to hold of a world of causes acting locally, if these causes adhere to Reichenbach's model.) Thus the philosopher of quantum mechanics performs for the philosopher of scientific explanation what Reichenbach (1938) identifies as the "advisory" task the competent epistemologist performs for the positive scientist (8–16). This is the task of explicating the decisions (not obviously) entailed by the different options.

In connection with philosophy's advisory role, it is interesting to observe that practicing physicists often take stands, implicit or explicit, on issues in the general philosophy of science. Consider the "horizon problem" (Earman 1995, Ch. 5) of contemporary cosmology: In standard Big Bang models, regions of the universe between which correlations (e.g., in their blackbody spectra) are observed that have no common past, so that no causal mechanism, acting locally, could have established those correlations. Whether these correlations therefore lack explanation depends on one's account of explanation. In the Reichenbachian account of the last paragraph, the correlations cannot be explained, and the horizon problem is genuine; in other accounts, the correlations are explained, and the horizon problem is illusory. By and large, practicing cosmologists have stalwartly Reichenbachian intuitions: They hail inflationary cosmologies for solving the horizon problem, and devote ongoing work to constructing explanations for features—for example, fine tuning (see Anthropic Principle)—that do not (according to many philosophically respectable views of explanation) require explanation. The lesson from this is that philosophers should refrain from evangelism: Many a fruitful scientific research program grows from seeds of suspect philosophic pedigree.

The scientific-realism debate furnishes another illustration of mutual constraint exerted between theses in the general philosophy of science and interpretations of particular theories. One question is: Do grounds, adduced in lofty abstraction, for realism about successful scientific theories support realism about particular successful physical theories? Here quantum mechanics provides a cautionary tale. Fantastically successful, it appears the archetype of a theory for which might be run an "explanationist" defense of scientific realism—an abductive inference that the theory is true because its truth is the best explanation of its empirical success. Suppose someone were convinced by such an argument to be a realist about quantum mechanics. What would that person believe? One needs an interpretation of quantum mechanics to give content to one's realism, but none are unexceptional, and some undermine one's grounds for realism. For instance, the standard collapse interpretation solves the measurement problem by suspending quantum dynamics for the duration of measurement (see Quantum Measurement Problem). On this view, the truth of quantum mechanics hardly explains its empirical adequacy, because the theory must break down for measurements to occur at all.

Physics Centrism

It is sometimes complained, usually against caricatured "positivists," that enthroning physics as the paradigm of science has led to a partial and antiseptic dominant philosophy of science (Harding 1991). One form of the complaint is: If a general philosophy of science takes as its criterion of adequacy only the capacity to deal successfully with physical sciences, then that philosophy exaggerates the role of theoretical structure, particularly mathematical structure (and the variety of precision accompanying it), in sciences. Whatever other justice this complaint might have, it underestimates the extent to which physical theories themselves fail to conform to "positivist" models of physics. For example, the question of the circumstances under which the behavior of some particle physics apparatus counts as evidence, or constitutes a phenomenon to be saved by physical theory, is a highly nontrivial one. Its resolution draws upon interactions between models and simulations of many sorts, as well as vast collaborations of theorists, experimentalists, technicians, and their machinery (Galison 1987) (see Particle Physics). This might be taken to suggest, *pace* the "positivist," that epistemology should extend its scope beyond the "context of justification" to include, at least, the context of evidence generation as well. It might also be taken to suggest that there are epistemological questions to which theory is not central, a suggestion reinforced by consideration of chemistry and astronomy (see Chemistry, Philosophy of; Astronomy, Philosophy of), where skill and observation call for epistemological analyses not directly supplied by traditional epistemologies of science.

Metaphysics

Metaphysics aims to characterize the possible. Physical sciences can be interpreted to characterize possibilities within their domains of applicability. Physical sciences can thereby serve both to stimulate and to check metaphysics. Interpreted physical sciences supplement the reservoir of possibilities (traditionally constituted by the metaphysician's meager imagination) worth taking seriously. The possibilities evinced by interpreted physical theories include challenges to metaphysical principles: Einstein's field equations can be interpreted to describe consistent time travel scenarios, and quantum mechanics can be interpreted to hold of nexus of events temporally but not causally ordered. What is more, the situations described by interpreted physical sciences have real complexity and antecedent interest. They are thus splendid occasions to flex metaphysical muscle, as a mereologist might in attempting to account for how the properties of a compound relate to the properties of the elements it comprises (see Chemistry, Philosophy of; Emergence). This is not to suggest that interpreted physical sciences serve as neutral data grounding conclusive tests of metaphysical principles. Interpretations can reflect metaphysical as well as epistemological predilections. Thus the clash of an interpreted physical theory with a favored metaphysical principle can be cited as reason for rejecting that interpretation, that theory, or both. Just as with theses in the general philosophy of science, the point of bringing metaphysics into contact with the philosophy of the physical sciences is not to settle disputes in either discipline, but to transform them, by unraveling (and yanking on) the conceptual ties that bind one sort of inquiry to another.

LAURA RUETSCHE

References

Belot, Gordon (2000), "Geometry and Motion," in Peter Clark and Katherine Hawley (eds.), *Philosophy of Science Today*. Oxford: Oxford University Press.

Belot, G., and John Earman (1999), "From Metaphysics to Physics," in Jeremy Butterfield and Constantine Pagonis (eds.), *From Physics to Philosophy*. Cambridge: Cambridge University Press, 166–186.

Earman, John (1995), *Bangs, Crunches, Whimpers, and Shrieks*. Oxford: Oxford University Press.

Fine, Arthur (1982), "Hidden Variables, Joint Probabilities, and the Bell Inequalities," *Physical Review Letters* 48: 291–294.

Galison, Peter (1987), *How Experiments End*. Chicago: University of Chicago Press.

Geroch, Robert (1977), "Prediction in General Relativity," in John Earman, Clark Glymour, and John Stachel (eds.), *Foundations of Spacetime Theories* (*Minnesota Studies in the Philosophy of Science*, Vol. XIII). Minneapolis: University of Minnesota Press, 81–93.

Harding, Sandra (1991), *Whose Science? Whose Knowledge?* Ithaca, NY: Cornell University Press.

Malament, David (1984), "'Time Travel' in the Gödel Universe," in *PSA Proceedings of the Biennial Meeting of the Philosophy of Science Association* [PSA] *1984, Volume Two: Symposia and Invited Papers*. East Lansing, MI: PSA, 91–100.

Reichenbach, Hans (1971), *The Direction of Time*. Berkeley and LA: University of California Press.

——— (1938), *Experience and Prediction*. Chicago: University of Chicago Press.

Sellars, Wilfrid (1963), "Philosophy and the Scientific Image of Man," in *Science, Perception, and Reality*. Atascadero, CA: Ridgeview Publishing, 1–40.

Sklar, Lawrence (1993), *Physic and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

van Fraassen, Bas (1980), *The Scientific Image*. Oxford: Oxford University Press, 1980.

——— (1989), *Quantum Mechanics: An Empiricist View*. Oxford: Oxford University Press.

Wald, Robert (1984), *General Relativity*. Chicago: University of Chicago Press.

See also **Anthropic Principle; Astronomy, Philosophy of; Causality; Chemistry, Philosophy of; Classical Mechanics; Complementarity; Conventionalism; Determinism; Experiment; Explanation; Hilbert, David; Irreversibility; Kinetic Theory; Laws of Nature; Locality; Observation; Particle Physics; Physicalism; Prediction; Reichenbach, Hans; Quantum Field Theory; Quantum Logic; Quantum Measurement Problem; Quantum Mechanics; Reductionism; Scientific Models; Scientific Realism; Space-Time; Statistical Mechanics; Theories; Time; von Neumann, John**

PHYSICAL NECESSITY

See **Laws of Nature**

PHYSICALISM

If any grand metaphysical thesis can lay claim to the title of a received view in contemporary philosophy, it is physicalism—roughly stated, the claim that everything is ultimately physical. The view is most familiar from the philosophy of mind, where the position frequently known as *materialism* is set out as one option available on the mind–body problem. The view is not confined to the philosophy of mind, however; its influence is quite widespread. One can find projects motivated by physicalism in the metaphysics of color, the philosophy of biology, meta-ethics, and so on. While the terms 'materialism' and 'physicalism' are often used interchangeably, the latter term is more likely to be used in referring to the more general thesis that has this widespread influence. The term in fact has a curious history, as it was introduced into the analytic tradition by the later work of the logical positivists. They were, of course, concerned to eliminate anything recognizable as metaphysics, while the term "physicalism" now denotes an unmistakably metaphysical view.

As used by the positivists, physicalism was a position regarding the sorts of sentences that should be used as the touchstone of the process of verification. If, as the positivists thought, the cognitive significance of a sentence is determined by the ways in which it may be empirically verified, it is important to determine what sort of "observation sentence" may aptly describe those events of verification. While earlier positivists tended to think that the relevant observation sentences were reports solely of one's own immediate experience, this approach eventually came to seem untenable. Led by Otto Neurath, the later positivists proposed physicalism as the view that the key verifying sentences are those describing events that are intersubjectively verifiable. The contrast between physical and mental or "phenomenal" was here understood primarily in epistemic terms: The mental is that which is essentially private, verifiable only by one person, and the physical is that which is locatable in the broader spatiotemporal framework, available for inspection by more than one observer. Unlike contemporary physicalism, which is usually put forward as a metaphysical view motivated in some way by the empirical success of physical science, the positivists' physicalism had nothing especially to do with *physics,* but was motivated, rather, by more general considerations about the nature of verification and the communicability of meaning (Uebel 1992).

The shift in the usage of 'physicalism' from the logical empiricists' sense to that found in contemporary philosophy is nicely illustrated in Herbert Feigl's (1968) long essay *The "Mental" and the "Physical,"* first published in 1958, which introduces two senses of physicality. By 'physical$_1$' Feigl means that which is needed to account for the spatiotemporal order and is intended to capture what is intersubjectively available. By 'physical$_2$' he means that which is needed to explain inorganic processes. This latter sense comes closest to what has been operative in the discussions of physicalism as a metaphysical thesis. The claim made by the contemporary physicalist is not a semantic claim about the role of reports of what is intersubjectively available; it is a claim about the sorts of entities discerned in the study of "mere matter," *viz.,* that those entities provide the ultimate building blocks of everything there is. The unqualified term 'physicalism' will hereafter in this essay denote the contemporary metaphysical view (not that of the positivists) summed up in the slogan "Everything is ultimately physical."

Two Questions About Physicalism

Two main questions about physicalism may be spotlighted. How exactly is the thesis to be *formulated?* And how might it be *justified?*

Consider first the slogan "Everything is ultimately physical." It contains a deliberate hedge. A simpler statement would be "Everything is physical," and the claim would then be that every entity in the appropriate domain is a physical entity. The hedge is needed, however, to accommodate the fact that there are recognizably physicalist positions that allow for entities that are, strictly speaking, not physical but that bear some appropriate relationship to the physical. More precisely, such positions allow for entities the existence and character of which are determined and explainable by the physical in such a way that it is appropriate to say that the nonphysical entities are not entirely distinct from physical ones, or that they are nothing over and above the physical entities involved in that determination.

To clarify the slogan "Everything is ultimately physical," one should answer three sets of questions:

1. What does 'everything' range over? What sorts of entities are supposed to be ultimately physical?
2. What does it mean to classify one of those entities as physical in the first place? In that domain of entities, how is the special class of physical entities to be picked out?
3. What exactly does the physicalist want to say about the relation of the entirety of that domain to that special class of physical entities? Is the claim simply that everything in the intended domain is itself a member of that special class? If not, how are the nonphysical entities related to the physical ones?

The first set of questions has occasioned the least controversy. Most physicalists agree that the doctrine must be at least strong enough to imply that *the way the world is generally* is determined by *the way the world is physically*. Hence, the thesis must be strong enough to make a claim about properties (and relations) and their distribution in the world. It is for this reason that the idea that the doctrine might be formulated as the claim that all events are physical events, which enjoyed a run of popularity in the wake of Davidson (1970), has lost its appeal. Unless events are understood as themselves nothing more than property exemplifications, that claim seems not to imply anything significant about the distribution of physical and other properties. In light of this, the focus in this article will be just on claims about properties. The next section will concern the controversies attending the second set of questions about the meaning of the physical, followed by a discussion of the third set of questions about the relation that *all* properties might be said to bear to physical properties.

While the question of formulation has received the most attention in the literature, the question of justification is no less important. One may distinguish two justification projects here. There is, first, the project of defending physicalism against apparent counterevidence. Consciousness, intentionality, ethical properties, biological functions—these are all phenomena that have been thought incompatible with physicalism. There is no question that philosophers have devoted considerable energy to this defensive project. Discussion of these defenses of physicalism is beyond the scope of this article (see Consciousness; Function; Intentionality). The other project is to provide some positive reason to adopt physicalism in the first place. The most important arguments of this sort range from the (alleged) history of successful physicalist theorizing to various causal and counterfactual arguments. The last three sections of this article will provide a survey of these arguments and the challenges they face.

Specifying the Physical

What is meant by classifying a property as physical? One may be tempted to appeal in some way to the traditional definition of matter as that which is extended in space, perhaps by defining a physical property as one the instantiation of which confers spatial extension on its bearer. But this approach will not engage with contemporary physicalism, which gives pride of place to the science of physics, not spatial extension. (This point is reflected in the shift, earlier noted, from Feigl's physical_1 to physical_2.) It is worth bearing in mind, too, that it is not at all clear that such a criterion will coincide with one that appeals to physical theory. The entities to which physical theory is committed cannot be counted as straightforward material in the sense of having spatial extension; forces and fields, for instance, do not seem to have spatial extension. If one takes into account, further, the picture of space found in quantum mechanics, it becomes even less clear how to think about any spatial extension criterion. For these reasons, most physicalists are inclined to appeal in some way to physical theory without invoking spatial extension at all.

Physics and the Physical

How exactly should the physicalist appeal to the entities at issue in physical theory? One approach is simply to consult the details of the actual theory and define the physical in those terms—perhaps by a simple list. Of course, there is a question about which physical theory exactly is at issue here. Is physical science to include not only microphysics but all disciplines that might be thought of as physical, such as chemistry, astronomy, geology, and so forth? Though not all physicalists have been explicit about this, most seem to opt for the more restricted theory. There are two good reasons to take this option. First, the properties at issue in the other physical sciences seem likely to be appropriately nothing over and above the more basic properties delineated by microphysical theory, so including them in the domain of the properly physical is unnecessary (see Chemistry, Philosophy of; Emergence; Reductionism). Second, and more important, the justification of physicalism that appeals to causal considerations makes the most sense when the causal considerations make reference specifically to the causal completeness of physics, which implies that the physicalist should focus solely on microphysics (Sturgeon 1998).

If it is agreed that the physical properties should be limited in this way to the microphysical, one must bear in mind the fact that much discussion of physicalism has been conducted as if physicality included a broader range of properties. For instance, Smart's (1959) celebrated defense of the identity theory focused on the identity of mental properties with *neurophysiological* properties, which would not count as properly physical on the present suggestion. Nor is it plausible to suppose that neurophysiological properties can be identified with strictly physical properties. What might be plausible, instead, is the thesis that such physical properties can be defined in terms of aggregations of basic individuals having certain physical properties and standing in certain physical relations to each other (see Kim 1998 for a discussion of such "micro-based" properties). In light of this, certain philosophical discussions about physical properties might best be interpreted as being about properties that are either physical or defined in this way in terms of the physical.

Skeptical Worries

The idea that the physical should be defined by reference to physics has prompted a number of philosophers to argue that no adequate account of the physical can be given (Crane and Mellor 1990). The core argument can be given as a dilemma. The theory by reference to which the physical is to be defined is either the physical theory that actually exists in the present or some envisaged future or ideal theory. In light of the history of science, it is likely that present theories are false, even if they are in some sense closer to the truth than past theories. As a result, the physicalist who takes the former option will be defining physicalism by reference to a false theory, which seems an unhappy result. On the other hand, if the physicalist chooses the latter option, the doctrine will be defined by reference to an unformulated theory, which seems again to be undesirable. In that case, it is said, the relevant notion of the physical becomes objectionably obscure, or perhaps is such as to render the thesis of physicalism trivially true.

Neither horn of the dilemma is quite as sharp as it may first appear. Consider the first option. Even if a physical property is defined by reference to current physical theory, physicalism is not committed to the *truth* of that physical theory. That theory is being exploited only for its ontology, not for its doctrine. So long as it has the right inventory of properties, regardless of the claims it makes about them, this actual theory will not be problematic. Still, if actual theory is in error in including nonexistent properties, or in failing to include certain properties, those errors seem to count against this option.

Consider the second option, whereby a physical property is defined by reference to an ideal physical theory. Two objections should be distinguished here: the triviality objection and the obscurity objection. The triviality objection may be understood as follows. In what sense is the ideal physical theory "ideal"? One may suspect that its being ideal implies that it is successful as a theory about all those phenomena that the physicalist believes to be ultimately physical. In that case, however, all those phenomena are trivially counted as physical simply by virtue of being the subject matter of the ideal physical theory. The point can be made dramatic by considering some properties that are intuitively not physical at all, such as mental properties. If the ideal physical theory is defined as one that includes an account of everything of interest, including the mental, then such mental properties trivially count as physical—an intolerable result.

But this result is hardly inevitable. The physicalist who opts to define physicality by reference to an ideal physical theory need not understand the theory as ideal in the sense that it is successful about all phenomena of interest to the physicalist. Rather, the ideal physical theory may be better understood as ideal in the more limited sense of succeeding as a theory about all phenomena of interest to *physicists.* Of course, there remains work to do by way of clarifying just what this more limited domain is to include, but the present point is simply that the physicalist is not required to define the relevant theory in a way that trivializes physicalism.

The other objection from the obscurity point of view is more serious. An appeal to some ideal physical theory is an appeal to a theory not actually formulated. But if the details of this theory are in fact unavailable, how can anyone even begin to evaluate claims like "Everything is ultimately physical"? If philosophers do not have the ideal physical theory in hand, do they have any grasp of physicalism itself? If philosophers are profitably to discuss the doctrine, they need to have at least some grasp of what counts as prima facie physicality and what does not; the objection is that an appeal to an ideal physical theory does not allow philosophers to have any such grasp.

Responses to Skepticism

Responses to these skeptical worries can be divided into three categories: those that embrace the first option of appealing to present physical theory, those that embrace the second option of appealing to some ideal physical theory, and those that attempt to avoid the choice altogether.

The first response has been defended by Melnyk (2003), who argues that physicalism is itself a scientific hypothesis and, hence, that the attitude the physicalist takes toward it should be comparable to that which scientific realists take toward what they consider to be the best of current scientific theories (see Scientific Realism). As scientific realists usually recognize that current theories are likely to be false, the attitude in question need not be one that is rendered irrational by that recognition. In Melnyk's account of the appropriate attitude, one assigns the endorsed theory a higher probability than any of its *relevant rivals*, where the class of relevant rivals is limited in various ways—for instance, only actually formulated theories are included. As a result, the physicalist can take this attitude toward current physical theory without being committed to its truth or even likely truth.

The second response is perhaps the most popular (Papineau 1993; Poland 1994). The complaint that the resulting notion of physicality is too obscure can be blunted if there is good reason to suppose that certain concepts will not be needed to understand the ideal physical theory. If, for instance, no concepts of the mental are needed for this purpose, then the mental is not trivially included as part of the physical, and it is a good and substantive question whether the mental can be located in a fundamentally physical world. (This strategy may be generalized, of course, for other phenomena typically seen as needing special accommodation: One may suppose explicitly that no biological concepts, for instance, will be needed to understand the ideal physical theory, and so on.)

One important question for this strategy concerns the nature of this confidence that certain entities will not appear as such in the ideal physical theory: Is this exclusion to be imposed by definitional fiat, or is it to be understood as an empirically grounded prediction about the likely character of the ideal physical theory? Even if the former were chosen, a commitment would presumably be required to an empirical claim about some sort of continuity between actual physical theory and the envisaged ideal, for a lack of such continuity would make unclear how the success of actual physical theory plays a role in motivating the physicalism thus defined.

Finally, a different response that has recently been aired is to *replace* talk of the physical with that of the "nonmental" or other negatively defined notions. It can be seen as a natural descendant of one version of the previous strategy, whereby one defines the ideal physical theory so as to guarantee that certain phenomena are excluded. The suggestion is to drop the term 'physical' altogether and formulate a thesis negatively—as, say, the claim that everything is ultimately *nonmental.* Physicalism itself, on this suggestion, would presumably splinter into a variety of theses of this sort: In addition to the nonmental thesis, there would be the claims that everything is ultimately nonbiological, nonaesthetic, and so on.

In evaluating these responses to skeptical worries, one should bear in mind that an appropriate response should maintain contact with the motivation for physicalism. If physicalism is supposed to be motivated by something about the actual success of physical theory, that motivation should continue to make sense given the chosen interpretation of (or, as in the third option, replacement for) the physical. In light of this, it is worth noting that it is not entirely clear that even present physical theory avoids incorporating the mental as an essential part of its machinery. If, as some speculate, the best way to make sense of quantum mechanics requires the essential involvement of a mental event of observation, then actual, non-ideal physical theory itself includes the mental as such, and an attempt to define the physical by reference to that theory will run the risk of counting the mental as such as already physical.

The following sections explore the relation of all else to the physical.

Identity Theses

The most straightforward way of formulating physicalism is a property identity claim:

(PI) Every property is a physical propery.

To many, PI has seemed too strong to be a fair expression of physicalism. If physical properties are understood as limited to those that are *microphysical*, this thesis will be simply incredible. This consideration is not, however, responsible for convincing many to reject PI as too strong, as much discussion of PI and its problems has taken place without having that specific method of defining the physical in the background.

The more famous consideration regarding PI is that it seems to be inconsistent with the claim that some properties are *multiply realizable,* which does not seem to be inconsistent with physicalism itself. The multiple realizability claim may be set out thus:

(MR) There is at least one property *F* such that it is possible for *F* to be instantiated on different occasions by virtue of being realized by different physical properties.

Exactly how the relation of realization is to be understood is a good question (addressed below under "Realizationism"). It may be enough for the moment to point out that realization is meant by physicalists to be a sort of determination relation that licenses the claim that the realized property is not entirely distinct from its realizers, so that one could accept MR while remaining a physicalist.

MR seems to imply the following corollary, which one may call the *lack of coextension thesis:*

(LC) There is at least one property *F* such that there is no physical property with which *F* is necessarily coextensive.

Given MR, it seems that several different physical properties are all sufficient for the instantiation of the multiply realizable property; as a result, no single physical property is necessary for it. LC, of course, is inconsistent with PI; so if MR implies LC, it is inconsistent with PI.

The option of formulating physicalism in terms of identity is not thereby rendered hopeless, however. There are three positions worth considering that may be appealing to those originally tempted to PI. The first (the disjunction option) retains MR but maintains that MR does not imply LC; the second (the quasi-eliminativist option) rejects MR but retains a thesis similar to it; the third (the trope identity option) retains MR and LC but rejects PI in favor of a similar thesis.

The Disjunction Option

The disjunction option rejects the claim that MR implies LC; it aims to ensure that a physical property necessarily coextensive with the multiply realizable property can be found by constructing one from the various possible physical realizers of it. Suppose that pain is multiply realizable and that R_1, R_2, ... R_n are all the possible physical realizers of pain. Now consider the open sentence 'x has R_1 or x has R_2 or ... x has R_n' and call this the *disjunctive predicate.* The disjunction option identifies pain with the property allegedly expressed by the disjunctive predicate. There are three key questions about this option.

The first question is: Does the maneuver succeed in specifying a property that is necessarily coextensive with pain? The disjunctive predicate is constructed out of names of the physical realizers of pain, but if there are possible *nonphysical* realizers of pain, then the disjunctive predicate is not in fact necessarily coextensive with pain. However, if the nonphysical realizers are never instantiated in the actual world, it is not clear that physicalism runs counter to such possible nonphysical realization.

The second question is more fundamental: Does the disjunctive predicate succeed even in expressing a genuine property? If, as some argue, there is nothing genuinely in common among the various individuals that form the extension of the disjunctive predicate, then one might conclude that there is no property corresponding to the predicate. It is not obvious that one should suppose there is nothing in common, however, in light of the fact that each individual in the extension has in fact been grouped together (at least by humans) as an exemplifier of pain.

Finally, the third question concerns the classification of the alleged property. Supposing that the disjunctive predicate indeed expresses a genuine property, it is not obvious that this property will count as a *physical* property. If the disjunction option is to save PI, the property with which it identifies pain has to be, of course, a physical property. The fact that the disjunctive predicate is constructed out of the names of physical properties is no guarantee that the resulting construction captures a physical property. The predicate '*x* does not have mass' is constructed from the name of a physical property, but it is doubtful that the property it captures should count as physical.

The Quasi-Eliminativist Option

The quasi-eliminativist option rejects MR but replaces it with a similar-looking thesis about predicates:

(MR*) There is at least one predicate *F* such that it is possible for *F* to be satisfied on different occasions by virtue of different physical properties being instantiated.

Suppose again that pain is one of the allegedly multiply realizable properties. The quasi-eliminativist does not insist that pain is realized in only one way; he allows that different things can be in pain by virtue of exemplifying distinct physical properties, but he rejects the claim that there is a genuine property of being in pain. There is a generally applicable predicate 'is in pain,' but there is no single property shared by all the individuals in the extension of that predicate. In effect, the quasi-eliminativist takes up the skepticism earlier expressed toward the alleged property expressed by the disjunctive predicate and applies it to the multiply realized property itself. The multiply realizable property is eliminated while the realizing properties are retained.

In fairness, the position is not described *simply* as eliminativist, since it allows that statements using the 'is in pain' predicate can be true. Indeed, one might adopt a version of this position according to which the various physical realizers of pain are themselves described as *kinds* of pain, thus allowing one to say that certain kinds of pain are genuine properties.

The quasi-eliminativist option has the undoubted attraction of providing a simple formulation of physicalism while not denying what seems to be the real substance of multiple realizability. It is hardly without cost, however, since the claim that there is no genuine property of being in pain is not very plausible at the outset. Perhaps, however, the advocate of this option could argue that insofar as one finds it plausible to think of a given alleged property as multiply realizable, one will also find it plausible to think that the property is merely alleged. In other words, if it is not plausible to think that there is no genuine property of being in pain, this may be because it is not plausible to think that pain is multiply realized.

The Trope Identity Option

The trope identity option abandons PI in favor of the thesis that every *trope* is a physical trope. This thesis is distinct from PI only if it is coherent to suppose that an individual exemplifies two distinct properties—the instantiations of which are nonetheless identical. If one understands tropes as being derivative on the more fundamental existence of properties as universals, so that they inherit their individuation conditions from the properties of which they are instances, it is not clear that this supposition is coherent. If, however, properties as universals are themselves constructed out of tropes understood as more fundamental entities, the supposition may be coherent. So long as there are at least two distinct equivalence relations by which tropes might be grouped into properties, the same trope could belong to two different groups and therefore be counted as an instantiation of two distinct properties.

The trope identity thesis can accommodate multiple realizability in a straightforward fashion. If a property *F* is multiply realized, no physical property is necessarily coextensive with *F;* nonetheless, each instance of *F* might be identical with an instance of some physical property—perhaps the physical property that realizes *F* on that occasion.

Supervenience Theses

While there are various ways to maintain an identity thesis in the face of MR, it may seem simpler to abandon identity for a rather different approach. The notion of supervenience has seemed to many to be a promising alternative. Supervenience is a relation that holds between families of properties: The *A*-properties supervene on the *B*-properties just in case there can be no difference in *A*-properties without some difference in *B*-properties (see Supervenience).

A good way to understand the attraction of supervenience is to think of it as an appropriately minimal way of talking about sufficiency. If A-properties supervene on B-properties, then there is a total state of each individual with regard to its B-properties that suffices for its total state with regard to its A-properties. Such sufficiency is just what one would expect if the A-properties were to be explained as being nothing over and above the B-properties. By appealing to supervenience, however, one avoids having to say just *which* of the explanatory properties do the explaining. Further, one allows that different physical properties do the explaining in different circumstances, thus accommodating multiple realizability.

The claim that the nonphysical properties supervene on the physical properties admits of more than one interpretation, however, and much of the literature on supervenience can be seen as devoted to distinguishing those interpretations and assessing them as expressions of physicalism. When it is said that there can be no difference in A-properties without a difference in B-properties, there are three key parameters to be specified. First, of course, one must specify the families A and B. Second, one must specify the modal status of the claim. Third, one must specify how the relevant comparisons are to be made. When it is said that a difference in A-properties requires a difference in B-properties, what sorts of objects are to be compared for sameness and difference, and how exactly are the relevant pairs to be selected? Permutations of these parameters can result in a wide variety of distinct supervenience claims. (See Supervenience for an overview of the varieties and their significance.)

The sort of supervenience thesis that has gained widespread popularity is global in form, comparing entire possible worlds for sameness or difference in the relevant families of properties. A *global supervenience* thesis seems well placed to capture the idea:

(GS) For any two possible worlds W_1 and W_2, if W_1 and W_2 are physically exactly alike, then W_1 and W_2 are exactly alike in all respects.

This is in accord with the notion mentioned earlier as central to physicalism, that *the way the world is generally* is determined by *the way the world is physically*. Nonetheless, GS is rejected by most physicalists as too strong. Most think of the doctrine as a contingent truth, so that there are possible worlds in which it is false. Given this, it seems that GS is false. Consider a possible world in which traditional Cartesian dualism is true and in which a particular body B is associated with a Cartesian soul. Now consider another possible world in which dualism is true just as in the first, with the sole exception that body B is, in this world, unconnected to any mental substance. This pair of worlds is a counterexample to GS, yet the physicalist may want to allow them as possible.

In order to focus on the character of the actual world, the physicalist is better off with a global supervenience thesis that compares other possible worlds with this actual world. If any world is just like this one in its physical character, then it must be just like this one generally. The desired thesis can be made more precise with the notion of a "minimal physical duplicate" (Jackson 1998). A minimal physical duplicate of a world W is, intuitively, what one would get if one were to take the physical description of W as a recipe, building a new world out of nothing but the ingredients spelled out in it, with no additional individuals or properties. Now consider the following supervenience thesis:

(MPD) Any minimal physical duplicate of the actual world is indiscernible from the actual world generally.

MPD appears to specify a supervenience thesis that is at least necessary for physicalism to be true. Whether it is *sufficient* is another question. A consensus appears to be emerging that *no* supervenience thesis is sufficient for physicalism. Supervenience is in itself simply a sufficiency claim. But physicalism is committed to an explanatory claim: If there are any nonphysical properties, their instantiation is to be explained by reference to the distribution of physical properties. Since it is a familiar idea from philosophy of science that sufficiency is not itself sufficient for explanation, supervenience does not seem capable of guaranteeing the explanatory import of physicalism.

This point has been emphasized by several philosophers in recent years (Horgan 1993; McLaughlin 1995). One way of making the point vivid is to demonstrate that there may be explanations of the truth of supervenience theses that are inconsistent with the sort of explanation desired by physicalists. Consider, for instance, how a traditional property dualist might explain the fact that any two physically indiscernible individuals are indiscernible with respect to their mental properties; the dualist might say that there are natural laws governing mental properties that dictate that they appear and evolve according to the physical states of individuals. Thus, a position plainly incompatible with physicalism seems compatible with supervenience.

But matters are not quite that simple, as much depends on the modal strength of the supervenience claim. The dualist explanation just considered depends on the natural laws thought to link the mental and the physical. In the two formulations given above (GS and MPD), there was no restriction on the domain of possible worlds. If there are possible worlds in which the actual laws of nature are not laws, then the dualist explanation just given will not carry over to those worlds. Of course, one might maintain that the actual laws of nature are necessary in the strongest sense, so that there are no possible worlds in which they fail to be laws. Barring that, it is not obvious how an unrestricted supervenience thesis like MPD might be explained in a way not consistent with physicalism. In any case, the point to be appreciated is that the relevance of supervenience theses depends both on one's general views about modality and the stipulated modal strength of the supervenience thesis.

Realizationism

An increasingly popular rival to the identity and supervenience approaches to formulating physicalism is *realizationism* (Melnyk 2003). This is the thesis that every property either is physical or, on every occasion of instantiation, is realized by a physical property. Because the claim that one property realizes another seems to carry explanatory force, realizationism seems to avoid the weakness of supervenience approaches while not being as strong as the property identity thesis.

The notion of realization is not grounded in ordinary talk but has flourished in the context of philosophical discussion of functionalism. As such, realization talk should be understood by reference to the notion of a second-order property. Suppose F is a second-order property defined in the following way:

x has $F =$ there is some property P such that
(i) x has P and
(ii) x's having P meets condition C.

If the definitive condition C is that of playing a certain causal or functional role, then F may be said to be a functional property. Advocates of realizationism are likely to be functionalists as well as physicalists, taking advantage of the fact that functional properties are second order.

In one standard way of talking about realization, one can say that F is realized by a property P on the occasion of being instantiated by an individual x just in case x has P and x's having P meets condition C. If 'realizer' is used in this way, however, one should bear in mind that even if F is realized by P on a given occasion, the property P is not thereby sufficient for the instantiation of F. It is the combination of P and the fact that the individual's having P meets condition C that is sufficient for F.

The point is highlighted by some terminology introduced by Shoemaker (1981). Shoemaker distinguishes between "core" and "total" realizers. In the example given above, P is the core realizer of F on that occasion of instantiation. The total realizer can be specified as the property of having *P and being such that one's having P meets condition C.* Given the definition of F, it is trivially necessary that any individual that exemplifies a total realizer of F will also exemplify F. Further, it seems plausible to say that when F is instantiated in this fashion, that instance of F is nothing over and above the instantiation of that total realizer. By contrast, it does not seem appropriate to say that the instance of F is nothing over and above the instantiation of the core realizer P. In light of the core/total distinction, it seems that the realizationist thesis should be understood as the claim that every nonphysical property has, on every actual occasion of instantiation, a physical total realizer.

Realizationism thus understood is not as distant from the identity approach as one may have expected. If the total realizers of F are themselves physical, then the definitive condition C is itself physical. In that case, one may well want to say that F is physically definable. After all, F would then be definable using nothing but logical connectives, quantification over properties, and terms for physical properties. One may say that the identity thesis has been vindicated, for the multiply realizable properties are all identical with second-order properties, which can be seen to be themselves physical (Field 1992).

The Question of Justification

The History of Successful Physicalist Theorizing

One route to justifying physicalism is quite straightforward: Argue that since a wide variety of phenomena have already been shown to be ultimately physical in nature, the physicalist is justified in making the inductive leap to conclude that all phenomena are ultimately physical.

One way of thinking about this is in terms of the history of reductive success. That is certainly the justification imagined by those philosophers writing in the middle of the twentieth century who were concerned to defend a physicalist theory of the mind. They took it more or less for granted that scientific results had already demonstrated that sciences other than psychology reduced to "lower-level" sciences, and ultimately to physics. In effect, in this view, physicalism as a view about everything else (other than the mental and those things dependent on the mental) had already been demonstrated by a string of successful reductions. All that remained was to justify the extension of physicalism to the realm of the mental and the like.

This way of justifying physicalism is, however, hampered by controversies over the notion of reduction. The claim that biology has been shown to be completely reducible to chemistry, for instance, will not command immediate assent (see Reductionism). No one doubts that all sorts of interesting links have been found and developed, but the controversy is over the proper interpretation of those achievements.

Nonetheless, the general drift of the argument is clear: There is a history of what might be called successful physicalist theorizing—a history of success in showing that various properties are indeed to be counted as ultimately physical, as well as an invitation to generalize to all properties. Whether the argument is a good one depends on how exactly the relation of being "ultimately physical" is understood and whether, in light of that understanding, it is the sort of feature one can find in the actual results of science and should confidently project to other cases.

A somewhat different way to use the history of science to argue for physicalism is via what one can call the *argument from proven methodological utility.* Instead of claiming that what history displays is a string of successful theories showing that various properties are ultimately physical, this argument strategy turns on the claim that scientific practice has often *presupposed* that physicalism is true, and this presupposition is at least partly responsible for its success. The history of science shows that it pays to presume physicalism; the best explanation of this proven utility is that physicalism is true.

One advantage of this argument is that it does not require its advocate to find in the history of science any explicit theories setting out how apparently nonphysical properties are to be explained as ultimately physical. All that is needed is evidence that confidence in the existence of such an explanation played a key role in advancing scientific inquiry. The disadvantage is that it is not easy to adjudicate claims about the role of presuppositions. Even if it is plausible to say that actual scientists have presumed physicalism, more work is needed to show that this presumption (and not perhaps some less weighty or less significant one) played a role in dictating the theories constructed, experiments designed, or the like.

The Causal Impact Argument

The causal impact argument for physicalism is easy to state in a rough form. There are three premises. The first is the causal impact thesis: Every event is a cause of some physical event or other. The second is the causal completeness of physics: Every physical effect that has a sufficient cause has a sufficient *physical* cause. (Physical events that are not causally determined may be said to have their objective chances of occurrence determined by physical causes.) The third is the claim that causal overdetermination is not rampant.

Together, these three premises seem to imply that any apparently nonphysical property must be ultimately physical; any such property is involved in some event and thereby has a causal impact on the physical. Given the causal completeness of physics, however, anything that has a causal impact on the physical must be itself physical—unless, of course, the physical event is causally overdetermined, which option is ruled out by the third premise.

While the first and third premises seem to be plausible on casual examination, the second—the causal completeness of physics—needs special comment. It should be stressed that the thesis does not imply that the only legitimate causal explanations of physical events are ones that appeal exclusively to physical conditions and laws; it implies only that for every physical event that can be causally explained at all, there exists a *sufficient* causal explanation of it that appeals only to physical conditions and laws. The thesis can therefore be supported by considerations internal to physical theory, by judging the success of physical theory in producing such explanations, without having to rely on any claims about the relation of physical theory to anything else. This is a considerable advantage, as any such claims are likely to be philosophically contentious.

The causal impact argument should be sharply distinguished from the traditional attack on Cartesian dualism that questions the intelligibility of causation spanning the mental and physical divide. The argument does not concern the mental/physical divide in particular, nor does it turn on claims about intelligibility at all. It should also be distinguished from Davidson's (1970) famous argument for the thesis that every event is a physical event. While there are a few (indirect) relations between this argument and Davidson's, the latter turns fundamentally on considerations about the availability of strict causal laws to subsume cause/effect pairs, whereas the causal impact argument in no way relies on such considerations.

The causal impact argument has gained ground as the most popular route for justifying physicalism. There is one very important way in which the argument needs further careful development, however. How exactly should its conclusion be understood? If causation relates events, then one may take the argument as concluding that every event is a physical event. But that conclusion seems insufficient for physicalism, as discussed earlier. If the argument is to give us a conclusion that is recognizably physicalist, its premises must be, so to speak, appropriately calibrated, so that they concern properties and their role in causation. For example, if physicalism is understood as realizationism, the advocate of the argument must find a way both to refer to the role of properties in causation *and* to show that the only alternative to genuine overdetermination is for the causally relevant property to be either itself physical or, on that occasion, physically realized.

The Manifestability Argument

The manifestability argument is akin to the causal impact argument in that it, too, makes use of the causal completeness of physics. It does not, however, rely on the assignment of causes; it turns solely on counterfactual claims. The key premise is that the nonphysical is physically *manifestable* in the sense that if two individuals differ in some nonphysical respect, this difference is capable of showing up in a difference of physical conditions. More precisely, if x and y are discernible in some nonphysical respect, then there is some physical context C and some physical event type E such that if x were in C, an E-type event would occur, while if y were in C, no E-type event would occur.

The premise has considerable intuitive plausibility. Consider the infamous case of the mental. It seems that differences in mental states, even if they do not actually manifest themselves physically, are capable of doing so. (If, for instance, two people differ in their beliefs as to whether there are ducks present, the physical condition that amounts to asking "Are there ducks about?" would elicit different behaviors.) If the manifestability premise is granted, any nonphysical difference is reflected in a difference in potential physical consequences. But given the causal completeness of physics, any difference in potential physical consequences is grounded in actual physical differences. Hence, any nonphysical difference is reflected in some actual physical difference. In other words, the nonphysical supervenes on the physical (Papineau 1993; Loewer 1995).

Although the manifestability argument has not received the same attention given the causal impact argument, it is accused of being incapable of supporting an appropriately strong supervenience thesis (Witmer 1998). Even if the argument is not successful on its own, however, those physicalists inclined to rest their convictions on the causal impact argument should keep it in mind. Since counterfactuals are intimately related to causal claims, the manifestability argument may contain resources relevant to the proper development of the causal impact argument.

D. Gene Witmer

References

Armstrong, David (1968), *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.

Crane, Tim, and D. H. Mellor (1990), "There Is No Question of Physicalism," *Mind* 99: 185–206.

Davidson, Donald (1970), "Mental Events," in L. Foster and J. W. Swanson (eds.), *Experience and Theory*. Amherst: University of Massachusetts Press. Reprinted in Davidson, *Essays on Actions and Events*. Oxford: Oxford University Press, 1980.

Feigl, Herbert (1968), *The "Mental" and the "Physical": The Essay and a Postscript*. Minneapolis: University of Minnesota Press.

Field, Hartry (1992), "Physicalism," in J. Earman (ed.), *Inference, Explanation and other Frustrations*. Berkeley and Los Angeles: University of California Press, 271–291.

Gillett, Carl, and Barry Loewer (eds.) (2001), *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.

Hellman, Geoffrey, and Frank Thompson (1975), "Physicalism: Ontology, Determination, and Reduction," *Journal of Philosophy* 72: 551–564.

Horgan, Terence (1987), "Supervenient Qualia," *Philosophical Review* 96: 491–520.

——— (1993), "From Supervenience to Superdupervenience: Meeting the Demands of a Material World," *Mind* 102: 555–586.

Jackson, Frank (1998), *From Metaphysics to Ethics*. Oxford: Oxford University Press.

Kim, Jaegwon (1993), *Supervenience and Mind*. Cambridge: Cambridge University Press.

——— (1998), *Mind in a Physical World*. Cambridge, MA: MIT Press.

Kirk, Robert (1979), "From Physical Explicability to Full-Blooded Materialism," *Philosophical Quarterly* 29: 229–237.

Loewer, Barry (1995), "An Argument for Strong Supervenience," in E. Savellos and Ü. Yalçin (eds.), *Supervenience: New Essays*. Cambridge: Cambridge University Press, 218–225.

McLaughlin, Brian (1995), "*Varieties of Supervenience*," in E. Savellos and Ü. Yalçin (eds.), *Supervenience: New Essays*. Cambridge: Cambridge University Press, 16–59.

Melnyk, Andrew (2003), *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge University Press.

Papineau, David (1993), *Philosophical Naturalism*. Oxford: Blackwell.

Pettit, Philip (1993), "A Definition of Physicalism," *Analysis* 53: 213–223.

Poland, Jeffrey (1994), *Physicalism: The Philosophical Foundations*. Oxford: Oxford University Press.

Post, John (1987), *The Faces of Existence: An Essay in Nonreductive Metaphysics*. Ithaca, NY: Cornell University Press.

Robinson, Howard (ed.) (1996), *Objections to Physicalism*. Oxford: Oxford University Press.

Schiffer, Stephen (1990), "*Physicalism*," in J. Tomberlin (ed.), *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*. Atascadero, CA: Ridgeview, 153–185.

Shoemaker, Sydney (1981), "*Some Varieties of Functionalism*," *Philosophical Topics* 12: 83–118. Reprinted in Shoemaker, *Identity, Cause and Mind*. Cambridge: Cambridge University Press, 1984.

Smart, J. J. C. (1959), "Sensations and Brain Processes," *Philosophical Review* 68: 141–156.

Sturgeon, Scott (1998), "Physicalism and Overdetermination," *Mind* 107: 411–432.

Uebel, Thomas (1992), *Overcoming Logical Positivism from Within: The Emergence of Neurath's Naturalism in the Vienna Circle's Protocol Sentence Debate*. Amsterdam and Atlanta: Editions Rodopi.

Walter, Sven, and Heinz-Dieter Heckmann (eds.) (2003), *Physicalism and Mental Causation*. Exeter, UK: Imprint Academic.

Witmer, D. Gene (1998), "What Is Wrong with the Manifestability Argument for Supervenience?" *Australasian Journal of Philosophy* 76: 84–89.

See also **Carnap, Rudolf; Cognitive Science; Consciousness; Function; Intentionality; Logical Empiricism; Neurath, Otto; Neurobiology; Phenomenalism; Psychology, Philosophy of; Reductionism; Supervenience; Vienna Circle**

JULES HENRI POINCARÉ

(29 April 1854–17 July 1912)

Jules Henri Poincaré was a prolific mathematician, mathematical physicist, and philosopher. Often compared to Gauss, he made significant contributions to many areas of mathematics, including topology, non-Euclidean geometry, Lie groups, and differential equations. He also published four volumes of philosophical and popular writings on science and mathematics.

As a philosopher of science, Poincaré is sometimes thought of as a conventionalist (see Giedymin 1982). Poincaré did argue that certain central decisions in science have a strongly conventional element. But as a general label, "conventionalism" does not accurately account for the variety and complexity of views that make up his philosophy of science. His clearest and most convincing arguments for conventionalism are found in his philosophy of geometry. Poincaré argued that the choice between Euclidean and non-Euclidean geometry is radically underdetermined. Either can be chosen, for either hypothesis can be made to fit any data. The result is thus a convention, for only convenience can determine the choice. But even here, at his most conventional, Poincaré (1982) believed that the candidates for a geometric choice are a priori constrained by the nature of human minds; furthermore, which choice appears to be the most convenient is influenced by the nature of the empirical world (276). In this way, all reference to "convention" in Poincaré must be balanced by his semi-Kantianism on the one side and his empiricism/realism on the other. Conventions in science were not in general "mere" for Poincaré.

Geometric Conventionalism

Poincaré argued that geometry is not synthetic a priori as Kant thought because it is not possible to have intuitions of points, lines, or spatial distances. Nor is geometry in any straightforward sense analytic a priori, for its truths are not determined by the nature of the fundamental geometric concepts alone. In the absence of axioms, the concepts of point, line, and plane do not distinguish one geometrical system from another. On the contrary, the fundamental geometric concepts become "implicitly defined" only after the geometric axioms are chosen.

Geometry is not empirical either. In order to test whether space is Euclidean or non-Euclidean, some measurements would have to be made. But such measurements would have to be based on assumptions in physics—for example, that light travels in straight lines. So they would have to depend on a prior understanding of 'straight line' and its physical realizations. This point was brought out very clearly by the parallax result, which showed that a fixed star observed during a solar eclipse appears in a different location than its ordinary location observed at night. The intuitive response is that the light from the star bends as it travels past the sun, owing to the sun's gravitational field. Poincaré famously pointed out that this means that geometry cannot be empirically tested in any straightforward manner. In the light of the parallax result, a choice had to be made whether to revise the physical hypothesis that light travels in straight lines or the Euclidean hypothesis that space is flat. In such a case, which choice is "correct" is undetermined by the evidence, for it depends only on what is most convenient given the global background theory. That Euclidean geometry can be chosen no matter what the data say means that geometry is more like language than an empirical claim. It is in this sense that it is conventional. The argument relies on the thesis that it is possible to test only the conjunction of physics plus geometry. Generalized by Duhem into the view that it is possible to test only a cluster of hypotheses rather than a single one, and further generalized by Quine, one can see that Poincaré's "conventionalism" was in some ways a foundation for contemporary holism (see Duheim Thesis).

An important limitation on Poincaré's candidates for geometry was that they had to be of constant curvature. This was of course rejected by the theory of general relativity, according to which the curvature of space depends on the distribution of its matter. However, Poincaré regarded the very idea of spatial measurement as requiring rigid body motion, where rigid bodies are simply idealizations of physical bodies. People know that ordinary bodies, including their own, move. Indeed it is precisely the possibilities underlying this sort of motion by which the difference between change of state and change of place is understood. And understanding the idea of motion—change of place—is central to geometrical thinking.

Also central to geometric thinking is the idea of a group, which Poincaré regarded as a form of understanding, given a priori. When conjoined with the presupposition that rigid body motion is possible, the group concept yields the Lie groups, which entail that there are three basic geometries of constant curvature: Euclidean, Riemannian, and Lobachevskian. According to Poincaré, experience helps us to pick out which group is merely the most convenient among the three possible. The indefinite iterability of the geometric operations means that geometry also presupposes arithmetic and its a priori intuitive basis (Poincaré 1982, 276) (see Conventionalism).

Intuition in Pure Mathematics

In his philosophy of pure mathematics, and especially of arithmetic, Poincaré was a self-declared defender of Kant against both logicism and Hilbert's program. The extent to which he was truly Kantian is a tricky question. Like Kant, Poincaré appealed to intuition as a necessary—yet nonlogical and nonconceptual—source of knowledge. Unlike Kant, the a priori intuition he focused on is indefinite iteration rather than space or time, where 'indefinite iteration' means the capacity of the mind to conceive the unlimited repetition of certain acts. There was an important connection between intuition and sense experience for Poincaré as for Kant, in that indefinite repeatability was presupposed in his displacement account of motion in, and experience of, space. But iterability seems a little less immediate than Kant's spatiotemporality. (The spatiotemporality of the world of experience seems obvious; the indefinite iterability of certain features of experience may seem downright false.) Poincaré (1913, 44) later also endorsed the intuitive nature of the continuum, claiming that experience would be impossible without it. Why this is so is left mostly unexplained and remains somewhat obscure.

Though the intuitive nature of the continuum is not clearly defended, Poincaré did argue vehemently for the intuitive basis of arithmetic. The nature and role of intuition in Poincaré's philosophy of mathematics is relevant to his philosophy of the natural sciences, and in particular to understanding what sort of a so-called conventionalist he was for the following reason. He believed that there is a hierarchy of sciences with arithmetic at the most fundamental level. Analysis builds on arithmetic, geometry builds on analysis, and physics builds on geometry. The epistemology of arithmetic thus provides a foundation for the epistemology of the rest of science in Poincaré's hierarchy (Poincaré 1902; Folina 1992 and 1995).

The main opposition to the Kantian paradigm for arithmetic was logicism, represented to Poincaré by people such as Russell and Couturat. In contrast to the logicists, Poincaré regarded mathematics as irreducibly mathematical, and intuition was his answer to the question of what the essence of mathematics is. Poincaré's strongest arguments against logicism and for intuition were circularity objections against any formal attempts to symbolically reconstruct arithmetic. His claim was that all such formal systems, if they are adequate for deriving the theory of the natural numbers, presuppose the intuition they are trying to avoid. This is a strong claim and one that is hard to defend in general.

His defense of intuition came in two stages. Prior to 1906, his most potent argument was that symbolic logic after Frege, Boole, etc., is itself inherently combinatorial, and so inherently mathematical in that it presupposes iteration in the form of recursive procedures. Thus, any derivation of arithmetic from logic has already presupposed something inherently mathematical. So it is circular.

By 1906, his circularity objection was modified in light of the set-theoretic paradoxes. Hereafter, he was no longer so concerned with mere circularity; his objections instead focused on the vicious circularity of certain definitions available in naive set theory and elsewhere in classical mathematics. The problematic definitions typically involve a property the scope of whose quantifier includes the object being defined. These definitions, via writings by both Poincaré and Russell, became known as "impredicative."

As it turned out, impredicative definitions are very common in mathematics. So the class of definitions affected by the critique was surprisingly broad. The problem in accepting Poincaré's point of view is that not many mathematicians would be willing to sacrifice mathematical methods for a philosophically based restriction—especially not when there are other available ways to avoid paradox, such as axiomatic set theory. Poincaré himself did not restrict his work to predicative mathematics. With his other circularity arguments he sometimes argued that the ineliminability of impredicativity simply showed the need for intuition in certain domains (see Heinzmann 1985).

The main point for this entry is that Poincaré persisted in defending the intuitive basis of mathematics, in his efforts to defend Kant's basic picture of mathematics from the attacks of logicism. In this account, at least part of the nature of mathematics is determined by the kinds of minds that human beings have: finite, yet capable of infinite combinatorial thinking. Since the rest of science appeals to mathematics and rests on it, science is not governed *simply* by conventions. The existence of certain conventional choices in science does not, therefore, mean that science as a whole is "merely" conventional.

Scientific Realism

If Poincaré's modified Kantianism balances his conventionalism from the a priori side, his empiricist-based realism balances it from the other. Whereas a priori intuition can be thought of as dictated by the a priori nature of the mind, empirical facts are dictated by the *a posteriori* nature of the physical world. It should be noted that by 'fact' in science Poincaré meant an intersubjectively verifiable state of affairs. The empirical world is in this account relevant to the "conventional" choice between Euclidean and non-Euclidean geometry. Which geometry is most convenient will depend on the use to which it needs to be put. Physics uses geometry, so the best current theories in physics will contribute to determining what counts as the simplest, most convenient geometry. And the best current theories in physics are at least in part determined by the way the world is.

Of course, one must not overemphasize the sense in which Poincaré was a realist. His conventionalism was not strictly limited to his philosophy of applied geometry. Poincaré used the term 'convention' outside of geometry as well, to include certain principles that originated as part of experimental science. His view was that when a bit of science is particularly well confirmed and particularly fundamental, it can end up being called a "principle" and treated as a convention. There is obviously some affinity here with Quine's "web of belief" metaphor. Though Poincaré would not have carried the view so far as to eliminate distinctions such as analytic/synthetic, or to eliminate all hierarchy in science, he did regard the boundary between the conventional and the experimental parts of science as both elastic and a little bit fuzzy. Not only can an experimentally based principle become a convention; a convention rooted in a principle can still be undermined by experiment—even if not directly falsified (Poincaré 1982, 318–319). Nevertheless, at any one time, there is a difference between the conventional and the empirical parts of science; and, in emphasizing this distinction, Poincaré resembles the logical positivists whom he influenced.

Despite the appeal to convention further up in Poincaré's hierarchy from geometry, the real world is "out there" and cannot be changed at will. That is, despite Poincaré's anti-realist sympathies, he argued (against the more radical, exaggerated conventionalism of Le Roy) that science is neither whimsical nor wholly conventional (Poincaré 1982, 335). Its theories are not mere creations. Scientific facts are "crude facts" translated into scientific language; but scientific facts are not *created* by scientists (208). The existence of conventions in a scientific theory does not, therefore, mean that the theory as a whole is a mere convention, nor does it mean that any part of it is trivial. Furthermore, change in science is governed and limited by the facts. Progress in science can involve a kind of revolution, where a whole paradigm is overturned for another. But Poincaré argued that even in this sort of case, the success of the new theory usually depends on some kernels of truth in the old. He thus emphasized continuities in science underlying the changes and even the revolutions. Objective truth in science and in everyday life consists of the enduring relations between things. Even an overturned theory can contain many truths about the relations between things—especially when what is overturned is the conception of the "things" themselves rather than the relations, and especially when these relations can be expressed mathematically (Poincaré 1982, 350–351). Poincaré's realism is thus expressed in his beliefs in genuine progress in science and that science discovers objective truths, even if these truths are largely structural, or relational, in nature.

Conclusion

Poincaré's philosophy of science is both complicated and interesting (see Stump 1989). He appears in different guises as a conventionalist, empiricist, Kantian, constructivist, logical positivist, anti-realist, and even a realist with a robust concept of truth and an interesting conception of objectivity. Though his popular writings often appear glib and polemical, they contain much wisdom, and remain influential, for philosophers of science today.

JANET FOLINA

References

Folina, Janet (1992), *Poincaré and the Philosophy of Mathematics*. London: Macmillan.

——— (1995), "Poincaré on Mathematics, Intuition and the Foundations of Science," *Proceedings of Philosophy of Science Association* [PSA] *1994*. East Lansing, MI: PSA, 217–226.

Giedymin, Jerzy (1982), *Science and Convention: Essays on Henri Poincaré's Philosophy of Science and the Conventionalist Tradition*. Oxford: Pergamon.

Heinzmann, Gerhard (1985), *Entre Intuition et Analyse: Poincaré et le concept de prédicativité*. Paris: Librairie Scientifique et Technique Albert Blanchard.

Poincaré, Jules Henri (1902), *La Science et L'Hypothèse*. Paris: E. Flammarion.

——— (1905), *La Valeur de la Science*. Paris: E. Flammarion.

——— (1908), *Science et Mèthode*. Paris: E. Flammarion.

——— (1913), *Dernières Pensées*. Paris: E. Flammarion.

——— (1982), *The Foundations of Science*. Washington, DC: The Science Press.

Stump, David (1989), "Henri Poincaré's Philosophy of Science," *Studies in History and Philosophy of Science* 20: 335–363.

——— (1991), "Poincaré's Thesis of the Translatability of Euclidean and non-Euclidean Geometries," *Noûs* 25: 639–654.

KARL RAIMUND POPPER

(28 July 1902–17 September 1994)

Karl Popper (Sir Karl Raimund) is widely considered to have been one of the greatest philosophers of science of the twentieth century. He was also an eminent social and political philosopher, a metaphysical indeterminist, and a relentless critic of authoritarianism. Popper's famously trenchant defense of the "open society" has been widely influential, not least for the highly original manner in which he traced the roots of totalitarian ideologies to historicist presuppositions. However, his most original and lasting contributions to the philosophy of science were his rejection of the Baconian observationalist-inductivist view of scientific methodology, his emphasis on the centrality of problem solving to the scientific enterprise, and his advocacy of the view that only a system of theories that is falsifiable by experience should be accorded genuine scientific status.

Life

Karl Raimund Popper was born in Vienna, the youngest child of middle-class parents of Jewish origin. His father, a barrister, communicated to him an interest in social and political issues that he was never to lose, while his mother cultivated in him a deep and abiding passion for music. He was educated at the University of Vienna, where he studied mathematics, physics, psychology, music, and philosophy. He obtained a Ph.D. in philosophy in 1928 and taught in secondary schools from 1930 to 1936. Although he was friendly with some members of the Vienna circle of logical positivists, whose concern with science he shared, he was sharply critical of many of their central principles, particularly what he took to be their misplaced concern with the theory of meaning. In 1937, concerns about the growth of Nazism led him to emigrate to New Zealand, where he became a lecturer in philosophy at Canterbury University College, Christchurch. In 1946, he moved to England to become a reader in logic and scientific method at the London School of Economics; he became a professor there in 1949. As his reputation and influence grew, Popper received many honors. He was knighted in 1965 and elected a Fellow of the Royal Society in 1976. He retired from academic life in 1969, though he remained intellectually active until his death.

Works

Popper's philosophy of science was first articulated in *Logik der Forschung* (Popper 1935), which was published in the Vienna Circle's series *Schriften zur wissenschaftlichen Weltauffassung*. This circumstance contributed to the initial misconception in the anglophone world that Popper was a positivist. Popper (1945 and [1957] 1961) widened his focus to cover historical, social, and political issues, and advanced a powerful critique of historical determinism as the theoretical assumption underpinning totalitarian ideologies. He published two collections of thematically linked papers (Popper [1963, 1965] 1969 and [1972] 1979), a reply to his critics in the volume on his work in the Schilpp (1974) Library of Living Philosophers, an intellectual autobiography (Popper 1976), and an examination of the mind–body problem with John Eccles (Popper and Eccles 1977). Most of these works have gone through several editions, in some cases incurring significant modification in the process. In the 1990s, Routledge published a number of posthumous editions of Popper's work and thought. These include collections of his lectures and essays (Popper 1992 and 1994 [including a critique of contemporary irrationalism]), an interview (Popper 1997), a series of essays on early Greek philosophy (Popper 1998), and the translation *All Life Is Problem Solving* (Popper 1999). Popper's manuscripts and correspondence are collected in the archives of the Hoover Institution, Stanford University, where he was a senior research fellow.

Science and Metaphysics: The Problem of Demarcation

Popper's main enterprise was to construct a model of scientific rationality that embodied an account of the logical relationships between theoretical and observation statements in science, and, associated with this, a prescriptive methodology. He was an epistemological fallibilist who recognized that the "central problem of epistemology has always been and still is the growth of knowledge" but who held that the growth of knowledge "can be studied best by studying the growth of scientific knowledge" (Popper [1935] 1959, 15). Consequently, his concern with the problem of demarcation in philosophy of science was intended to be seen as having the widest-ranging philosophical implications, since he proposed that epistemology "should be identified with the theory of scientific method" (49). He described how the issue of demarcation relates to his own general objectives as follows:

> [My] business, as I see it, is not to bring about the overthrow of metaphysics. It is, rather, to formulate a suitable characterization of empirical science, or to define the concepts "empirical science" and "metaphysics" in such a way that we shall be able to say of a given system of statements whether or not its closer study is the concern of empirical science. ([1935] 1959, 37)

Demarcation, then, is the problem of distinguishing the empirical sciences from nonempirical areas such as logic, pure mathematics, and metaphysics, as well as pseudosciences such as astrology and phrenology. However, Popper dissented from the approach of the logical positivists, who had addressed this problem by effectively identifying it with that of meaning, thus making the verifiability of a proposition a criterion of both its meaningfulness and its scientific status. This, in Popper's eyes, reduces a real philosophical issue to a trivial verbal issue. It is also a naturalistic error, because it treats demarcation as a problem amenable to the method of discovery and presupposes the existence of a clear-cut, impenetrable line between science and metaphysics. Thus it ignores the fact that metaphysical theories can generate, or even develop into, scientific theories, as, for example, with the classical theory of atoms. By contrast, Popper's approach to formulating a criterion of demarcation was nonnaturalistic or prescriptive; it involved offering a "proposal for an agreement or convention" ([1935] 1959, 37). For him, the line between science and metaphysics was to be drawn by agreement and decision, not by discovery. Such an agreement, however, must be informed by both logical and methodological considerations, and while he acknowledged that "a reasonable discussion of these questions is only possible between parties having some purpose in common" (ibid), he was optimistic about the possibility of rational assent on the part of those who shared his goal of adequately characterizing empirical science.

Popper's rejection of the naturalistic approach to the problem of demarcation taken by the logical positivists has a counterpoint in his repudiation of their view that science is characterized by its inductive methods, in which universal laws are supposedly inferred from a set of singular statements such as experimental observation reports. This is the traditional observationalist-inductivist paradigm of scientific investigation; the repudiation of this paradigm links the problems of induction and demarcation in Popper's philosophy. His case against it rests on three central contentions. First, there are no "pure" or theory-free observations. "Observation is always selective. It needs a chosen object, a definite task, an interest, a point of view, a problem" (Popper [1963, 1965] 1969, 46). Therefore, observation is theory laden and involves applying theoretical terms, a descriptive language, and a conceptual scheme to particular experiential situations. Second, scientific laws are strictly unverifiable. All such theories are universal in nature, and no finite collection of observation statements, however great, is logically equivalent to, or can justify, an unrestricted universal proposition. Third, induction, conceived of as a system of logical inferences that generates scientific law from the particularity of experimental results, is a "myth" (53). Such inferences, Popper held, play no role in scientific investigation or in human life generally.

At one level, then, Popper concurred with Hume's critique of induction. However, a crucial counterpart, which he believed Hume himself had missed, is that while no number of positive outcomes at the level of experimental testing can demonstrate the truth of a scientific theory, a single genuine counterinstance is logically decisive. "Hume showed that it is not possible to infer a theory from observation statements; but this does not affect the possibility of refuting a theory by observation statements" (Popper [1963, 1965] 1969, 55). By the canonical *modus tollens* rule of classical logic, it is possible to deductively infer the falsity of a universal proposition once the truth-value of an appropriately related singular proposition is established. For Popper, this means that the central kind of inferences involved in science are deductive ones from observation reports of the form "This *A* is not *X*" to the *falsity* of the corresponding universal hypotheses, and such inferences occur in the critical testing of such hypotheses rather than in their generation.

Accordingly, Popper's view of the relationship between scientific theory and experience was both anti-inductivist and anti-Humean: Theory is not logically derived from, nor can it be confirmed by, experience, though experience can and does delimit it. He argued that human knowledge generally, including scientific theory as one of its most refined forms, is both fallible and wholly hypothetical, and it is produced not by logical inference but by the creative imagination. The central rational activity in science is problem solving, whereby new hypotheses are imaginatively projected to solve problems that have arisen with respect to a preexisting theoretical framework. This process may be retrospectively retraced though "more and more primitive theories and myths … [to] unconscious, inborn expectations" (Popper [1963, 1965] 1969, 47). The critical role of experience in science is to show us not which theories are true but which theories are false. However, a theory that has successfully withstood critical testing is thereby "corroborated" and may be regarded as preferable to falsified rivals. In the case of rival nonfalsified theories, for Popper, the higher the informative content of a theory, the better it is scientifically, because every gain in content brings with it a commensurate gain in predictive scope and testability. For that reason, he held that a good scientific theory will be more *improbable* than its rivals, because the probability and informative content of a theory vary inversely: "If our aim is the advancement or growth of knowledge, then a high probability (in the sense of calculus of probability) cannot possibly be our aim as well: *these two aims are incompatible*" (218).

It is thus not too much to say that Popper's perception of the asymmetrical logical relation between verification and falsification lies at the heart of his philosophy of science: A universal scientific theory cannot, in principle, be verified, but a single counterinstance can and does decisively falsify it. Accordingly, he held that from a logical perspective, a system of theories is scientific only if it is refutable or falsifiable:

> I shall not require of a scientific system that it shall be capable of being singled out, once and for all, in a positive sense; but I shall require that its logical form shall be such that can be singled out, by means of empirical tests, in a negative sense: *it must be possible for an empirical scientific system to be refuted by experience.* ([1935] 1959, 40–41)

Popper defined this demarcation criterion most clearly in terms of the relation between a scientific theory and "basic statements," which are to be understood as singular existential statements of the form "There is an *X* at *Y*." In this definition, where a theory is scientific, it must exhaustively divide basic statements into two nonempty classes:

1. the class of basic statements that are consistent with the theory, or that the theory "permits" (class 1); and
2. the class of basic statements that the theory rules out, or "prohibits" (class 2).

The latter class is, in Popper's account, by far the more important, as it constitutes the theory's potential falsifiers; that is, the truth of any statement in class 2 implies the falsity of the theory. In short, for Popper ([1935] 1959), "[A] theory is falsifiable if the class of its potential falsifiers is not empty" (86).

One of the most controversial implications of this criterion, which Popper strongly affirmed, is that psychoanalytic theory and the contemporary Marxist theory of history cannot be deemed scientific. Ironically, the all-inclusive nature of psychoanalytic explanation, which Freud and his followers saw as the principal basis for the scientific status of psychoanalysis, turns out on this account to be a critical weakness, for it entails that psychoanalysis is not genuinely predictive. Psychoanalytic theories, Popper argued, are insufficiently precise to be "prohibitive," that is, to have negative implications in the form of a class of potential falsifiers, and so are not subject to experimental falsification. Popper denied the claim to scientific standing of the Marxist theory of history on the grounds that while it was prohibitive and therefore testable in its original form, many of its implications turned out to be false. However, Marxists responded by reformulating the theory ad hoc to make it consistent with the falsifying evidence. They thus "gave a 'conventionalist twist' to the theory; and by this stratagem they destroyed its much advertised claim to scientific status" (Popper [1963, 1965] 1969, 37).

Falsificationism and Methodological Rules

Popper was aware that there is a significant disparity between the precision of the logical analysis of statements contained in his demarcation criterion and the complex, heterogeneous nature of actual scientific practice. Since observation is, as he insisted, itself fallible, an experimental result can always be questioned. His account of the logic of falsifiability was thus tempered by an explicit recognition that scientific theories are often retained in the face of conflicting or anomalous empirical evidence and that, in actual scientific practice, a single conflicting instance or counterinstance is never sufficient to force the repudiation of an established theory. He was also cognizant of the fact that against dogmatic or uncritical advocacy, "no conclusive disproof of a theory can ever be produced" ([1935] 1959, 50). In short, he recognized that a logical analysis of statements alone is not sufficient to recapitulate the unique character of empirical science.

To achieve that objective, Popper concluded, it is necessary to embed falsifiability in a normative methodology, which in this connection relates to decisions that must be made as to how to deal with scientific statements—decisions which, in turn, are determined by one's aims. For Popper ([1935] 1959), the aims of elucidating empirical science and constructing a model of scientific rationality bring with them a need to adopt a set of rules that will "ensure the testability of scientific statements; which is to say, their falsifiability" (49). Accordingly, the "supreme rule" associated with the falsifiability criterion, which functions as a norm with which all other methodological rules must accord, is "the rule which says that the other rules of scientific procedure must be designed in such a way that they do not protect any statement in science against falsification" (54). This rule prohibits, as unscientific, any ad hoc reformulation of a theory to meet contradictory evidence. The recognition of the hypothetical and fallible nature of human knowledge, and with this recognition the willingness to subject even one's most cherished theory to a critical test that could conceivably show it to be false, became, for him, the defining characteristic of the true scientific mentality: "The wrong view of science betrays itself in the craving to be right; for it is not his *possession* of knowledge, of irrefutable truth, that makes the man of science, but his persistent and recklessly critical *quest* for truth" (281).

Science, History, and Society

Popper's emphasis on the epistemological and methodological importance of conditions that foster the critical evaluation of deeply held beliefs is an important link between his philosophy of science and his social and political philosophy. In *The Open Society and Its Enemies* and *The Poverty of Historicism* he launched a powerful attack on historicism (Popper 1945 and [1957] 1961). Historicism encompasses the views that historical processes and events are governed by immutable deterministic laws and that history itself evolves inexorably toward a teleological goal, a conception that is fundamental to most "dialectical" theories of history. Popper saw it as the principal theoretical assumption underlying political authoritarianism, and he considered the theories of Plato and Marx particularly illustrative in this regard. Associated with historicism is the "historicist doctrine of the social sciences," the view that the main task of the social sciences is "to make historical predictions, such as the predictions of social revolutions" (Popper [1963, 1965] 1969, 338). For the historicist, just as the natural sciences allow the prediction of eclipses, appropriate knowledge of the "laws of history" would allow the social sciences to predict social processes such as revolutions. Popper argued that this presupposes an incorrect view of the nature of scientific law and scientific prediction—a view in which the unconditional prediction of eclipses is wrongly taken as typical, whereas it actually applies only to systems, such as our solar system, that are "well-isolated, stationary, and recurrent" (339). Because of the universality of scientific theory, prediction in natural science is typically conditional and limited in scope to a particular aspect of the system under investigation. In general, Popper further argued, a predictive science of human history is impossible, because the "course of human history is strongly influenced by the growth of human knowledge … [and] we cannot predict, by rational or scientific methods, the future growth of our scientific knowledge" ([1957] 1961, v–vi).

Popper saw "large-scale social planning"—for instance, revolutionary attempts to restructure the social order—as an inevitable consequence of historicism, and he accordingly rejected it, advocating instead "piecemeal social engineering" as the central mechanism for social reform. In this latter mode, he argued, intentional actions should be directed toward a limited number of objectives, facilitating the observation of any adverse unintended effects. Here Popper's conservatism contrasts strongly with his advocacy of bold scientific conjectures, but the strategy outlined does parallel the critical testing of theories in scientific investigation. It is also linked with Popper's commitment to negative utilitarianism, the moral principle that one should seek to reduce suffering to a minimum. In his view, the function of the state is not to impose a preconceived ideal of the good but rather to ameliorate generally acknowledged social ills. Popper held that an open society, providing as it does for peaceful changes of government and a milieu in which the critical appraisal of policy is fostered and encouraged, is an essential prerequisite for critical thought and the emancipatory goals flowing from such thought. It thus happens that problem solving is as characteristic and reflective of common humanity at the social and political levels as at the level of scientific investigation—an insight that is one of the integrating forces in Popper's thought.

Later Developments: Objective Knowledge, the Third World, and Verisimilitude

In his later works Popper looked to biology rather than physics, seeing the growth of human knowledge through the critical appraisal of competing theories in terms that are strongly Darwinian:

> The growth of our knowledge is the result of a process closely resembling what Darwin called "natural selection"; that is, *the natural selection of hypotheses*: our knowledge consists, at every moment, of those hypotheses which have shown their (comparative) fitness by surviving so far in their struggle for existence; a competitive struggle which eliminates those hypotheses which are unfit. (Popper [1972] 1979, 261)

This led him to advance an evolutionary epistemology, in which he attempted to account for the very existence of the philosophical and scientific quest for truth in terms of natural selection. In this he represented biological adaptations as forms of problem solving and, ultimately, of knowledge. This epistemology related in turn to a new, radically pluralist metaphysics, in which Popper construed objective knowledge as residing in neither the world of physical objects and states (the "first world") nor in the world of minds and mental processes (the "second world"), but in an autonomous objective world containing products of the human mind such as theories, possible objects of thought, arguments, and values, which he called a "third world" or "World 3" (Popper [1972] 1979, 154–160). These products of the human mind are encoded in such material first-world objects as books, journals, and mathematical tables but are as autonomous relative to them as they are relative to minds: Their logical interrelations and properties have no direct counterparts in the first or second worlds, and their development transcends the minds in which they originate and the media in which they are instantiated. Knowledge thus stored is objective, he argued (controversially), in the sense that although it is a product of the human mind, its continued existence is nonetheless independent of its being accessed by any mind: It is knowledge without a knowing subject (109). He contended that scientific knowledge, properly understood, is objective in this sense, and that much of traditional epistemology is misguided in its focus on knowledge in the subjective sense of the mental states of the second world.

One of the more important later developments of Popper's thought was an explicit engagement with the notion that theoretic progress in science is explicable in terms of ever-closer approximations to the truth. A lifelong realist, Popper was nevertheless initially relatively silent regarding the truth of scientific theories, largely because he was aware of the contemporary metaphysical difficulties associated with that concept. However, under the influence of Alfred Tarski, Popper came to endorse the correspondence theory of truth as the fundamental idea underpinning rational criticism (Popper [1972] 1979, 263–264). This in turn precipitated a move on his part toward formalizing the intuitive notion of "truthlikeness" or "verisimilitude," on the grounds that "we simply cannot do without something like … [the] idea of a better or a worse approximation to truth" (Popper [1963, 1965] 1969, 232). Accordingly, Popper formulated the metalogical concept of verisimilitude by linking the ideas of truth and content. The content of any statement is the class of all statements that follow from it, which is zero only in the case of tautologies. Hence, the class of true logical consequences of a theory is its "truth content," while its "falsity content" is the class of its false consequences. Given two rival theories t_1 and t_2, and assuming that their truth content and falsity content are comparable, then, Popper ([1963, 1965] 1969) asserted:

> We can say that t_2 is more closely similar to the truth, or corresponds better to the facts, than t_1, if and only if either: (a) the truth-content but not the falsity-content of t_2 exceeds that of t_1, or (b) the falsity-content of t_1, but not its truth-content, exceeds that of t_2. (233)

Popper's purpose here was to address, positively, the problem of scientific progress within a falsificationist frame of reference. This is a crucial issue, not least because the working scientist frequently has to operate with theories that are approximations and, as such, are known to be, strictly speaking, false. Given this, and his insistence that empirical testing can demonstrate the falsity but not the truth of a theory, and that it is consequently impossible to know whether any theory is in fact true, Popper needed to show that there are rational grounds for preferring some false theories over others. With the theory of verisimilitude he sought to provide just such an account. In this theory, it is legitimate to regard a falsified scientific theory t_2 that has a higher truth content than a rival falsified theory t_1 as a better theory than t_1, provided the falsity content of t_2 is not also greater than that of t_1. Moreover, "better" in this context is now understood to mean not merely that theory t_2 is more testable or has greater explanatory force than theory t_1, but that it has a higher level of verisimilitude or is closer to the truth than t_1. The growth of scientific knowledge, in other words, is here represented as progress toward the truth, even where such progress takes place through falsification of theories.

However, the work of Miller (1974a, b), Harris (1974), and Tichý (1974) demonstrated that Popper's two conditions for comparing the truth and falsity contents of theories are both satisfiable only when the theories concerned are true. In the crucially important cases of false theories, Popper's account is formally defective, in that where any false theory t_2 has excess content over a rival false theory t_1, the truth content and falsity content of t_2 will both exceed those of t_1. Fatally, in the case of false theories, the conditions outlined by Popper can never be satisfied.

Popper ([1972] 1979) subsequently acknowledged this deficiency in his definition of verisimilitude, but he argued that it had never been his intention to imply "that degrees of verisimilitude … can ever be numerically determined, except in certain limiting cases" (59). The central merit of the concept is heuristic and intuitive, he maintained, and the failure of his formal definition of verisimilitude should not preclude its utilization in evaluating theories whose content is relativized to problems considered relevant by the practicing scientist (368, 371–372).

Whereas some commentators have found that this response acceptably reflects the position of verisimilitude in Popper's philosophy, many others remain critical, seeing the failure of Popper's definition as indicating a more general fragility in his philosophy of science. It is clear, however, that the deficiencies in this account of verisimilitude leave unresolved the important problem of defining scientific progress through falsification in a formally satisfactory way.

Quine and Lakatos, following an argument first suggested by Poincaré (1902 and 1905) and Duhem ([1906] 1914), have both also attacked the key notion in Popper that the falsification of scientific theories can be yielded by discrete critical tests. Quine (1963), who called into question the assumption of a clear-cut distinction between analytic and synthetic statements as one of the two "dogmas of empiricism," also vigorously advocated a holistic view of empirical tests, contending that "our statements about the external world face the tribunal of experience not individually but only as a corporate body" (41). In this view, the entire system of human knowledge impinges on experience "only along the edges" (42); hence, contradictory empirical evidence has no necessary connection with any given theory and may in fact force a reassessment of a range of elements within the system.

In a similar vein, Lakatos (1970) argued that many of the most respected scientific theories cannot be refuted by individual critical tests, as they are not in themselves prohibitive; rather, they "forbid an event occurring in some specified finite spatio-temporal region ... only on the condition that no other factor ... has any influence on it" (101). Such theories, in other words, require the addition of an implicit *ceteris paribus* clause if they are to have prohibitive implications at all, and the *ceteris paribus* clause can always be replaced by another to make the theory consistent with apparently falsifying evidence. Hence, Lakatos contended, theories are falsified not in isolation but as integral elements of a "degenerating" research program that is supplanted by a rival program with equal predictive success and additional "heuristic power" (155).

Popper's response to these and related criticisms was to emphasize the significance of assumed background knowledge in the encounter between theory and experimental results, acknowledging that such assumed knowledge is fully as open to challenge and revision as the theory that its tentative acceptance permits us to test. He argued, however, that such considerations do not entail holistic conclusions, because in many cases it is quite possible to determine which hypothesis or group of hypotheses "is responsible for the refutation" ([1963, 1965] 1969, 239). This reply indicates Popper's awareness, in his later works, that his falsifiability criterion requires supplementation by extraneous pragmatic considerations—an awareness that extended to his recognition of ad hoc modification of theory as a recurrent, albeit undesirable, feature of general scientific practice. Consequently, while maintaining his advocacy of proposals that enshrine the critical spirit in scientific investigation, he went so far as to state that "the methodology of science (and the history of science also) becomes understandable in its details if we assume that the aim of science is to get explanatory theories which are as little ad hoc as possible: a 'good' theory is not ad hoc, while a 'bad' theory is" ([1963, 1965] 1969, 61).

This contention is not at all implausible, and, like many of Popper's other methodological precepts for science, it has been influential. However, "bad" theories here are to be taken as including bad scientific theories, and this is a significant move away from Popper's earlier view that the incorporation of ad hoc elements, as in the Marxist theory of history, suffices to make a theory unscientific. It also remains questionable whether the concept of the ad hoc can be made sufficiently determinate to do the work required of it in Popper's later work, in which there is thus an evident weakening of his formal demarcation criterion in the direction of pragmatic interpretation (Stokes 1998, 21).

Stephen P. Thornton

References

Ackermann, R. (1976), *The Philosophy of Karl Popper*. Amherst: University of Massachusetts Press.

Baudoin, J. (1989), *Karl Popper*. Paris: PUF.

Duhem, P. ([1906] 1914), *La théorie physique, son objet et sa structure*. Paris: Editions Rivière.

Harris, J. (1974), "Popper's Definitions of Verisimilitude," *British Journal for the Philosophy of Science* 25: 160–166.

Jacobs, S. (1991), *Science and British Liberalism: Locke, Bentham, Mill, and Popper*. Aldershot, UK: Avebury.

James, R. (1980), *Return to Reason: Popper's Thought in Public Life*. Shepton Mallet, UK: Pen.

Johannson, I. (1975), *A Critique of Karl Popper's Methodology*. Stockholm: Scandinavian University Books.

Lakatos, I. (1970), "Falsification and the Methodology of Scientific Research Programmes," in Lakatos and A. Musgrove (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 91–196.

——— (1978), *The Methodology of Scientific Research Programmes*. Edited by J. Worrall and G. Currie. Cambridge: Cambridge University Press.
Magee, B. (1977), *Popper*. London: Fontana.
Miller, D. (1974a), "On the Comparison of False Theories by Their Bases," *British Journal for the Philosophy of Science* 25: 178–188.
——— (1974b), "Popper's Qualitative Theory of Verisimilitude," *British Journal for the Philosophy of Science* 25: 166–177.
Munz, P. (1985), *Our Knowledge of the Growth of Knowledge: Popper or Wittgenstein?* London: Routledge.
O'Hear, A. (1980), *Karl Popper*. London: Routledge.
Poincaré, H. (1902), *La science et l'hypothèse*. Paris: Flammarion.
——— (1905), *La valeur de la science*. Paris: Flammarion.
Popper, K. R. ([1935] 1959), *The Logic of Scientific Discovery*. London: Hutchinson. Originally published as *Logik der Forschung*. Vienna: Julius Springer Verlag.
——— (1945), *The Open Society and Its Enemies* (2 vols.). London: Routledge.
——— ([1957] 1961), *The Poverty of Historicism* (2nd ed). London: Routledge.
——— ([1963, 1965] 1969), *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.
——— (1972), *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.
——— (1976), *Unended Quest: An Intellectual Autobiography*. London: Fontana.
——— (1992), *In Search of a Better World: Lectures and Essays from Thirty Years*. Translated from German by Laura J. Bennett, with additional material by Melitta Mew. London: Routledge.
——— (1994), *The Myth of the Framework: In Defence of Science and Rationality*. Edited by M. A. Notturno. London: Routledge.
——— (1997), *The Lesson of This Century: With Two Talks on Freedom and the Democratic State*. Interviewed by Giancarlo Bosetti. Translated from Italian by Patrick Camiller. London: Routledge.
——— (1998), *The World of Parmenides: Essays on the Presocratic Enlightenment*. Edited by A. Petersen and J. Majer. London: Routledge.
——— (1999), *All Life Is Problem Solving*. Translated from German by Patrick Camiller. London: Routledge.
Popper, Karl, and John C. Eccles (1977), *The Self and Its Brain: An Argument for Interactionism*. Berlin: Springer Verlag.
Quine, Willard Van (1963), *From a Logical Point of View* (2nd ed., rev). New York: Harper and Row.
Schilpp, P. A. (ed.) (1974), *The Philosophy of Karl Popper* (2 vols.). LaSalle, IL: Open Court.
Simkin, C. (1993), *Popper's Views on Natural and Social Science*. Leiden, Netherlands: Brill.
Stokes, G. (1998), *Popper: Philosophy, Politics, and Scientific Method*. Cambridge, UK: Polity.
Tichý, P. (1974), "On Popper's Definitions of Verisimilitude," *British Journal for the Philosophy of Science* 25: 155–160.

See also **Cognitive Significance; Corroboration; Demarcation, Problem of; Duhem Thesis; Evolutionary Epistemology; Induction, Problem of; Logical Empiricism; Lakatos, Imre; Quine, Willard Van; Verisimilitude**

POPULATION GENETICS

Evolutionary population genetics is the study of the dynamics of change in the genetic constitution of populations. The discipline grew out of the need to establish the Darwinian theory of evolution by natural selection on Mendelian hereditary principles. Prior to 1900, the view that the diversity of life arose from common ancestry was widely accepted in the scientific community, but Darwin's hypothesis that natural selection was the main mechanism of descent with modification was controversial (see Bowler 1983). This was due in part to Darwin's confused views about heredity. If, as Darwin thought, any character in the offspring was a blend of the corresponding characters in the parents, then in every generation, the character would regress toward the mean. This would make natural selection ineffective at generating change beyond the "sphere of variation" of the species (Jenkin 1867). So, for Darwin's hypothesis to be vindicated, it was necessary to establish a theory of heredity according to which variation was not lost in every generation. Mendelism, rediscovered in 1900, met this need. Unfortunately, Mendelism was not immediately accepted. There was a disagreement between two competing schools of thought in the early 1900s on the nature of heredity and of evolutionary change. Two critics in particular, the biometricians Pearson and Weldon, accepted Darwin's claim that changes due to selection were gradual and that selection acted on variations in quantitative characters, or characters like height or weight (Provine 2001). In contrast, Mendelians by and large rejected

Darwin's claim of gradual change. Their theory of heredity focused primarily on qualitatively varying characters, or characters like color or shape. On the whole, Mendelians held that evolution was the result of selection acting on major mutations, and not gradual selection on slightly varying traits. Gradually, however, biologists came to accept that Darwinian gradual selection was compatible with a Mendelian theory of inheritance. The development of a quantitative theory of evolution relying upon Mendelian principles of inheritance (population genetics) was crucial to the acceptance of Darwin's hypothesis that natural selection played a significant role in evolution and thus in generating the diversity of life.

The early population geneticists, R. A. Fisher, J. B. S. Haldane, and Sewall Wright, used primarily single-locus algebraic models to describe changes at the population level (see Evolution for an example). These were *prospective* models—given a set of values for selective parameters, migration rates and mutation rates, equations could be solved indicating, for instance, the rate at which evolutionary change would occur, or predicting genotype frequencies from one generation to the next. The above parameters describe deterministic factors effecting change in allelic frequencies. However, allelic frequencies change because of purely random factors as well. Fisher (1918) was the first to use diffusion methods to consider the stochastic changes in gene frequencies arising in finite populations, and Wright (1931) made "drift" a factor in his overall evolutionary theory. *Drift* refers to the random changes in gene frequency brought about by the random sampling of genes from one generation to the next, that is, the chance survivorship and reproduction of individuals irrespective of their fitness relative to their cohort. Any population that is sampled from one generation to the next will show some shift in distribution of characters due to chance alone. The effects of drift are accelerated in smaller populations; that is, the smaller the population, the more quickly will random sampling tend to make a population homogeneous, or uniformly of one or another genotype. In sum, prospective models describe how allele frequencies may change as a result of five different factors: mutation, migration, assortative mating, drift, and selection (cf. Haldane 1924). (Genotypic frequencies can change with or without changes in allelic frequency; for instance, as a result of assortative mating, inbreeding, and [in multilocus systems] recombination between gene loci.) Since the 1950s, multilocus models have been developed, which represent the change from one generation to the next at two or more loci. For the most part, however, many evolutionary questions can be answered using simple one-locus models. In the past twenty-five years, *retrospective* or "coalescent" models have been developed to assist in drawing inferences about the history of some lineage. Given some DNA sequence data, one can use retrospective models to answer questions like: "Given this information, when did the most recent common female ancestor of all humans alive today live?"

Evolutionary population genetics is but one part of population genetics generally. The mathematical component of population genetics is used not only in the evolutionary context but also in plant and animal breeding theory and in theoretical aspects of human genetics, especially in the search for the chromosomal location of disease genes. This article will focus on evolutionary questions and their mathematical analysis. First, there will be a summary of several of the major results of early population genetics theory and some of its more controversial aspects. Second, there will be an overview of some recent developments in population genetics—in particular, the influence of developments in molecular biology on theoretical population genetics. In conclusion, there will be a brief discussion of the scope and limitations of modeling in evolutionary genetics more generally.

The History of Population Genetics

The Hardy-Weinberg Law and the Maintenance of Variation

As mentioned above, the Darwinian theory of evolution by natural selection requires genetic variation. Variation is ultimately caused by mutation and subsequently also by chromosomal rearrangements, but it must be preserved for long periods for natural selection to act. The hereditary theory assumed by Darwin, that the characteristic of any child is in some sense a blend of that characteristic in the two parents, leads to rapid dissipation of variation. Thus, the very variation needed by the Darwinian theory is not supplied by the hereditary mechanism that he assumed. The Mendelian hereditary mechanism was rediscovered some forty years after the publication of *On the Origin of Species* and seventeen years after Darwin's death. Not only did this prove to be the correct hereditary model: It was one of the early triumphs of the mathematical theory to show that the Mendelian hereditary system is a variation-preserving one. Indeed, Mendelism supplies possibly the only hereditary mechanism maintaining the variation that is necessary for the Darwinian theory to work.

Weinberg and Hardy independently established the "law of panmictic equilibrium," today known as the Hardy-Weinberg law or principle. The law might be better described as a neutral or equilibrium model—a mathematical derivation starting from assumptions (some known to be false) for the purposes of evaluating the baseline state of a Mendelian system absent perturbing forces. Interestingly, the consequences of the segregation law were issues that Mendel himself explored but did not follow through to the case of random mixing—he was working with self-pollinating plants, and his "law of disjunction" treated only the case of reversion to type. In 1902, and later in 1903, Yule and Pearson independently examined the consequences of Mendel's law of segregation for a randomly mating population. However, their examinations of the question were yet again specific to a case in which the two factors were at the same initial frequency.

In 1908, Punnett, then a geneticist at Cambridge, asked Hardy, a mathematician, to derive the consequences of Mendel's laws for a randomly breeding population. Hardy demonstrated that *whatever the genotype frequencies* might be in a population, stable frequencies will result after one generation of random mating. The significance of this result is that given a particulate, or Mendelian, system of heredity, variation will be maintained in a population. The initial genotype frequencies in a population will remain unchanged from one generation to the next. This simple consequence of Mendel's law had been discovered a few months earlier by Weinberg. The derivation is as follows.

First, assume a diploid organism, sexual (or hermaphrodite) reproduction, nonoverlapping generations, perfectly random mating (no assortative mating), infinite population, and no migration, mutation, or selection. Let the two alleles at a locus be A and a. Suppose that in any generation the proportions of the three genotypes AA, Aa, and aa are P, Q, and R, where $P + Q + R = 1$. Correspondingly, let the frequencies of the A allele equal p, where $p = (2P + Q)/2 = P + Q/2$, and, for the a allele, the frequency $q = (2R + Q)/2 = R + Q/2$. The frequency of matings of $AA \times AA$, given random pairing of individuals, will be P^2. Likewise, the probability of an $AA \times Aa$ mating is $2PQ$, and the probability of an $Aa \times Aa$ mating is Q^2. Only these three matings can produce AA offspring, and they do so with respective probabilities 1, $\frac{1}{2}$, and $\frac{1}{4}$. Table 1 lists the genotypic frequencies resulting from each mating.

Table 1. Frequencies of offspring genotypes in a randomly mating population

Mating	Frequencies of mating	Offspring genotype frequencies		
		AA	Aa	aa
$AA \times AA$	P^2	1	0	0
$AA \times Aa$	$2PQ$	$\frac{1}{2}$	$\frac{1}{2}$	0
$AA \times aa$	$2PR$	0	1	0
$Aa \times Aa$	Q^2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$Aa \times aa$	$2QR$	0	$\frac{1}{2}$	$\frac{1}{2}$
$aa \times aa$	R^2	0	0	1

It follows that the frequency of AA offspring after one generation will be:

$$\begin{aligned} P' &= P^2 + \frac{1}{2}(2PQ) + \frac{1}{4}(Q^2) \\ &= (P + Q/2)^2 = p^2. \end{aligned}$$

Similarly, the frequency of Aa and aa after one generation will be:

$$\begin{aligned} Q' &= \frac{1}{2}(2PQ) + 2PR + Q^2/2 + 2QR/2 \\ &= 2(P + Q/2)(R + Q/2) = 2pq \\ R' &= Q/4 + 2QR/2 + R^2 = (R + Q/2)^2 = q^2. \end{aligned}$$

Thus, the frequency of each genotype after one generation of random mating will be p^2, $2pq$, and q^2. Replacing the values P', Q', and R', in the above equations, in order to determine the values for P'', Q'', and R'' in the subsequent generation, the same frequencies result. In other words, the genotype frequencies obtained after one generation of random mating are maintained in all subsequent generations. Thus, Hardy and Weinberg demonstrated that given the assumptions above, after one generation of random mating, stable genotype frequencies will result and be maintained. The key point here is that if there is no action by external forces (selection, mutation, migration, or random drift), then variation will be preserved in a population. This simple mathematical demonstration of the consequences of Mendel's law on the assumption of random mating thus answers one of the longstanding objections to Darwinism, *viz.*, that given a blending theory of inheritance, the variation needed for evolution through natural selection would rapidly be dissipated (Jenkin 1867). In contrast, under a Mendelian or particulate scheme of inheritance, variation will be preserved, *ceteris paribus.*

The Correlation Between Relatives

The Mendelian theory did not win immediate acceptance upon its rediscovery in 1900. One reason why it was not accepted quickly was that it was widely felt that biometrical data, including in particular the correlation between parent and offspring for characters such as height and weight, could not be explained on Mendelian grounds. Fisher (1918) showed not only that the broad pattern of these correlations could be explained assuming a Mendelian hereditary system, but that the numerical values for the correlations could also be explained. A Mendelian system of inheritance had to account for the observations of the normal distribution of most quantitative characters (e.g., height, weight) and the measurements of correlations between relatives with respect to these same characters. Fisher's (1918) paper showed that Mendelism did just that. First, by assuming that the character value for the heterozygote could be halfway between those of the two homozygotes, that the relevant Mendelian factors were entirely independent in their effects, and that the number and effects of such factors affecting any particular trait were quite large, Fisher showed how a normal distribution of measurements of some trait followed from a particulate scheme of inheritance. Second, and more significantly, Fisher demonstrated the consistency of the biometricians' observation of correlations between continuously varying traits and the Mendelian theory. There is no doubt that Fisher's specific genetical models were simplified. However, by showing that a reasonable fit to the observed correlations could be obtained under the Mendelian scheme, Fisher's work was a major force in leading to the acceptance of that hereditary scheme. A model does not have to be too precise to be useful.

The Fundamental Theorem

Having fused the biometrical and Mendelian viewpoints, Fisher then tried to establish general principles of evolution as a Mendelian process. Perhaps the best-known of these is his *fundamental theorem of natural selection:*

> The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time. (Fisher 1930, 37)

Careful attention to Fisher's intended meaning shows that although this is a true theorem, its significance is perhaps more circumscribed than Fisher claims. With the fundamental theorem, Fisher believed that he had discovered a universal generalization akin to the second law of thermodynamics. He believed the theorem to be a law of nature. Fisher (1930) describes his object in the opening pages of his chapter on the fundamental theorem:

> [t]o combine certain ideas derivable from a consideration of the rates of death and reproduction in a population of organisms with the concepts of a factorial scheme of inheritance, so as to state the principle of natural selection in the form of a rigorous mathematical theorem, by which the rate of improvement of any species of organisms in relation to its environment is determined by its present condition. (22)

Despite the appearance of progressivist language here, the fundamental theorem is not a statement about the unending or necessary adaptation of the species to its environment, but an expression of a fundamental relationship between the reservoir of genetic variation available and accessible to selection and the rate of increase in fitness in a population. Fisher was well aware that genetic interactions, rapid changes in or deterioration of the environment, overpopulation, and many other factors could affect whether or not a population of organisms would increase in numbers or continue to adapt over time. The fundamental theorem is thus a statement not of the necessary improvement of the species, but about the relation between genetic variance in some trait and increase in numbers of individuals possessing such a trait.

What Fisher's demonstration actually shows is simply that the additive variance in fitness (or that portion of the genetic variance that contributed to the correlation of relatives) is equal to that component of the increase in mean fitness in the population brought about by changes in gene frequencies only. This change was called the "partial change" by Ewens (1989), following a clarification of the meaning of the theorem by Price (1972). However, almost all commentators, starting with Wright (1930), have misunderstood the meaning of the theorem. Wright, for example, "corrected" the theorem as follows: "The total variance in fitness of a population is ascribable to the variance in fitness due to natural selection, which excludes the effects of dominance, epistasis, mutation, migration, change in environment, and drift." Subsequent commentators, and indeed the majority of textbooks in population genetics through the 1970s (Li 1955; Moran 1962; Crow and Kimura 1970; Jacquard 1974), misinterpreted Fisher's theorem along the same lines. The "received" interpretation thus came to be that "the increase in mean fitness of a population is approximately the current additive genetic variance in fitness, and this is non-negative"(Edwards 1994). This takes the theorem to refer to the mean fitness of the population and to be an approximate result. However, Price (1972), Ewens (1989), and Lessard (1997) have shown that the theorem, as correctly interpreted, is exact, not approximate.

Wright Versus Fisher

A continuing point of controversy in population genetics theory is the relative significance of two different models of the evolution of adaptation: Wright's and Fisher's. According to Fisher, evolution takes place for the most part in large, panmictic populations, and the factor of greatest significance in shaping adaptation is selection acting on alleles, even those with small selective effects. According to Wright, the landscape of gene combination consists of multiple adaptive peaks, separated by maladaptive valleys, or gene combinations that are less fit. The most effective means of traversing such peaks is via a three-phase process of isolation of small subpopulations and intrademic and interdemic selection. Wright called this process the "shifting balance model" of evolution.

The diagnosis and resolution of this controversy is contentious. Some argue that at the core are differing views about the nature and extent of genetic interaction, tied to the presuppositions behind Wright's model of the adaptive landscape (Whitlock et al. 1995). If indeed genetic variation is held tightly in "balance," or if there are many epistatic interactions for fitness, then it would seem that a mechanism like shifting balance is necessary for populations to move from suboptimal, or lower, to higher peaks in the adaptive landscape. On the other hand, it may be the case that whatever the extent of epistatic interactions for fitness, populations may always find "ridges" to traverse adaptive valleys via selection. For instance, assortative mating may permit the traversion of valleys (Williams and Sarkar 1994).

Others argue that the core of the divide between Wright and Fisher has to do with the rather delicately timed balance of isolation, selection, and migration Wright requires for shifting balance to go forward. In particular, it seems unduly restrictive to expect no migration between demes for the time necessary for them to diverge significantly for there to be a difference in fitness between them, followed suddenly by migration. The controversy over the shifting balance model continues today. Coyne, Barton, and Turelli (1997), neo-Fisherians, and Wade and Goodnight (1998), neo-Wrightians, continue to debate the extent of empirical support and the interpretation of mathematical and metaphorical models such as Wright's adaptive landscape.

The Introduction of Molecular Biology and the Neutral Theory

In the mid-1960s, molecular methods were introduced into the study of evolution. Protein sequencing revealed that the number of amino acid substitutions among species increases approximately linearly with time since divergence (Zuckerkandl and Pauling, 1962). Electrophoretic studies by Lewontin and Hubby (1966) demonstrated that there was a great deal of genetic variation at the protein level within natural populations. These observations eventually led to Kimura's (1968) proposal of the neutral theory of molecular evolution, which says that most changes detected at the molecular level were not acted upon by natural selection, but were neutral with respect to selection (that is, did not affect fitness). Kimura's reasoning was as follows. First, he examined molecular data on the variation among hemoglobins and cytochromes *c* in a wide range of species. Second, he calculated the rates of change of these proteins. Third, he extrapolated these rates to the entire genome. When he saw the rapidity of change that this implied, Kimura concluded that there simply could not be strong enough selection pressures to drive such rapid evolution. He therefore hypothesized that most evolution at the molecular level was the result of random processes like mutation and drift. Kimura called this hypothesis the *neutral theory of molecular evolution.*

Kimura's theory met with a great deal of controversy, as many interpreted it to run counter to the neo-Darwinian view that selection was the main agent of evolutionary change. This false impression was exacerbated by a paper published immediately after Kimura's by King and Jukes (1969), defending roughly the same thesis. Many were led to the mistaken view that the neutral theory denies the fact of adaptive evolution. However, it simply states that a large quantity of the turnover of molecular variation within populations has nothing to do with adaptation—it is simply neutral with respect to selection; that is, it has no effect on an organism's survival.

Today, most biologists accept that there is a great deal of molecular variation that is neutral with respect to selection. However, the rate of sequence change in evolution varies considerably with the DNA region examined. The more important the function of the region, the lower the rate of sequence change, as would be expected (Nei 1987). Much of systematics today uses the rapid turnover of some relatively neutral regions, such as the region that controls cytochrome *c*, to reconstruct phylogenetic relationships.

Kimura was in part inspired by the work of Sewall Wright, and in particular by Wright's emphasis on drift as a significant factor in evolutionary change. It should be noted, however, that it is a confusion to equate the neutral theory with Sewall

Wright's view on adaptive evolution in populations, the shifting balance theory. In other words, there may well be a great deal of turnover in a population at the molecular level, whether or not selection or drift is the main force changing the genotypic constitution of such a population over the long term. "Random drift" in the classical sense refers to chance fluctuations in the genetic constitution of a population, or sampling error. By chance alone, some individuals, irrespective of their selective advantages, may not survive to reproduction. In this way, one or another allele may become fixed in a population, irrespective of its selective advantage (or disadvantage). Reduction in population size accelerates the effects of drift in the sense that the average time to fixation of an allele is shorter, the smaller the population. So, over the short term, and in smaller populations, drift will be of greater significance in any population relative to selection. The neutral theory simply describes the turnover of sequence at the molecular level. Changes in some loci are effectively neutral with respect to selection, so there is significant turnover at the molecular level in such loci. This does not preclude that over the longer term, the effect of selection may significantly change the constitution of populations.

Retrospective Models, Molecular Genetics, and Coalescence Theory

Much of the early modeling in population genetics theory was prospective: Given certain fitness values, mutation rates, etc., equations could be solved indicating the rate at which evolution could occur. In the early years of this century such an analysis was needed, largely to support the Darwinian theory. But the Darwinian theory is now in effect accepted, and with the information provided by DNA sequence data, theory and modeling have branched into a retrospective analysis, as noted above.

Coalescent theory uses mathematical models and molecular data to determine times since most recent common ancestor of different lineages, or time to *coalescence.* While the mathematical demonstration of coalescence theory is beyond the scope of this article, the following are some basic premises of coalescence theory. All genes in a population ultimately trace their way back to a single ancestor gene, so that their ancestry coalesces at that gene. However, the allelic types of the genes in the population might differ from that of this common ancestor, because of mutation. These mutational differences help answer questions about the size, structure, and history of populations.

Coalescence theory assumes, in part for reasons of simplifying the mathematics, that most changes in the genome are neutral, so that most of the changes seen are a result of drift. This is in fact a very reasonable assumption when the scope of investigation is shorter time frames, or changes in populations over thousands, as opposed to tens of thousands, of years. For shorter time frames, the effects of drift will predominate. With the development of molecular methods, and of coalescence theory, there has thus been a shift in focus of the models of evolutionary genetics from longer to shorter time frames, and in these models, the significance of selection relative to drift will be negligible. For longer-term evolutionary questions of the sort that interested Wright and Fisher, selection will, relatively speaking, be a more significant factor, changing the genetic constitution of populations.

Conclusion

Theoretical population geneticists use mathematical models to investigate the dynamics of evolutionary change in populations. Essentially, they describe and explain the conditions on the possibility of evolution. Thus, population genetics constitutes that theoretical core of evolutionary biology. While the mathematical models of theoretical population genetics are necessarily idealized, they nonetheless constitute a useful tool for describing the main mechanisms of evolutionary change and answering questions about the relative significance of this or that factor in evolution, under some description of initial conditions. Rather than attempt to capture all the subtleties of inheritance (cytoplasmic as well as nuclear), development, and gene expression, classical population genetics treats evolution as simply change in allele frequency. In the context of investigating loci that contribute to disease, classical Mendelian models represent the disease of interest as a product of a single allele that is either dominant or recessive. Of course, many loci contribute in the expression of most diseases (and most traits), and the same allele may be expressed differently in different genetic contexts. Given what is known now about the nature and extent of genetic interaction, one may think that Mendelian "beanbag" genetics is obsolete (Provine 2001).

To the contrary, simplified treatment is necessary—first, because the complexity of the genetics of evolving populations ensures that a completely accurate description of reality is impossible; and second, were such descriptions possible, they would be mathematically intractable (see Crow 2001 for further discussion). Theoretical population genetics gives mathematically tractable ways to begin to describe the evolutionary process. Such models are in some sense idealizations, but they are useful tools to answer many simple questions, providing a framework for looking at phenomena that often take place over the lifetimes of many individual scientists.

As Wimsatt (1987) has pointed out, null (or false) models can be enormously useful tools for arriving at true theories. For instance, oversimplified models may serve as the starting point in a series of models of greater complexity and realism and may provide a simpler arena for answering questions that would be impossible to answer in more complex models. Or false models may describe extremes of a continuum in which the real case is presumed to lie (Wimsatt 1987, 30–31). For example, the neutral theory claims that all change at the molecular level is neutral, but using the neutral theory as a null model, biologists have now found that different regions of the genome turn over at different rates, indicating that the truth lies somewhere in between the continuum of complete neutrality and selection at every locus. Coalescent theory assumes that all change is neutral, but this strictly false assumption allows biologists to determine times since divergence of modern taxa.

Population genetics models, despite their many simplifications of genetic systems, provide real insight not otherwise obtainable into the evolutionary process. Such models may enable one to describe the common features of many systems that all differ in detail, determine how varying outcomes depend on the relative magnitude of one or another parameter, and decide which factors may legitimately be ignored, given the question or time frame under consideration. For example, population genetics theory shows that selection is more effective than drift in populations of large size (where $4Ns \gg 1$), whereas the effects of drift will overpower those of selection when the opposite is the case.

One may use such models to conclude which outcomes are very unlikely or impossible given some initial conditions. And mathematical analysis may serve to generate conclusions that could not be arrived at by empirical research at a given stage of inquiry. Lewontin (2000) describes the role of modeling in population genetics as delimiting what is possible and what is prohibited in microevolutionary change. And some biologists have extended the use of these models to answer questions about change above the species level, such as in speciation (e.g., Barton and Charlesworth 1984).

Exact prediction in population genetics is a near impossibility. While the models of Newtonian mechancs may be used to predict the motions of the planets or the trajectory of a projectile here on Earth with a high degree of accuracy, one ought not to expect this sort of predictive power from population genetics models. While one might hope that models of biological evolution will help with short-term predictions, one cannot hope that they will lead to explicit long-term predictions. In effect, the allele passed on by a parent to a child at any locus results from a random choice of one of the two alleles that the parent has at that locus, and one can never know which one this will be for each gene in each individual. At best, one may predict trends, given initial population sizes and rates of mutation and migration. Thus, theoretical population genetics is an irreducibly probabilistic theory. However, population genetics theory provides a rigorous way to determine the relative significance of different factors over long time frames in the changes of the genetic constitution of populations.

Which simplifications to employ in modeling a biological system will depend upon the context and the question at issue. For example, when considering the effects of geographical dispersal, it might be reasonable to assume only one sex; and when addressing questions asking why sexual dimorphism exists, it might be reasonable to ignore geographical distribution. In general, the problem of finding a balance between a model's being sufficiently complex to describe reality adequately and at the same time being sufficiently simple to allow a mathematical analysis is not only a question of philosophical interest, but also a serious one faced in the everyday practice of theoretical biology.

Anya Plutynski
Warren J. Ewens

References

Barton, N. H., and B. Charlesworth (1984), "Genetic Revolutions, Founder Effects, and Speciation," *Annual Review of Ecology and Systematics* 15: 133–165.

Bowler, P. (1983), *The Eclipse of Darwinism: Anti-Darwinian Evolution Theories in the Decades around 1900*. Baltimore, MD: Johns Hopkins University Press.

Coyne, J., C. Barton, and M. Turelli (1997), "Perspective: A Critique of Sewall Wright's Shifting Balance Model of Evolution," *Evolution* 51: 306–317.

Crow, J. (2001), "The Beanbag Lives On," *Nature* 409: 771.

Crow, J. F., and M. Kimura (1970), *An Introduction to Population Genetics Theory*. New York: Harper & Row.

Edwards, A. W. F. (1994), "The Fundamental Theorem of Natural Selection," *Biological Review* 69: 443–475.

Ewens, W. (1989), "An Interpretation and Proof of the Fundamental Theorem of Natural Selection," *Theoretical Population Biology* 36: 167–180.

Fisher, R. A. (1918), "The Correlation Between Relatives on the Supposition of Mendelian Inheritance," *Transactions of the Royal Society, Edinburgh* 52: 399–433.

——— (1930), *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Haldane, J. B. S. (1924), "The Mathematical Theory of Natural and Artificial Selection, Part I," *Transactions of the Cambridge Philosophical Society* 23: 19–41.

Jacquard, A. (1974), *The Genetic Structure of Populations*. Berlin and New York: Springer-Verlag.

Jenkin, F. ([1867] 1973), "*Review of On the Origin of Species," in D. L. Hull (ed.), Darwin and His Critics. Cambridge, MA: Harvard University Press, 303–350. Originally published in North British Review 44: 277–318.*

Kimura, M. (1968), "Evolutionary Rate at the Molecular Level," *Nature* 217: 624–626.

King, J., and T. H. Jukes (1969), "Non-Darwinian Evolution," *Science* 164: 788–798.

Lessard, S. (1997), "Fisher's Fundamental Theorem of Natural Selection Revisited," *Theoretical Population Biology* 52: 119–136.

Lewontin, R. C. (2000), "*What Do Population Geneticists Know and How Do They Know It?*" in R. Creath and J. Maienschein (eds.), *Biology and Epistemology*. Cambridge: Cambridge University Press.

Lewontin, R. C., and J. L. Hubby (1966), "A Molecular Approach to the Study of Genetic Heterozygosity in Natural Populations II: Amount of Variation and Degree of Heterozygosity in Natural Populations of Drosophila pseudoobscura," *Genetics* 54: 595–609.

Li, C. C. (1955), *Population Genetics*. Chicago: Chicago University Press.

Moran, P. A. P. (1962), *The Statistical Processes of Evolutionary Theory*. Oxford: Clarendon Press.

Nei, M. (1987), *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Pearson, K. R. (1898), "Mathematical Contributions to the Theory of Evolution: On the Law of Ancestral Heredity," *Proceedings of the Royal Society* 62: 386–412.

Price, G. R. (1972), "Fisher's 'Fundamental Theorem' Made Clear,"*Annals of Human Genetics* 36: 129–140.

Provine, W. (2001), *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press.

Wade, M., and C. J. Goodnight (1998), "Perspective: Theories of Fisher and Wright in the Context of Metapopulations: When Nature Does Many Small Experiments," *Evolution* 52: 1537–1553.

Whitlock, M., P. C. Phillips, F. B. C. Moore, and S. J. Tonsor (1995), "Multiple Fitness Peaks and Epistasis," *Annual Review of Ecology and Systematics* 26: 601–629.

Williams, S. M., and S. Sarkar (1994), "Assortative Mating and the Adaptive Landscape," *Evolution* 48: 868–875.

Wimsatt, W. (1987), "*False Models as Means to Truer Theories*," in H. Nitecki and A. Hoffman, (eds.), *Neutral Models in Biology*. Oxford: Oxford University Press, 23–55.

Wright, S. (1931), "Evolution in Mendelian Populations," *Genetics* 16: 97–159.

Zuckerkandl, E., and L. Pauling (1966), "*Evolutionary Divergence and Convergence of Proteins*," in V. Bryson and H. J. Vogel (eds.), *Horizons in Biochemistry*. New York: Academic Press, 189–225.

POSITIVISM

See **Logical Empiricism**

PREDICTION

Whether one predicts rainfall, recessions, or racetrack winners, predicting an event or state of affairs often, perhaps even typically, involves saying that it will happen before it occurs, and this common association is presumably responsible for the idea that predictions must be about the future. But in scientific contexts one often characterizes a theory's predictions as its implications or entailments without regard for temporal constraints, as when one says a successful theory of cosmology predicts the existence of cosmic background radiation at all times. The language of prediction is also used to describe declarative assertions about past and present events made in light of a theory, as when evolutionary theory was used to predict that marsupial mammals must once have lived in what is now Antarctica and left fossilized remains there. A temporal element might be preserved by insisting that these are really cases of postdiction, retrodiction, or even shorthand predictions about future evidential findings. But perhaps these comfortable extensions of predictive language more naturally suggest that the central element in prediction is not temporal but epistemic. To predict is to make a claim about matters that are not already known, not necessarily about events that have not yet transpired.

Of course, prediction cannot be as simple as that, because one way to know something is to predict it correctly on the basis of a well-confirmed theory. Predictive language seems most appropriate when one makes claims about unknown matters using tools (like inductive generalization, scientific theorizing, or sheer guesswork) that can be contrasted with more direct methods of ascertaining the same information (like simply observing in the right place and/or at the right time and/or under the right conditions, or looking for physical traces of some past state of affairs). Although specific philosophical and scientific conceptions of what is immediately given in experience or known directly have shifted over time, predictive language has continuously respected the fundamental idea that a prediction is a claim about unknown matters of fact whose truth or falsity has not already been independently ascertained by some more direct method than that used to make the prediction itself (see Phenomenalism; Physicalism).

As this account suggests, successful prediction is valuable because it goes beyond what is already known most directly, but this same feature renders prediction inherently risky. The most interesting and useful predictions typically concern matters to which more direct intersubjective access is ultimately expected, so prediction is characteristically something that one can be caught out on given the shared standards of the community of inquirers.

This idea that scientific prediction involves risk led Karl Popper (1963) to single out the willingness to make risky predictions as what distinguishes genuine science from pseudoscience (see Popper, Karl Raimund). Pseudoscientific theories, he suggested, typically include the resources to explain any outcome in their intended domain of application after it is known. Marxist history, Freudian psychoanalysis, and Adlerian "individual" psychology were among Popper's favorite examples. He urged that such theories not be regarded as genuinely confirmed by passing tests that they could not possibly have failed. Confirmation, or for Popper "corroboration," requires that a theory succeed where it might have failed (see Confirmation Theory; Corroboration). Thus, Popper argued, genuine science requires theories that rule out some states of affairs and make risky predictions about unknown cases, exposing themselves to the serious possibility of refutation.

In empirical science, the requirement of shared epistemic access to the success or failure of a prediction means that the fate of a prediction is typically decided in the court of experiment and observation.

The Problems of Induction

The Scottish empiricist David Hume may have posed the problem of the rational justification for prediction in its starkest form. Hume's empiricism led him to regard the most general problem about knowledge to be how one comes to know anything whatsoever "beyond the present testimony of our senses, or the records of our memory" (Hume [1748] 1977, 16). Hume pointed out that the mere occurrence of one event or sense impression never deductively implies that another will occur. From this he concluded that it must be on the basis of experience that one learns which particular events reliably cause, precede, or are otherwise associated with others. One is thereby able to make predictions about events or states of affairs beyond those immediately perceived (see Empiricism).

But how can one possibly justify assuming that the regular associations or even causal relationships that have been noted between past events will persist into the future? Again there is no logical contradiction in supposing that things will change. That the sun will not rise tomorrow, Hume notes (15), is no less intelligible a proposition than that it will rise—indeed, the future will almost certainly be quite unlike the past in innumerable particular respects. And any attempt to justify this assumption by appeal to past experience of uniformity in nature, Hume claims, will be "going in a circle, and taking that for granted, which is the very point in question" (23). That the future has been like the past *in the past* constitutes evidence only about what one's own future will be like if one already assumes that how things have been in the past is a good guide to what they will be like in the future, which was the very assumption needed to justify the inferential practice in the first place (see Causality).

Efforts to solve or dissolve Hume's problem of induction are a topic of continuing debate (see Induction, Problem of). For his part, Hume concluded that there can be no rational justification whatsoever for predictions concerning unexperienced matters of fact, and he took this to illustrate that reason or rational justification does not play anything like the role usually supposed in the cognitive lives of human beings. In his skeptical solution to the problem, Hume argues that what generates expectations about unknown cases is a primitive or instinctive psychological disposition he calls *custom*, which is not itself mediated by any process of reasoning at all. Custom leads one, automatically and without reflection, to expect an event of type *B* on the appearance of an event of type *A* just in case *B*s have followed *A*s reliably in the past. Thus, Hume offers a naturalistic *explanation* of the psychological mechanism by which empirical predictions are made but not any rational *justification* for this practice. But this is not to say that it is a mistake to rely on custom: Not only do we have no choice in the matter, Hume ([1748] 1977) argues, but "[c]ustom . . . is the great guide of human life. It is that principle alone which renders our experience useful to us. . . . Without the influence of custom, we should be entirely ignorant of every matter of fact, beyond what is immediately present to the memory and senses" (29). The fact that there is no rational justification for such an important and useful cognitive function, he suggests, simply illustrates that nature has secured "so necessary an act of the mind, by some instinct or mechanical tendency" rather than leaving it "to the fallacious deductions of our reason" (37). The most central aspects of human cognitive lives, he suggests, are neither products of nor even subject to reason. Instead they are "a species of natural instincts, which no reasoning or process of the thought and understanding is able either to produce or to prevent" (30).

A further problem of inductive justification, arguably anticipated in Hume's treatment, is clearly articulated by Nelson Goodman (1954). Here the problem is not how to justify the belief that unexperienced cases will resemble experienced ones, but how to understand, categorize, or describe experienced cases so as to know just what it would be like for unexperienced cases to resemble them. Present inductive evidence fully supports the claim that all emeralds are green, for example, but it equally well supports the claim that they are all grue, where 'grue' means "green if first observed before 2050 and blue if not observed before 2050." Those who believe that emeralds are grue rather than green, however, will have expectations concerning the appearance of emeralds that diverge significantly from the customary one starting in 2050. Nor can one say that the predicate 'grue' is somehow artificially conjunctive or really disguises a change, Goodman argues, for it is relative only to a set of predicates that regards green and blue as natural categories that it does so. If one takes 'grue' and, say, 'bleen' (understood as "blue if first observed before 2050, and green if not observed before 2050") as natural and primitive predicates of a language it will be "green" that must be defined in an artificially conjunctive way (i.e., "grue if first observed before 2050 and bleen if not"). But, of course, it was the choice of green and not grue as natural, primitive, or singularly appropriate for law-like generalization for which a defense was sought in the first place. Goodman thus argues that any attempt to use inductive evidence to project future or unknown cases relies on a set of entrenched predicates, and it is controversial whether the entrenchment of one set of predicates rather than another can be rationally defended. Like Hume's custom, Goodman's entrenchment may offer a kind of naturalistic explanation of how humans come to make the predictions they do, but not one that seeks or provides any rational justification for the practice.

Models of Empirical Prediction

Hume's empiricist approach to the foundations of knowledge proved attractive to such later theorists of science as the logical empiricists, many of whom held that the aim of empirical science was to determine the dependence of observable phenomena on one another; indeed, some famously insisted that every meaningful statement derived its meaning from its implications regarding observable phenomena (see Cognitive Significance; Verificationism). In this broad view, empirical predictions were required to be statements (i) in a specified observation language, (ii) entailed by one's theory together with one's past observations, and (iii) concerning unobserved but observable phenomena. It is important to recognize, however, that the logical empiricists did not always agree even among themselves about how to characterize the nature of empirical predictions. To take just one example of controversy, in Carnap's (1967) *Logische Aufbau der Welt*, the empirical predictions made by a scientific theory do not concern the "given" of sense experience but rather *structural* features of the intersubjective domain constructed from experience (see Carnap, Rudolf).

Carl Hempel's (1965) model of scientific knowledge was both deeply influenced by the earlier logical empiricist tradition and itself widely influential in turn. In the simplest, deductive-nomological case, predictions and explanations are logical deductions of the form

$$\frac{\begin{array}{l} C_1 \wedge C_2 \wedge \ldots \wedge C_k \\ L_1 \wedge L_2 \wedge \ldots \wedge L_r, \end{array}}{E}$$

where $C_1 \wedge C_2 \wedge \ldots \wedge C_k$ are statements of particular occurrences (e.g., the positions and momenta of certain celestial bodies at a time), $L_1 \wedge L_2 \wedge \ldots \wedge L_r$ are general laws (e.g., of Newtonian mechanics), and E is the sentence stating whatever is being explained, predicted, or postdicted (e.g., the time of the next solar eclipse), in Hempelian terms. Hempel (1965) also allows for what he calls inductive-statistical predictions, where the argument has the same basic form, but the laws invoked are statistical probability statements. Here a specific event is not logically implied by the boundary conditions and laws, but only supported to a certain degree (175–177). For Hempel, the conclusion of any argument of this form qualifies as a prediction if E refers to an occurrence at a time later than that at which the argument is offered. A fascinating and controversial feature of this account is the symmetry it asserts between prediction and explanation: To explain an event by appeal to a set of laws and conditions is simply to show that it could have been predicted using them (see Hempel, Carl Gustav; Explanation).

More recent accounts of empirical prediction have moved progressively away from the logical empiricists' original requirement of a neutral "sense datum" language for reporting observation or representing experience. In van Fraassen's (1980) constructive empiricism, for example, presenting an empirical theory involves specifying a model for the language of the theory: a domain of objects together with a description of the properties they can have and the relations they can bear to one another. In presenting the theory, one also specifies those substructures of the model that are candidates for representing observable phenomena. The theory is empirically adequate just in case the appearances given in phenomenal experience are isomorphic to the observable substructures of the model (64) (see Empiricism; Instrumentalism). As in the empiricist tradition more generally, then, the distinction between observable and unobservable phenomena does significant work here, but this distinction is not drawn in linguistic terms. Rather, for van Fraassen, the distinction is supposed to be grounded in the actual observational capacities of human observers, and it is natural science itself that shows what those observational capacities are (see Phenomenalism; Perception).

The naturalistic suggestion that observability is a question to be settled by natural science is perhaps promising. But how could one's best theories determine what is observable? If they characterize important features of the natural world and one's place in it, then they also might be expected to specify how and the circumstances under which reliable inferences from measurements are possible for human observers. It is presumably in just those circumstances for which one's theories indicate that measurements will provide the resources for reliable inferences about the presence or absence of some entity that one is inclined to characterize the entity as observable. In such a naturalistic view, an empirical prediction might in principle concern any feature of the world that one's best theories indicate can be reliably detected.

But herein also lies a problem for the naturalist. What one judges to be observable will depend on one's current best understanding of the natural world, but this best understanding will itself depend on what one believes one has observed. Since the naturalist's account of what is observable itself depends on the theories the naturalist accepts, observations cannot test the truth or falsity of theories in any direct or simple way. As Quine (1951) and others have noted, one can always respond to a failed test of a theory by blaming background assumptions, presumably including the assumptions used to characterize what empirical observations are and the conditions under which they can be reliably made, rather than admitting that a particular prediction was mistaken. But if empirical predictions need never be given up, then they cannot, strictly speaking, test the theory that makes them (see Quine, Willard Van).

In practice, however, this general epistemic problem is more often a point of logic rather than a real obstacle to naturalistic inquiry, as Quine himself noted in developing his own naturalistic position. Testing a given empirical prediction to the satisfaction of the scientific community requires only that there be a sufficient context of shared background assumptions to provide the understanding of and rules for the empirical test. The understanding and rules might be implicit, change over time, and be subject to challenge, but none of this undermines the possibility of testing predictions in principle and, consequently, the possibility of testing the theories that make them. That empirical predictions are in fact often taken by the scientific community to be thoroughly tested and that theories are in fact accepted or rejected on this basis suggest that there are often, perhaps typically, unambiguous standards for checking them.

The Epistemic Significance of Prediction

As the preceding discussion of the relationship between theories and their predictions suggests, testing a theory's predictions is often taken to be a crucial aspect of how it is confirmed or disconfirmed. The most persistent question here concerns whether the ability to predict *novel* phenomena is of fundamental significance in the testing and confirmation of specific theories in the special sciences, that is, whether it counts in favor of a theory's confirmation that it has predicted novel phenomena rather than merely accommodating, explaining, or anticipating phenomena already known to occur. In this context the relevant sense of prediction involves not anticipating when and where familiar phenomena will recur but rather discovering the existence of phenomena unlike those that are already familiar.

The roots of this debate reach back at least to the foundations of modern science itself; perhaps its most famous iteration pitted William Whewell against John Stuart Mill, who expressed amazement at Whewell's view that

> an hypothesis . . . is entitled to a more favourable reception, if besides accounting for all the facts previously known, it has led to the anticipation and prediction of others which experience afterwards verified. Such predictions and their fulfillment are, indeed, well calculated to impress the uninformed. . . . But it is strange that any considerable stress should be laid upon such a coincidence by persons of scientific attainments. (*System of Logic,* III, xiv, 6, cited in Musgrave 1974, 2)

Mill's amazement notwithstanding, versions of this Whewellian intuition have been defended by "persons of scientific attainments" as diverse as Clavius, Descartes, Leibniz, Huygens, Peirce, and Duhem. By contrast, Mill defended the view that confirmation depends only on the match between a theory's entailments and the phenomena. While decidedly less popular, this competing view also recruited influential champions, such as John Maynard Keynes (Giere 1983, Sec. 3).

Enthusiasts have sometimes gone so far as to claim that only predictions of novel phenomena are of any confirmational significance at all or that any prediction of a novel phenomenon is of greater confirmational significance than any amount of accommodation of existing evidence. But the claim of a special confirmational significance for prediction does not require such extremes. For prediction as such to enjoy a special confirmational privilege, it seems sufficient that predicting a given phenomenon provides (or would have provided) greater confirmation for a theory that does so than the mere accommodation of *that same phenomenon* does (or would have). A view having this consequence, including the extremes just described, may be described as a form of *predictivism.* Predictivist themes have recently loomed large in debates over the progressiveness of research programs, the adequacy of various approaches to confirmation (especially Bayesianism), and the so-called miracle defense of scientific realism.

Imre Lakatos is widely credited with having reintroduced this concern over the confirmational significance of novel prediction, specifically in connection with his "methodology of research programs" (see Lakatos, Imre). Lakatos's bold claim was that it is *only* the ability of the successive theories in a research program to make successful novel predictions that bears on its progressiveness or acceptability. But even Lakatos's own work includes several competing lines of thought about the nature of novelty (Gardner 1982, 2–3). At times he seems to construe the novelty of a prediction for a theory purely temporally, though his most famous account holds novel prediction to consist in predicting phenomena that are "improbable or even impossible in the light of previous knowledge" (Lakatos 1970, 118), and he later accepted Zahar's (1973) revisionist proposal that the novelty of a fact for a hypothesis requires only that it "not belong to the problem-situation which governed the construction of the hypothesis" (103). Each of these lines of thought has been more fully developed by later thinkers, even as they have lost any immediate connection to concerns about the evaluation of research programs (see Research Programs).

The second issue of recent interest concerns whether standard philosophical approaches to confirmation can recognize a special confirmational significance for novel prediction, and if not, whether this weighs against such approaches to confirmation or against the legitimacy of predictivist intuitions instead. Such approaches are described as taking into account only "logical" and not "historical" relations between theory and evidence, or alternatively, only the content of theories and evidence and not historical facts about them. It has sometimes been claimed that a logical approach to confirmation is strictly inconsistent with predictivism; but this is too strong, for the fact of successful prediction can itself simply be treated as part of the evidence supporting a theory. In Bayesian terms, one need only treat the fact that a novel result was *predicted* as part of the evidence on which the theory's probability is conditionalized to allow a special confirmational role for novel prediction. This brute-force solution to the problem invites the complaint that a special epistemological significance for novel prediction still finds no expression in the formal *machinery* of either Bayesian or any other extant logical accounts of confirmation. But even this is far from uncontroversial (see Bayesianism; Confirmation Theory).

There has been widespread discussion among Bayesians concerning the nature and plausibility of the further assumptions that must be granted in order to accord novel prediction a special confirmational significance within the Bayesian framework. Central to this discussion has been Glymour's (1980) "problem of old evidence": The Bayesian approach to confirmation suggests that known evidence cannot provide any support for a theory because probability 1 is conferred on that evidence by background knowledge alone (see Bayes's theorem below). To make matters worse, it is difficult to see how one could conditionalize on or even specify what one's background knowledge "would have been" without the evidence in question. Indeed, it has variously been argued that Bayesianism is legitimate because it recognizes a special confirmational significance for novel prediction, that it is legitimate because it does not, that it is illegitimate because it does, and that it is illegitimate because it does not (Brush 1995).

Finally, it has sometimes been argued that the ability of a theory to make successful novel predictions is the one form of scientific success for which only the truth of a theory can provide any explanation. This grounds a specific form of the traditional miracle or explanationist argument for scientific realism on behalf of theories enjoying success in making novel predictions. Leplin's (1997) account of novel prediction, for example, is an explicit effort to pick out just those forms of scientific success that only the truth of the successful theory could explain (see Scientific Realism).

A related point of contention concerns whether predictivist convictions have in fact played any role historically in the confirmational judgments made by actual scientific communities. Theorists have appealed to such famous cases of novel prediction as the Poisson "bright spot" by Fresnel's formulation of the wave theory of light, the gravitational bending of light by the general theory of relativity, and the existence and properties of three new elements by Mendeleev's periodic law to argue that particular novel predictions have or have not been accorded exceptional confirmational weight by actual scientific communities relative to the mere accommodation of existing evidence (see Scerri and Worrall 2001 for references and discussion). Such claims about scientific practices are also invoked to either bolster or defuse the further claim that an adequate account of confirmation will have to respect predictivist intuitions. This historical debate serves to underscore the contentious character of the *explanandum* for which accounts of novel prediction are supposed to provide explanations. For even if it is true that scientific communities have not historically weighted novel predictions over other kinds of evidence, this itself would seem to call for some kind of explanation, in light of the grip that predictivist intuitions seem to hold on ordinary thinking about confirmation.

Indeed, any serious assessment of the epistemic significance of novel prediction seems to invite a stark conflict of powerful normative intuitions. There is something especially impressive about such famous cases of novel prediction as gravitational light bending and the Poisson bright spot, but it seems perfectly fair to ask why the temporal order or other historical circumstances of discovery should have any bearing on the confirmational significance of the evidence for a theory. After all, whether a phenomenon was already known does not have any impact whatsoever on how convincing the theory's account of that phenomenon is. Why should it make any difference whether the data were predicted by a theory or acted as a constraint on the development or selection of that same theory in the first place? Features such as the theory's fit to the data, the auxiliary assumptions required to obtain that fit, and the theory's intrinsic plausibility remain precisely the same whatever the order, manner, or other circumstances of their discovery. It seems perverse to treat such apparent historical accidents as relevant to the degree of confirmation conferred on that theory by the evidence at hand.

Predictivism's defenders have turned, therefore, to specifying criteria for genuinely novel prediction in a way that seeks to avoid dependence on apparently arbitrary or epistemically insignificant features. Meanwhile, their opponents have sought to show that the apparent significance of novel prediction is a product of its confusion, conflation, or frequent association with something else that is of genuine epistemic importance. It is, however, sometimes hard to see more than a rhetorical or terminological difference between the positions of those who seek to creatively refine the conception of novel prediction so as to guarantee its epistemic significance and those who seek to explain the apparent importance of novel prediction as dependent upon the genuine epistemic significance of something else altogether. Sometimes both camps appeal to the same or similar relationships between theory and evidence, and it is not always clear whether a given author even means to explain the epistemic significance of novel prediction or explain it away. A similar ambiguity infects the discussion of the confirmational role of novelty in actual historical cases.

Complicating this dialectical situation are two competing strands of thinking about the epistemic significance of novel prediction, whether real or apparent. The first, sometimes called "heuristic," holds that the epistemic significance of genuine novelty is a matter of the independence, in some sense, of a given result from the formulation of the theory for which it counts as a novel prediction. Various formulations count a given phenomenon as novel for a given theory if and only if it was not part of the problem situation that led to the theory (Zahar 1973), was not actually used in the formulation of the theory (Worrall 1978 and 1989), was not known to some theorist who formulated the theory (Gardner 1982), or fits the hypothesis despite its not having been designed for that purpose (Campbell and Vinci 1983). These accounts differ most centrally in the precise role that known data must play in the formulation of a hypothesis in order for it to lose the special confirmational significance associated with novel prediction.

The second approach, sometimes unfortunately called "epistemic," proposes instead that novelty be understood as a matter of unexpectedness or low probability in light of what is believed absent the theory. Examples include Lakatos's (1970) construal of novel prediction as the prediction of phenomena that are "improbable or even impossible in the light of previous knowledge" (118) and Musgrave's (1974) suggestion that a novel prediction of a theory is one that either conflicts with or is at least not also made by its competitors or predecessors (see also Popper 1963, 36). These accounts differ centrally over what should form the foundation for the expectations that a predicted phenomenon must violate in order for it to enjoy the exceptional confirmational significance associated with novel prediction, and thus recall the problem of old evidence for Bayesianism.

Each approach is sometimes motivated by canvassing weaknesses or challenges to various versions of the other, but it is worth noting that each encapsulates one of two quite different phenomena that might reasonably be called novel prediction. The first is aimed at a theory's entailment of a result not involved in its own development, that is, a result that is novel *for the theory*, while the second concerns a theory's prediction of phenomena unlike those with which an epistemic community is already familiar, that is, a novelty *for an epistemic community*.

Among theorists who seem inclined to argue that the apparent epistemic significance of novel prediction is a product of its confusion with some other condition of genuine epistemic significance, the most influential proposal has been that the evidence must provide a "severe test" of the theory, that is, one that the theory is likely to fail if it is in fact false (Popper 1963; Horwich 1982; Giere 1984; Mayo 1991). Other analyses propose alternative sources of confusion, such as the assurance novel prediction typically provides that the hypothesis be well supported by earlier subsets of the data as well as by the whole (Schlesinger 1987), that there be no opportunity for "fudging" the hypothesis to fit the data (Lipton 1991), or that the hypothesis itself not be an arbitrary conjunction of facts (Lange 2001).

One natural way to unify these divergent intuitions about the epistemic significance of novelty is to suggest that each is concerned to rule out a different possible explanation of the evidential situation that would undermine the support a given piece of evidence would otherwise provide for a given theory. The idea here is that the various attempts to define novelty and to explain or to explain away its confirmational significance appeal to different ways in which the *prima facie* support that evidence provides for a theory can be undermined by further information, such as finding out that the theory was constructed, manipulated, or chosen so as to yield its supporting data, that there are reasons besides the theory to expect the results reported in the data, that the theory itself is simply an arbitrary conjunction of unrelated facts, and so on. If this view is right, the various competing accounts of novel prediction reflect the variety of possible confirmation-undermining explanations of the evidential situation, and it was a mistake all along to insist on just a single criterion of "genuine" novel prediction or even a single analysis of its epistemic significance, real or apparent. This suggests that paradigmatic cases of novel prediction like the Poisson bright spot are particularly impressive precisely because they preclude nearly all of the most likely confirmation-undermining possibilities. In a similar spirit, Leplin's effort to pick out the sort of predictive success that could only be plausibly explained by the truth of a theory that enjoys it includes criteria of both heuristic and epistemic varieties.

Any such pluralistic proposal regarding the epistemic significance of novelty must face up to an argument given by Horwich (1982), who grants that the confirmation provided for a theory by a given result would be compromised by the existence of plausible competing explanations for that same result, and endorses a Bayesian version of the severe-tests conception of confirmational significance. But he denies that the explanations ruled out by the heuristic novelty of a result (e.g., that the theory was formulated or even manipulated so as to entail the result) are actually competitors to the theory itself as explanations *of the available data.* Instead, he suggests, the explanations ruled out by heuristic novelty are those of *why a given theory fits the data as well as it does*, and thus do not compete with the theory itself to explain the data at hand. For this reason, he suggests, neither heuristic novelty nor the explanations of our evidential situation that it is able to preclude carry any genuine confirmational significance.

But what about the latter explanatory demand alone? Even if the fact that a theory was formulated or manipulated so as to entail the data simply offers an explanation for why the theory fits the data, this would seem to compete with the alternative explanation that the theory fits the data because it is true. Horwich resists this suggestion by way of an intriguing analogy designed to illustrate that even different explanations of the same state of affairs need not always be genuine competitors. He points out that being out of gas and having a broken starter compete to explain why Smith's car will not start: Both could obtain, but the probability reasonably assigned to one will be dramatically lowered upon learning that the other is true, because they answer the same explanatory demand in the same way. But not all explanations of the same state of affairs compete in this way. Consider, he suggests, the following explanations for the fact that his car is green: Is it because he buys only green cars or because the previous owner painted it green? In this case, the candidate explanations do not compete, in the sense that the fact that one obtains does not reduce the probability that the other also obtains. He further suggests that the explanations of fit precluded by heuristic novelty and provided by the truth of a theory are like the second case and not the first: The fact that a theory fits the data because one requires or even manipulates the hypothesis to ensure that this is so simply does not compete with the explanation that the resulting hypothesis fits the data because it is true. If so, there is no prior reason to think that hypotheses that merely accommodate existing data are less likely to be true than those that successfully predicted the same data as novel phenomena.

It is far from clear that this claim must be accepted as it stands. Perhaps there is no competition if data simply constrained the formation of a theory in the first place. But if one manipulated a theory's variable parameters to get it to fit the data, and if one believed that such adjustment could have accommodated most any data of the kind in question, this might indeed seem to compete with the claim that the theory fits the data because it is true. But even if one accepts Horwich's claim, all need not be lost for heuristic novelty. It remains possible that one might find a promising *inductive* justification for the epistemic significance of heuristic novelty. But this would require making the case that theories successfully predicting heuristically novel phenomena go on to enjoy especially impressive track records.

How to Make Predictions

In the spirit of Hume's skeptical solution to the problem of induction, one might wonder whether questions about the rational justification of predictions are not best dealt with by investigating how successful predictions have in fact been made. What inferential techniques and assumptions are actually used to move from known facts to predictions about unknown cases?

It may help to start with simple cases. Hans Reichenbach took the aim of inductive inference to be that of finding series of events whose frequency of occurrence converges toward a limit (see Reichenbach, Hans). If this is all one wants, one might simply keep track of the relative frequencies in any series of data one finds interesting. Suppose, for example, one wants to determine the probability of tossing a coin and having it come up heads. Start tossing it, keeping track of the relative frequency of heads to all tosses. If there is a well-defined relative frequency in the limit as the coin is tossed interminably, then it would be guaranteed that the probability would be found in this way. If the limiting relative frequency is undefined, then, for Reichenbach, it makes no sense to assign probabilities to the possible outcomes of any toss of the coin, and there is no solution to the problem of induction for that particular series of events. That is, the world is predictable insofar as it is sufficiently ordered to enable one to construct limiting relative frequencies from empirical data. While Reichenbach admits that "we do not know whether the world is predictable," keeping track of relative frequencies in a world that is in fact predictable is guaranteed eventually to deliver the right probabilities, and this sort of inductive inference will work; and if the world is not predictable, then nothing will work. Moreover, if something besides this sort of inductive inference does reliably work to predict future events, then this sort of inductive inference would track the success of the alternative method and warrant its use (Reichenbach 1938, 350).

While raw relative frequencies sometimes are the best estimates of probabilities, several caveats are in order. Although Reichenbach's procedure is guaranteed to deliver the actual relative frequencies in the long run if there are any, one never in fact performs an infinite number of observations calculating relative frequencies at each step, and it is not even clear that this is in principle possible (a real coin would not survive). Furthermore, a series of events might very well have well-defined limiting relative frequencies but be such that one would not find them in the short to medium run.

This last caveat helps to illustrate an important general principle: The kinds of tools that allow one to make predictions based on past evidence require the use of background assumptions about how a given segment of the time series of events one observes relates to the time series more generally. In this case, if the local relative frequencies are approximately equal to the relative frequencies in future segments of the time series, then keeping track of local relative frequencies clearly provides a good way of making future predictions. Of course, as a solution to the general problem of induction, such an assumption simply begs the question. But it would perhaps be surprising to find that accurate predictions could be made without any background assumptions whatsoever. Thus, one might usefully classify methods of prediction by the type of background assumptions that must be satisfied for the method's predictions to be reliable.

One might, for instance, roughly distinguish *basic tools*, which make predictions on the basis of empirical data with only the most basic statistical assumptions, from *model-based tools*, which make predictions by estimating unknown parameters in more complex or intricately structured predictive models (see Hamilton 1994 for a generous sample of both sorts of predictive tools). Reichenbach's suggestion that one take local relative frequencies as probability estimates is an example of a basic predictive tool. Here a prediction is a bet that one's short-to-medium-run evidence faithfully reflects longer-run relative frequencies. While one may never be certain that this is the case, it is easy to imagine evidence for or against the claim that the assumption is reasonable in a given context. If it is, one can confidently use this basic tool for empirical prediction.

Another relatively basic predictive tool is Bayesian updating. This tool requires that one have prior probabilities to be updated on the basis of one's new evidence. This is an advantage in that more than one's relevant prior beliefs can be used in making predictions, and a disadvantage in that one must have an appropriate set of prior probabilities in order to use the predictive tool at all.

There are two steps to Bayesian updating. One first calculates the probability of the hypothesis under consideration H being true given evidence E using Bayes's theorem:

$$P(H \mid E) = \frac{P(H)P(E \mid H)}{P(E)}.$$

For a Bayesian subjectivist, $P(H)$ is one's prior degree of belief in H, $P(E|H)$ is the degree to which H being true would explain evidence E, and $P(E)$ is one's prior degree of belief that E would occur. Bayes's theorem follows from the axioms of probability theory and the definition of conditional probability.

The total probability theorem can be used to expand $P(E)$, yielding Bayes's theorem in a form that is often more useful:

$$P(H \mid E) = \frac{P(H)P(E \mid H)}{\sum_i P(H_i)P(E \mid H_i)},$$

where H_i values form a mutually exclusive and exhaustive set of possible hypotheses (which typically includes H), and $P(E|H_i)$ is a measure of how well each rival hypothesis would explain the occurrence of E.

After calculating the old probability of hypothesis H given evidence E, one updates the probability of H given that E has in fact occurred. In the simplest case, where one's evidence is itself certain, one might use strict conditionalization:

$$P_{new}(H) = P_{old}(H \mid E).$$

For a subjective Bayesian, P_{new} (H) represents the degree of belief one ought to have in H after evaluating evidence E. The justification here is given in a series of Dutch Book arguments, where one shows why an agent would accept irrational wagers guaranteed to lose money if the agent adopts an incompatible strategy for revising degrees of belief (Howson and Urbach 1989) (see Dutch Book Argument).

The use of prior probabilities in Bayesian updating requires stronger initial assumptions than

Reichenbach's method. But in return it allows inferential use of more of one's relevant background beliefs about the nature of the world and the context in which a prediction is made.

Because predictive tools are characterized by the background assumptions needed to apply them, basic prediction and model-based prediction differ only in the number, detail, and complexity of the background assumptions they require. Model-based prediction begins with a model for a selected set of empirical data on which reliable future predictions can be made if its parameters are correctly set. As one might expect, model-based predictive tools are both numerous and diverse. Adopting a predictive model might involve anything from assuming that the data should fit to a straight line (the parameters to be estimated here might be the line's slope and y-intercept) to assuming that the data should fit a broad range of specific parameters in a detailed causal description of the system being observed (the parameters to be estimated here might be the state of a physical system at a time and its Hamiltonian). Stronger background assumptions may permit the use of predictive tools that yield more detailed and/or accurate predictions, but the stronger assumptions are also more likely to be mistaken.

The methods one might use to set the parameters of a predictive model are similarly diverse. Basic predictive tools may be used to estimate the values of the unknown parameters in a more complex predictive model, or another model might very well be used to estimate these parameters. And the accuracy of one's subsequent predictions will depend upon whether the background assumptions required by the models are satisfied, the number of parameters estimated, the accuracy of the estimations, and the sensitivity of the models to specific failures in accuracy. In short, then, the reliability of the predictions of the model will typically depend upon a host of nontrivial background assumptions.

One might take from Hume's problem of induction the general lesson that information about the past cannot guide rational expectations about the future without some additional background assumptions about the system under consideration. It is perhaps not surprising that there are genuine choices to be made concerning what background assumptions and associated predictive tools are applicable in a given context. In real cases of justification of a particular set of background assumptions, the issue is not whether the future will resemble the past in some vague general sense. Rather, one typically finds a variety of concrete argumentative and evidential considerations weighing in favor of competing ways in which it might be expected to do so. And theories in the particular sciences discussed below are typically associated with one or more predictive models requiring various sets of both substantive and controversial background assumptions.

Prediction in the Empirical Sciences

Classical Mechanics and Chaotic Systems

Newton's classical mechanics is perhaps the most influential example of a predictive theory. It is deterministic in that the physical state (i.e., the position and momentum of each particle) of a closed and finite physical system at a time, $S(t_0)$, together with the energy properties of the system uniquely determine the physical state at all other times (see Determinism). Since the position of each particle can be given by 3 coordinates and the momentum can be given by 3 coordinates, the complete state of an N-particle system can be given by $6N$ coordinates or a single point in a $6N$ dimensional phase space. As the system evolves, the point representing its state in phase space moves in a continuous way. The past, present, and future history of a particular closed system is represented by the curve in phase space that represents the state of the system at each time. The dynamics can be represented as a set of differential equations that have as solutions the possible phase-space trajectories of the system. Since the history of the system is fully determined by the initial state $S(t_0)$ and the system's dynamical properties, this information and sufficiently precise calculations would, at least in principle, enable one to predict with perfect accuracy the state of the system at any time (see Classical Mechanics).

This ideal is compromised in application by the fact that observational error is always introduced in measuring continuous quantities like position, momentum, and energy with limited precision. Moreover, computational errors are nearly always introduced by rounding, since analytic solutions to the dynamical laws are rare in general and almost never perfectly applicable. Nonetheless, classical mechanics allows one to make very accurate empirical predictions in a wide variety of contexts, and Newton's *Principia* famously employs his theory of mechanics to explain and predict the future motions of the five primary planets, the moon, the satellites of Jupiter and Saturn, the precession of the equinoxes, tidal phenomena in the Earth's seas, and the motions of comets. And it does all this so successfully that many thought he had determined, as Edmund Halley wrote in his ode honoring Newton's accomplishment, "Jove's calculation and the laws / That the creator of all things, while he was setting the beginnings of the world, would not violate" (Newton [1713] 1999, 379).

Notwithstanding these and other remarkable successes, there are severe limits to prediction in classical mechanics that are consequences of its theories' nonlinear dynamics. The problem is that phase-space trajectories that are initially close may diverge exponentially with respect to time. Thus, the inevitable small errors in determining the initial state of a system or introduced into computation may generate large predictive errors. And it can happen that the expected error becomes so large over even relatively short times that one can predict almost nothing concerning the future state of the system from its current state. A chaotic system is one that exhibits such exponential sensitivity to initial conditions. More precisely, the chaotic domain of a system is that region of its phase space where the trajectories associated with infinitesimally displaced initial conditions separate from each other exponentially in time.

Chaotic behavior is exhibited by many familiar physical systems (Ott 1993). A dripping water faucet, for example, will often exhibit nearly equal times between drops for low, even inflows, then shift to an unpredictable sequence of times between drops when the inflow is increased. Chaotic behavior is also exhibited by some chemical and biological systems. And curiously, there is reason to expect the motions of the planets in the solar system, the paradigmatic example of clockwork regularity, to exhibit chaotic behavior. Given that the best estimates of the relevant continuous physical parameters are approximate and that numerical methods for performing computations invariably introduce error, there are strict limits on the reliability of the predictions concerning the motions of the planets obtainable from classical mechanics.

The behavior of chaotic systems is not, however, entirely unpredictable. In many cases, a chaotic system can be characterized for the purposes of prediction by its attractors, those sets of points in phase space (whether singular points, limit cycles, or more complex regions) that are attractive to all neighboring trajectories. Knowing the type and location of the attractors can allow one to predict the long-term behavior of even a chaotic system, although the nature and precision of such predictions will depend on the existence and type of attractors exhibited by the system. Some dynamical systems are associated with limit sets that are asymptotically attractive to neighboring trajectories but contain trajectories that are locally divergent within the attractive set. Such a limit set is called a *strange attractor*. If the system begins within the region attracted by the strange attractor, one would be able to predict convergence to the attractor but virtually nothing concerning the behavior within the attractor. Such attractors may even be typical in nonlinear systems of order higher than 2 (Cook 1994).

A rather different predictive problem in classical mechanics concerns the fact that it is deterministic for only finite, closed physical systems. Consider a particle that is moving at 1 m/s at $t = 0$, then is accelerated to 2 m/s at $t = \frac{1}{2}$ s, to 4 m/s at $t = \frac{3}{4}$ s, to 8 m/s at $t = \frac{7}{8}$ s, etc. After $t = 1$ s, the particle will be farther than any finite distance from where it started. Since any possible physical history can in classical mechanics be run backward as well as forward, consider the time-reversed version of this history. Here the particle starts farther than any finite distance from a system—for instance, one whose behavior one wishes to predict—and ends up crashing into it and ruining the prediction. In this case, one would not be able to make any predictions concerning the behavior of the system whatsoever, even after taking into account every particle that has a well-defined position at the beginning of the time-reversed story (Earman 1986).

Quantum Mechanics

Quantum mechanics is the most successful empirical theory ever, but unlike classical mechanics, it typically allows predictions that are only probabilistic (see Quantum Mechanics). In quantum mechanics the state of a physical system, S, is given by a vector, ψ_S, in an appropriate vector space. ψ_S is sometimes called the wave function of S. The state of the system almost always evolves in a linear, deterministic way that depends only on the energy properties of the system. In nonrelativistic quantum mechanics, this deterministic evolution is described by the time-dependent Schrödinger wave equation. Given the standard way of interpreting the quantum mechanical state, a physical system typically fails determinately to have or determinately not to have a given classical physical property at a time. But systems are found to have the determinate property being measured when a measurement is made. In the von Neumann–Dirac *collapse formulation* of quantum mechanics, this is explained by the collapse of the quantum mechanical state on measurement: When a system S, initially in state ψ_S, is measured, S instantaneously and randomly jumps to a state where the property being measured is determinate (see Quantum Measurement Problem). Which state S jumps to is taken to be an irreducible matter of chance. The probability of ending up in the determinate-property state χ_S is determined by the geometric relationship between the vectors ψ_S and χ_S (the probability is equal to $|\langle\psi\,|\,\chi\rangle|^2$). It is because the collapse dynamics is random that one is typically limited to making only probabilistic predictions concerning the results of future observations.

While there is disagreement concerning how one ought to understand quantum mechanics generally, and the collapse dynamics in particular, there is nearly universal agreement that one will never be able to make empirical predictions that do better than the standard quantum probabilities (Albert 1992). In this sense, the quantum probabilities are taken to represent a fundamental limitation to empirical prediction in physics.

Biological and Social Sciences

It is sometimes claimed that biological sciences in general, and evolutionary biology in particular, are not predictive, but this is at worst simply false and at best a simplistic description of a complex situation. There is no question, however, that there are systematic differences between the predictive capabilities characteristic of biological sciences and those familiar from the physical sciences. Some of these are illustrated by an example Mary Williams (1982) borrows of characteristic prediction in evolutionary theory: "Sexual dimorphism in the length and color of the furry body covering of bumblebees should show a latitudinal and altitudinal gradient among species of bumblebees, with tropical and low altitude species having more dimorphism" (293). As the example suggests, biological predictions more typically concern groups, species, populations, or ensembles than individuals, and they more often describe unknown past or present states of affairs than future ones. Moreover, the intended scope of predictive generalizations in biology is typically not spatiotemporally unrestricted: They are at a minimum restricted to circumstances in which particular (often unique and evolutionarily contingent) causal mechanisms operate, and they are typically exception-ridden or asserted *ceteris paribus*, even within the scope of such intended domains. Mendel's law of segregation, for example, can be used to make reliable predictions (of ratios for populations or probabilities for individuals), but even these predictions are both restricted to contexts in which a particular, evolutionarily contingent causal mechanism is operating (i.e., sexual reproduction) and are subject to exceptions (e.g., meiotic drive) even in that domain. Perhaps the most interesting question, then, concerns the source of these characteristic differences in predictive capabilities (see Population Genetics).

The first and most obvious is the relatively greater causal complexity of the natural systems studied by biological sciences as compared with those in the restricted domains in which physical science is able to deliver precise and accurate predictions. As Mitchell (2003, chap. 5) effectively documents, the relevant complexity here is of several kinds, including the compositional complexity of actual biological systems, the number and variety of causal processes operating in them, the sometimes dramatic and sudden shifts in the relationships between key variables across different ranges of their variation, and the characteristic embeddedness of biological entities in levels of organization with multiple weak, nonadditive forces and redundant mechanisms operating both within and between these various levels. The impact of such causal complexity is amplified by the long timescale of evolutionarily significant effects, during which such complexity must be modeled, controlled, or eliminated to allow effective prediction. Also significant is the characteristic contingency of the states of affairs studied by evolutionary biology and other historical sciences on the occurrence of particular (often rare) events, which themselves constrain and shape the course of future evolutionary change. Perhaps the simplest example here is the effect of mutation: At a given time, different mutational substitutions of just a single amino acid can easily lead to extremely different evolutionary outcomes over relatively short timescales, but the process of mutation is treated as a random variable by evolutionary theory, either because it is genuinely indeterministic or because one does not yet know enough about the process or relevant conditions in particular organisms to predict with any precision what, when, and how particular mutations will occur. It is for all these sorts of reasons (although not always in these terms) that philosophers of science have enthusiastically debated whether there are laws of any traditional variety in biological science and the ultimate source of the indeterministic character of evolutionary theory.

Taking these sorts of complexities and contingencies into account suggests that the physical analogue for biological sciences is not predicting the speed and position of a ball rolling down an inclined plane, but something more like predicting the path of a bag of feathers dumped out of an airplane. Physical sciences are able to make some predictions about what will happen in these circumstances, but these predictions will be relatively weak, more likely to concern the feathers as a group, and will be reliable only *ceteris paribus* or subject to the operation and/or interference of particular causal mechanisms that affect their trajectory (e.g., prevailing weather conditions). Of course, the extent to which these features are characteristic of the predictive application of the physical sciences generally is also controversial (see Cartwright 1983), but biological sciences appear to fare even worse in natural settings and to be even less amenable to the construction of specialized contexts that make precise and powerful prediction possible.

Rosenberg (1983 and 1994) and others have argued for a related but perhaps more fundamental kind of limitation on the predictive capabilities of biological sciences. Rosenberg suggests that such predictive limitations arise because in typical cases (arguably excluding some parts of molecular genetics) biologically significant categories must be characterized functionally or teleologically, and there is a wide diversity (which seems unlikely to admit of finite specification even in a disjunctive list) of possible physical realizations of such functionally characterized entities (see Function).

Put another way, the claims here are that

- there are no type identities between important explanatory biological categories (like fitness, mimicry, temperature-regulating mechanism, balanced polymorphism, regulatory gene) and the mechanistic physical descriptions of the tokens instantiating these functional kinds in particular cases; and
- it is these mechanistic descriptions that offer power and precision in predicting the range of conditions under which a mechanism will operate, what causal factors might interfere with it, its probable evolutionary trajectory, and the like.

Rosenberg thus argues that the goals of prediction and explanation pull in different directions here and that the diversity and potential infinity of the possible realizations of functionally characterized biological kinds conspire to ensure that biological sciences retain their weakly predictive character if they are to remain useful to us.

Perhaps a useful example of how this can be so is provided by Fisher's (1930) famous explanation of why the sex ratio in sexually reproducing diploid species at reproductive age is typically 1:1. Briefly, the explanation is that (assuming equal parental cost to produce offspring of either sex and ignorance of offspring quality, and setting aside complications) no matter what the mating system, a parent will spread more copies of its genes by producing offspring of the less numerous sex: Since every successful mating requires the genetic contribution of exactly one member of each sex, members of the less numerous sex are more easily able to obtain multiple successful matings, be more choosy about mates, or enjoy whatever reproductive advantages members of that sex enjoy in the mating system. Even if a species' mating is structured in such a way that a few successful males do all the mating and most males do not mate at all, when males are less numerous than females, a parent will do better on average by producing sons with a proportional *chance* of being one of the lucky few than daughters who are guaranteed to mate. Therefore, it pays to produce members of whichever sex is more rare, ensuring strong selective pressure against any mechanism that favors producing male offspring over female or vice versa.

Contrast this explanation of the sex ratio with that provided by conjoining the physical histories of each organism in each species. There is certainly a sense in which the sex ratio is thereby explained, for this complex history (including the random segregation of sex-determining chromosomes, the details of development and survivorship for each organism, and even the survival and propagation of the specific genetic, physiological, and ontogenetic mechanisms responsible for sex determination in each species in the first place) deductively imply that sex ratios are what they are (near 1:1 in each case). But this mechanistic explanation, provided in terms that would increase predictive power and precision regarding individual cases, provides nothing like Fisher's insight into the evolutionary reasons for the emergence and persistence of the 1:1 sex ratio. Similarly, a detailed molecular description of the operation of a particular DNA-repair mechanism may permit effective prediction of the conditions under which the mechanism will operate, but it will be unable to explain *why* such a mechanism exists and persists and perhaps even obscure the fact that it is a mechanism for repairing DNA (and thereby minimizing mutational changes). Of course, without this functional characterization, there would seem little sense left to be made of the question of under what conditions such a mechanism will operate (or do so effectively). Such examples illustrate why the relative predictive weakness of biological sciences might not simply be an unfortunate consequence of increased complexity and contingency in their domains of application, but also an aspect of those sciences intimately bound up with what renders them useful.

An interesting further question concerns to what extent any of these predictively limiting aspects of biological science also underlie the relative predictive weakness of the social sciences. Causal complexity and contingency are often invoked in this connection, perhaps most famously in an influential argument due to Karl Popper (1957) (helpfully discussed by Rosenberg 1993) to the effect that because the growth of scientific knowledge has persistently exerted dramatic effects on the course of history and human affairs, the unpredictable trajectory and directions of such growth preclude even the possibility of a predictively robust social science. However, the failure of type identities between important explanatory categories has also been invoked to explain the relative predictive weakness of the social sciences (see Rosenberg 1994). Here it is the *intentional* explanatory categories of the social sciences that are supposedly both ineliminable and multiply realized by tokens that are heterogeneous in the terms of more predictively precise and powerful sciences, including biology itself (but cf. Nelson 1990) (see Intentionality). These and closely related considerations are invoked to support a variety of predictively relevant conclusions concerning the social sciences, including the claims that the kinds of predictive limitations discussed above will prove to be ineliminable from them; social scientific predictions will remain merely "generic" or qualitative; genuinely scientific and predictive social science will have to eliminate any appeals to intentional notions; the study of social phenomena is autonomous and cannot be understood in terms of aggregate actions and dispositions of individuals; and social inquiry must or should be restricted to the interpretive study of others (see Social Sciences, Philosophy of).

JEFFREY A. BARRETT
P. KYLE STANFORD

References

Albert, D. (1992), *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press.

Brush, Stephen G. (1995), "*Dynamics of Theory Change: The Role of Predictions*," *Proceedings of the Philosophy of Science Association* 2: 133–145.

Campbell, Richmond, and Thomas Vinci (1983), "Novel Confirmation," *British Journal for the Philosophy of Science* 34: 315–341.

Carnap, Rudolf (1967), *The Logical Structure of the World: Pseudoproblems in Philosophy*. Translated by Rolf A. George. Berkeley and Los Angeles: University of California Press.

Cartwright, Nancy (1983), *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Cook, Peter A. (1994), *Nonlinear Dynamical Systems*, 2nd ed. Hemel Hempstead, UK: Prentice-Hall International.

Earman, John (1986), *A Primer on Determinism*. Dordrecht, Netherlands: D. Reidel.

Fisher, R. A. (1930), *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press.

Gardner, Michael R. (1982), "Predicting Novel Facts," *British Journal for the Philosophy of Science* 33: 1–15.

Giere, Ronald N. (1983), "Testing Theoretical Hypotheses," in John Earman (ed.), *Minnesota Studies in the Philosophy of Science*, vol. 10: *Testing Scientific Theories*. Minneapolis: University of Minnesota Press, 269–298.

——— (1984), *Understanding Scientific Reasoning*, 2nd ed. New York: Holt, Rinehart and Winston.

Glymour, Clark (1980), *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Goodman, Nelson (1954), *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Hamilton, James D. (1994), *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hempel, Carl G. (1965), *Aspects of Scientific Explanation*. New York: Free Press.

Horwich, Paul (1982), *Probability and Evidence*. Cambridge: Cambridge University Press.

Howson, Colin, and Peter Urbach (1989), *Scientific Reasoning: The Bayesian Approach*. LaSalle, IL: Open Court.

Hume, David ([1748] 1977), *An Enquiry Concerning Human Understanding*. Edited by Eric Steinberg. Indianapolis, IN: Hackett.

Lakatos, Imre (1970), "Falsification and the Methodology of Scientific Research Programmes," in Imre Lakatos and Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 91–195.

Lange, Marc (2001), "The Apparent Superiority of Prediction to Accommodation as a Side Effect: A Reply to Maher," *British Journal for the Philosophy of Science* 52: 575–588.

Leplin, Jarrett (1997), *A Novel Defense of Scientific Realism*. New York: Oxford University Press.

Lipton, Peter (1991), *Inference to the Best Explanation*. London: Routledge.

Mayo, Deborah G. (1991), "*Novel Evidence and Severe Tests*," *Philosophy of Science* 58: 523–552.

Mitchell, Sandra D. (2003), *Biological Complexity and Integrative Pluralism*. Cambridge: Cambridge University Press.

Musgrave, Alan (1974), "Logical versus Historical Theories of Confirmation," *British Journal for the Philosophy of Science* 25: 1–23.

Nelson, Alan J. (1990), "Social Science and the Mental," in Peter A. French, Theodore E. Uehling Jr., and Howard K. Wettstein (eds.), *Midwest Studies in Philosophy*, vol. 15: *The Philosophy of the Human Sciences*. Notre Dame, IN: University of Notre Dame Press, 194–209.

Newton, Isaac ([1713] 1999), *The Principia: Mathematical Principles of Natural Philosophy*, 2nd ed. Translated by I. Bernard Cohen and Anne Whitmen. Berkeley and Los Angeles: University of California Press.

Ott, Edward (1993), *Chaos in Dynamical Systems*. Cambridge: Cambridge University Press.

Popper, Karl R. (1957), *The Poverty of Historicism*. London: Routledge.

——— (1963), *Conjectures and Refutations*. New York: Harper.

Quine, Willard Van (1951), "Two Dogmas of Empiricism," in *From a Logical Point of View*. New York: Harper, 20–46.

Reichenbach, Hans (1938), *Experience and Prediction: An Analysis of the Foundations and Structure of Knowledge*. Chicago: University of Chicago Press.

Rosenberg, Alexander (1983), *The Structure of Biological Science*. Cambridge: Cambridge University Press.
——— (1993), "*Scientific Innovation and the Limits of Social Scientific Prediction*," *Synthese* 97: 161–182.
——— (1994), *Instrumental Biology or the Disunity of Science*. Chicago: University of Chicago Press.
Scerri, Eric R., and John Worrall (2001), "*Prediction and the Periodic Table*," *Studies in History and Philosophy of Science* 32A: 407–452.
Schlesinger, George N. (1987), "Accommodation and Prediction," *Australasian Journal of Philosophy* 65: 33–42.
van Fraassen, Bas (1980), *The Scientific Image*. Oxford: Clarendon Press.
Williams, Mary B. (1982), "The Importance of Prediction Testing in Evolutionary Biology," *Erkenntnis* 17: 291–306.
Worrall, John (1978), "*The Ways in Which the Methodology of Scientific Research Programmes Improve on Popper's Methodology*," in G. Radnitzky and G. Andersson (eds.), *Boston Studies in the Philosophy of Science*, vol. 58: *Progress and Rationality in Science*. Dordrecht, Netherlands: Reidel, 321–338.
——— (1989), "*Fresnel, Poisson and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories*," in D. Gooding, T. Pinch and S. Shaffer (eds.), *The Uses of Experiment: Studies in the Natural Sciences*. Cambridge: Cambridge University Press, 135–157.
Zahar, Elie (1973), "Why Did Einstein's Programme Supersede Lorentz's?" parts I and II, *British Journal for the Philosophy of Science* 24: 95–125, 223–262.

See also **Bayesianism; Causality; Confirmation Theory; Determinism; Empiricism; Explanation; Induction, Problem of; Laws of Nature; Logical Empiricism; Phenomenalism; Popper, Karl Raimund; Probability; Reichenbach, Hans; Theories; Verificationism**

PROBABILITY

There are two central questions concerning probability. First, what are its formal features? That is a *mathematical* question, to which there is a standard, widely, though not universally, agreed upon answer (reviewed in the next section). Second, what sorts of things *are* probabilities—what, that is, is the *subject matter* of probability theory? This is a *philosophical* question, and while the mathematical theory of probability certainly bears on it, the answer must come from elsewhere. To see why, observe that there are many things in the world that have the *mathematical structure* of probabilities (e.g., the set of measurable regions on the surface of a table) but that would never be mistaken for *being* probabilities. So probability is distinguished by more than just its formal characteristics. The bulk of this essay will be taken up with the central question of what this "more" might be.

Kolmogorov's Axiomatization

Probability theory was inspired by games of chance in seventeenth-century France and inaugurated by the Fermat–Pascal correspondence, which culminated in the *Port Royal Logic* (Arnauld [1662] 1964). Its axiomatization had to wait nearly another three centuries. The locus classicus of the mathematical theory of probability is Kolmogorov's ([1933] 1950) *Foundations of Probability.* Inspired by measure theory, Kolmogorov's axiomatization has become orthodoxy. Let Ω be a nonempty set. A *field* (*algebra*) on Ω is a set F of subsets of Ω that has Ω as a member and that is closed under complementation (with respect to Ω) and union. Assume for now that F is finite. Let P be a function from F to the real numbers, obeying the following axioms:

$$P(a) \geq 0 \text{ for all } a \in F, \qquad \text{A1}$$

$$P(\Omega) = 1, \text{and} \qquad \text{A2}$$

$$P(a\,\Omega\,b) = P(a) + P(b)$$
$$\text{for all } a, b \in F \text{ such that } a \cap b = \phi. \qquad \text{A3}$$

Call P a probability function, and (Ω, F, P) a probability space.

One could instead attach probabilities to members of a collection of *sentences* of a formal language, closed under truth-functional combinations. Either way, a kind of reflective equilibrium is achieved between these axioms, which are thought to be intuitively plausible, and various important interpretations of probability (to be discussed in the subsequent sections), which obey them and bring them to life in applications.

It is often thought that the only nonconventional part of the axiomatization is A3. That is too quick, for it is substantive that probabilities are

1. Defined by *functions* (rather than by one-many or many-many mappings);
2. Functions of *one variable* (unlike primitive conditional probability functions, which are functions of two variables);
3. Defined on a *field* (rather than a set with weaker closure conditions);
4. Represented *numerically* (rather than qualitatively, as is "possibility"; or comparatively, as is "similarity to a given world" in the Stalnaker/Lewis style of semantics for counterfactuals [Lewis 1986a]);
5. *Real* numbers (rather than those of some other number system);
6. *Bounded* (unlike other quantities that are treated measure-theoretically, such as lengths);
7. Bounded by *Maximal* and *minimal* values (thus prohibiting open or half-open ranges).

For a discussion of rival theories that relax or replace points 2, 3, 4, and 6 above, see Fine 1973. Complex-valued probabilities are proposed by Feynman and Cox (Mückenheim et al. 1986); infinitesimal probabilities (of nonstandard analysis) by Skyrms (1980) and Lewis (1986b), among others; and unbounded probabilities by Renyi (1970). Primitive conditional probability functions will be briefly discussed at the end of this section.

Kolmogorov extends his axiomatization to cover infinite probability spaces. Probabilities are now defined on a *σ-field* (*σ-algebra*)—a field that is further closed under *countable* unions—and A3 is correspondingly strengthened:

A3′ (Countable additivity) If $a_1, a_2, a_3, \ldots$ is a countable sequence of (pairwise) disjoint sets, each belonging to F, then $P(\bigcup_{n=1}^{\infty} a_n) = \sum_{n=1}^{\infty} P(a_n)$.

De Finetti (1990) is a notable opponent of countable additivity.

Kolmogorov then defines the conditional probability of a given b by the ratio of unconditional probabilities:

$$P(a|b) = \frac{P(a \cap b)}{P(b)}, \text{provided } P(b) > 0.$$

Note that this ratio is undefined if either or both of the unconditional probabilities are undefined, or if $P(b) = 0$. Yet in uncountable spaces there can be genuine, nontrivial events whose probabilities are undefined (so-called "nonmeasurable" sets), and others whose probabilities are 0 ("probability 0 does not imply impossible," as textbooks and Kolmogorov himself caution us). So Kolmogorov's definition does not guarantee that certain intuitive constraints on conditional probability are met—for example, that the probability of an event, *given itself*, is 1.

Kolmogorov addresses the probability-0 problem with a more sophisticated account of conditional probability as a random variable conditional on a sigma algebra, appealing to the Radon-Nikodym theorem to guarantee the existence of such a random variable (see, e.g., Billingsley 1995). A rival approach takes conditional probability $P(\_,\_)$ as primitive and defines the unconditional probability of a as $P(a, \mathbf{T})$, where $\mathbf{T}$ is a necessary (e.g., tautological) proposition. Various axiomatizations of primitive conditional probability have been defended in the literature, typically differing only in the handling of conditional probabilities with zero unconditional probability antecedents. In many ways, the most general and elegant of the proposed axiomatizations is Popper's (1959). (See Roeper and Leblanc 1999 for an encyclopedic discussion of competing theories of conditional probability, and Keynes 1921, Carnap 1950, Popper 1959, and Hájek 2003b for arguments that probability is inherently a two-place function.)

Versions of Bayes's theorem can now be proven (see Bayesianism):

$$P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)} = \frac{P(b \mid a)P(a)}{P(b \mid a)P(a) + P(b \mid \neg a)P(\neg a)}.$$

More generally, suppose there is a partition of hypotheses $\{h_1, h_2, \ldots, h_n\}$ and evidence e. Then for each i,

$$P(h_i \mid e) = \frac{P(e \mid h_i)P(h_i)}{\sum_{j=1}^{n} P(e \mid h_j)P(h_j)}.$$

The $P(e|h_i)$ terms are called likelihoods, and the $P(h_i)$ terms are called priors. Finally, Kolmogorov defines a and b to be *independent* iff (if and only if) $P(a|b) = P(a)$; equivalently, iff $P(b|a) = P(b)$; equivalently, iff $P(a \cap b) = P(a)P(b)$ (for $P(a) \neq 0 \neq P(b)$). The terminology suppresses the fact that such independence is really a *three*-place relation between an event, another event, and a *probability function.* This distinguishes probabilistic independence from such two-place relations as logical, causal, and counterfactual independence.

The next section turns to the so-called *interpretations* of probability—attempts to answer the central philosophical question: What is probability?

Frequentism

Ask a scientist what probability is, and one will typically get a *frequentist* answer: The probability of an event is the relative frequency of trials of a repeatable experiment on which that event occurs; sometimes the words "in the long run" are added. This leaves open important questions: Which are the trials to be counted? How long does the run have to be? One may confine one's attention to *actual* trials, realized in this world, or countenance *hypothetical* trials. And one may have merely finitely many trials to contend with or infinitely many, in which case probability will be identified with the *limit* of the relative frequency in a *sequence* of trials. One may thus immediately distinguish $2 \times 2 = 4$ variants of frequentism. However, the actual world typically delivers only finitely many trials of any given experiment. And it is often thought that if one is going to allow the trials to be hypothetical anyway, there is no obstacle to letting the sequence of trials be infinite, thus *guaranteeing* a "long run." So one may confine one's attention, as frequentists typically do, to just two of the possible positions: finite actual frequentism and infinite hypothetical frequentism.

In his discussion of the proportion of births of males and females, Venn (1866) contends that "probability is nothing but that proportion" (84)—a version of finite actual frequentism. Von Mises (1957), by contrast, insists that probabilities exist only relative to virtual infinite sequences of "attributes" called *collectives*. In a collective, the limiting relative frequency of any attribute exists and is the same on any recursively specified subsequence. (Von Mises' original definition, in terms of "place selections," is finessed by Church.) The probability of a given attribute, relative to a collective, is then identified with its limiting relative frequency in that collective. Von Mises' position is thus a version of infinite hypothetical frequentism, as are those of Reichenbach and van Fraassen.

Any version of frequentism faces the notorious *reference class problem.* Any event, in all its detail, occurs exactly once, so if nontrivial frequencies are to be associated with it, it must be regarded as a token of a more general event type, whose instances constitute its reference class. However, there are indefinitely many ways of typing a given event. This would not be a problem if its relative frequency were the same in each reference class, or if one such class stood out as natural or privileged. The problem gains teeth to the extent that various competing reference classes have equal claim to determining the probability and that they yield different relative frequencies for the event.

In some cases, the reference class problem may be solved for the actual, finite frequentist, but at the price of creating the equally notorious *problem of the single case:* Intuitively, the objective probability of a one-off event may be less than 1, but finite frequentism cannot respect this intuition. Many events occur only once by any reasonable standard of typing: the 2000 presidential election, the invasion of Iraq, the last Lakers–Bulls game, and so on. The only natural reference class for such an event is the singleton set consisting of itself, and thus it has relative frequency 1 (and its nonoccurrence has relative frequency 0). Nonetheless, it seems natural to think of nonextreme probabilities attaching to at least some of these "single-case" events.

The problem of the single case is particularly striking, but there is really a sequence of related "granularity" problems: the problem of the double case, the problem of the triple case, and so on. A finite reference class of size n can produce relative frequencies at only a certain level of "grain," namely $\frac{1}{n}$. Among other things, this rules out irrational probabilities; yet, the best physical theories say otherwise (for example, various decay probabilities delivered by quantum mechanics are irrational). Furthermore, there is a sense in which any of these problems can be transformed into the problem of the single case. Suppose that a coin is tossed a thousand times. This can be regarded as a *single* trial of a thousand-tosses-of-the-coin experiment. Yet one does not want to be committed to saying that *that* experiment yields its actual result with probability 1.

The move to infinite hypothetical frequentism makes the reference class problem only worse, for not only must a set of events be chosen in which to place a given event, but since the set is now infinite, an *ordering* among the events must be chosen. After all, in nontrivial cases a limiting relative frequency can be made whatever value one likes simply by reordering the results of a given sequence. Consider the limiting relative frequency of even numbers among positive integers. In the "natural" ordering $<1, 2, 3, \ldots>$ it is $\frac{1}{2}$; however, one can make it $\frac{1}{4}$ by reordering the integers so that the even numbers occur at every fourth place in the sequence: $<1, 3, 5, 2, 7, 9, 11, 4, 13, \ldots>$; and so on. Thus, limiting relative frequencies are sensitive to apparently arbitrary choices of ordering, while it appears that probabilities need not be. One might call this the *reference sequence problem.*

A sequence of events is said to be *exchangeable* with respect to a given probability function if all the joint probabilities of the events are invariant under finitely many permutations of the sequence: Every event has the same probability, every conjunction of two events has the same probability, every conjunction of three events has the same probability, and so on. A sequence of events is automatically exchangeable with respect to the relative frequency function: The frequency of an event is insensitive to *which* trials the event occurs at. Yet various events intuitively are not exchangeable with respect to the relevant probability function. Consider someone learning to throw a dart at a bull's-eye: The sequence <MISS, MISS, HIT> is presumably more probable than <HIT, MISS, MISS>, because the dart thrower's accuracy improves with practice. Yet the (finite) relative frequency of HIT is $\frac{1}{3}$ either way. Since relative frequencies *force* a kind of symmetry that probabilities need not obey, they cannot be the same thing. (Ironically, it was the failure of a more thoroughgoing "infinite exchangeability" that proved to be the undoing of hypothetical infinite frequentism in the previous paragraph.)

The Classical Interpretation

The brainchild of such founding fathers of probability as Pascal, Fermat, Huygens, and Leibniz, and clearly articulated in Laplace ([1814] 1951), the classical interpretation is the oldest interpretation of probability—indeed, it dates back to a time when the axiomatization and interpretation of probability were not clearly distinguished. It seeks to characterize the probability assignment of a rational agent in a state of epistemic neutrality with respect to a finite set of "equipossibilities": The agent has either *no evidence* or *symmetrically balanced evidence* regarding the possibilities. It appeals to the so-called principle of indifference: Whenever there is no evidence favoring one possibility over another, each should be assigned the same probability as the others. So

$$P(e) = \frac{\text{number of equipossibilities in which } e \text{ occurs}}{\text{total number of equipossibilities.}}$$

But the notion of "equipossibilities" seems to presuppose some prior notion of probability. After all, the most obvious characterization of "symmetrically balanced evidence" is in terms of equality of conditional probabilities: Given evidence e and possible outcomes $o_1, o_2, \ldots, o_n$, the evidence is symmetrically balanced with respect to the outcomes iff $P(o_1|e) = P(o_2|e) = \ldots = P(o_n|e)$. Perhaps, then, one should regard the classical interpretation as an attempt to reduce *quantitative* probability to *comparative* probability: All *numerical* probabilities are ultimately based on facts about *equalities* among probabilities.

Note the structural resemblance of the classical theory to finite frequentism. Both theories see probability as a matter of evenhanded counting and ratio taking:

$$P(e) = \frac{\text{number of cases favorable to } e}{\text{total number of cases}}.$$

It is just that for frequentism, the cases are *actual* outcomes of a *repeated* experiment, whereas for the classical theory they are *possible* outcomes of a *single* experiment. And indeed the classical theory faces many of the same problems as frequentism. There is the granularity problem: Clearly, every classical probability is some fraction of the form $\frac{m}{n}$, where n is the number of possibilities. There is the exchangeability problem: Classical probabilities are invariant under permutation of the labeling of the possibilities (for example, relabeling the faces of a die makes no difference to their probabilities of coming up). Thus, the classical interpretation cannot readily provide *asymmetric* probability distributions (e.g., for biased dice or coins), and it cannot handle distributions that evolve over time (e.g., for the dart thrower's hitting the bull's-eye).

Moreover, the reference class problem reappears. If one is truly ignorant about the results of some experiment, then presumably there is nothing to favor various competing choices of sample space. One should then be indifferent between, for example, {heads, tails} and {heads, tails, edge}. And one should be indifferent between various refinements of the original space: for example, between spaces that refine in different ways the heads outcome according to its final orientation relative to due north. Thus, probabilities will be determined by an apparently arbitrary choice of sample space. To adapt an example from physics: Bose-Einstein statistics, Fermi-Dirac statistics, and Maxwell-Boltzmann statistics each arise by considering the ways in which particles can be assigned to states and then partitioning the set of alternatives in different ways (see, e.g., Fine 1973). Someone ignorant of which statistics apply to a given type of particle can make only an arbitrary choice and hope for the best.

In typical applications of the classical theory (gambling, for example), one is not wholly ignorant, but the evidence that one has is symmetrically balanced regarding the possibilities. There are two problems here: in the evidence and in the symmetry. Classical probabilities are acutely sensitive to the evidence. If the evidence becomes *unbalanced*, favoring some outcomes over others, then classical probabilities are not merely revised; they are *destroyed.* And there may be competing respects of symmetry, each equally compelling. This problem arises especially when there are infinitely many possible outcomes. Then, the equipossibilities must be a finite partition of the outcomes. But *which* partition?

A tempting answer may be: the most "natural" partition. However, "Bertrand's paradoxes" show that there need not be any. The trick is to give competing parameterizations of a given problem that are nonlinearly related to one another but equally natural. Suppose one is told only that a car traveled 100 miles at an average speed between 50 and 100 mph. What is the probability that its average speed was between 75 and 100 mph? Perhaps 0.5, since (50, 75) and (75, 100) are equipossible intervals for the average speed. But the question could be equivalently formulated: A car took between 1 and 2 hours to travel 100 miles. What is the probability that it took between 1 hour and $1\frac{1}{3}$ hours? Now it seems that there are three equipossible intervals for the time taken: $(1, 1\frac{1}{3})$, $(1\frac{1}{3}, 1\frac{2}{3})$, and $(1\frac{2}{3}, 2)$; whence the answer should be $\frac{1}{3}$.

Logical Probability

Many philosophers—Leibniz, von Kries, Keynes, Wittgenstein, Waismann, Carnap, and others—have tried to explicate the following "logical" concept of conditional probability:

$$P(p \mid q) = \frac{\text{the proportion of logically possible worlds in which both } p \text{ and } q \text{ are true}}{\text{the proportion of logically possible worlds in which } q \text{ is true}}.$$

An obvious problem has been to justify a *measure* of the "proportion of logically possible worlds in which a proposition is true." Early attempts (including those by Carnap that will be the focus here) tried to apply the controversial principle of indifference (see Carnap, Rudolf). Carnap's (1950) early constructions are very similar to systems developed earlier by W. E. Johnson (1921) (see Inductive Logic for further references on Carnapian inductive logic and logical probability).

Begin with a first-order language L containing a finite number of monadic predicates: $F, G, H, \ldots$, and a finite or denumerable number of individual constants $a, b, c, \ldots$. Then define an (a priori) unconditional probability function $P(\bullet)$ over the sentences of L, in a way that appeals to only their *syntactic structure* (whence the name "logical" probability). Finally, use the standard ratio definition to construct a conditional probability function $P(\bullet \mid \bullet)$ over pairs of sentences of L.

The results of this procedure will be *language relative:* If one describes the same phenomena by means of a different language L^*—equipped with a different stock of monadic predicates—one will typically not recover the same probabilities. Consider two languages used to represent the outcomes of random draws from an urn filled with colored balls. Let L contain the color predicates "blue" and "green" and let L^* contain the predicates "grue" and "bleen." The intended interpretation is that a draw is grue just in case it is one of the first million and green or a later one and blue; a draw is bleen just in case it is one of the first million and blue or a later one and green. Starting with L, use whatever is the appropriate procedure to calculate

> P(draw 1,000,001 is green|the first 1,000,000 draws are green).

Starting with L^*, use this procedure to calculate

> P(draw 1,000,001 is grue|the first 1,000,000 draws are grue).

If syntax is all that matters, then these conditional probability values will be identical—and surely greater than $\frac{1}{2}$, at least if logical probability is to have a hope of modeling actual inductive reasoning (see Inductive Logic). The trouble is that the second conditional probability, translated into L, is just

> P(draw 1,000,001 is blue|the first 1,000,000 draws are green).

One can avoid contradiction, but only by explicitly insisting that probability is language-relative. And that raises a serious problem—really, the reference class problem in a new guise: If one wishes to employ logical probability as a foundation for inductive inference, which is the "right" language to use? The remainder of this discussion will presuppose that an answer to this question has been found (for Carnap [1980], this question was "external" to inductive logic anyway, and his later systems did not have this blatant form of language relativity).

Returning now to Carnap's early systems, consider a simple language with only two monadic predicates F and G and only two individual constants a and b. This language yields exactly sixteen maximally specific descriptions of the world—the *state descriptions* of L: $(Fa \wedge Ga \wedge Fb \wedge Gb)$, $(Fa \wedge Ga \wedge Fb \wedge \neg Gb)$, etc. Two state descriptions S_1 and S_2 are *permutations* of each other if S_1 can be obtained from S_2 by some permutation of the individual constants. For example, $Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$ and $\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb$ are permutations of each other. A *structure description* in L is a disjunction of state descriptions, closed under permutation. The L language provides these ten structure descriptions:

$Fa \wedge Ga \wedge Fb \wedge Gb$

$(Fa \wedge Ga \wedge Fb \wedge \neg Gb) \vee$
$(Fa \wedge \neg Ga \wedge Fb \wedge Gb)$

$(Fa \wedge Ga \wedge \neg Fb \wedge Gb) \vee$
$(\neg Fa \wedge Ga \wedge Fb \wedge Gb)$

$Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb) \vee$
$(\neg Fa \wedge \neg Ga \wedge Fb \wedge Gb)$

$Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb$

$(Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb) \vee$
$(\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb)$

$(Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb) \vee$
$(\neg Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb)$

$\neg Fa \wedge Ga \wedge \neg Fb \wedge Gb$

$(\neg Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb) \vee$
$(\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb)$

$\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb$

Now, assign nonnegative real numbers to the state descriptions, so that these sixteen numbers sum to 1. Any such assignment will constitute an (a priori) unconditional probability function $P(\bullet)$ over the state descriptions of L. To extend $P(\bullet)$ to the entire language L, note that the probability of a disjunction of mutually exclusive sentences is the sum of the probabilities of its disjuncts. Since every sentence in L is equivalent to some disjunction of state descriptions, and all the state descriptions are mutually exclusive, this gives a complete unconditional probability function $P(\bullet)$ over L—typically called a *measure function.* The standard ratio definition then yields a conditional probability function $P(\bullet \mid \bullet)$ over pairs of sentences in L. Carnap (1950) discusses two natural measure functions. The first, $m^\dagger$, treats each state description as equiprobable a priori: If there are N state descriptions in L, then $m^\dagger$ assigns $\frac{1}{N}$ to each. However natural this measure function may seem, it has the consequence that the resulting probabilities cannot undergird *learning from experience.* To see why, observe that

$$P(Fb \mid Fa) = \frac{m^\dagger(Fb \wedge Fa)}{m^\dagger(Fa)} = \frac{1}{2} = m^\dagger(Fb) = P(Fb).$$

So "learning" that one object has property F cannot affect the probability that any other object will also have property F. Indeed, it can be shown that *no matter how many objects are assumed to be F*, this will (according to probability functions based on $m^\dagger$) *always be irrelevant* to the hypothesis that a distinct object will also be F—a feature widely viewed as a serious shortcoming of $m^\dagger$.

As a result, Carnap formulated an alternative measure function m^*: First, assign equal probabilities to each *structure* description. Then, each state description entailing a given structure description is assigned an equal portion of the probability assigned to that structure description. So, in the present toy language, the state description $Fa \wedge Ga \wedge \neg Fb \wedge Gb$ gets assigned an a priori probability of $\frac{1}{20}$ ($\frac{1}{2}$ of $\frac{1}{10}$), but the state description $Fa \wedge Ga \wedge Fb \wedge Gb$ receives an a priori probability of $\frac{1}{10}$ ($\frac{1}{1}$ of $\frac{1}{10}$). Unlike $m^\dagger$, m^* *does* allow for learning from experience; for example, $P(Fa \mid Fb) = \frac{3}{5} > \frac{1}{2} = P(Fa)$. Still, even m^* can give unintuitive results in more complex languages (see Carnap 1952 for discussion). Also, note that the state descriptions are exchangeable with respect to m^*, an omen that logical probabilities will face some of the problems that plagued the frequentist and the classical probabilist.

Carnap (1952) presents a more complicated "continuum" of conditional probability functions. This continuum depends on a parameter λ intended to reflect the "speed" with which learning from experience is possible. $\lambda = 0$ corresponds to the "straight rule," which says that the probability that the next object observed will be F, conditional upon a sequence of past observations, is simply the frequency of F objects in that sequence; $\lambda = +\infty$ yields a conditional probability function much like that derived from the measure function $m^\dagger$ (i.e., $\lambda = +\infty$ implies that there is no learning from experience); $\lambda = \kappa$ (which is the number of independent families of predicates in Carnap's more elaborate [1952] linguistic framework) yields a conditional probability function equivalent to that generated by the measure function m^*.

Problems remain. None of the Carnapian systems allow universal generalizations to have nonzero probability. Carnap's early systems also failed to allow for *analogical effects*, since in these systems the fact that two objects have several properties in common is (in many cases) *irrelevant* to whether they have any *other* properties in common. Carnap's most recent (and most complex) theories of logical probability (1980) include two additional parameters designed to provide the theory with enough flexibility to overcome these (and other) limitations. Unfortunately, no Carnapian logical theory of probability to date has dealt successfully with the problem of analogical effects (see Maher 2001 for further discussion). The consensus now seems to be that the Carnapian project of constructing an adequate logical theory of probability is all but hopeless: The syntactical constraints implicit in any such theory will inevitably prevent the theory from being able to model certain essential features of statistical inference and/or inductive logic (see Inductive Logic).

Subjectivism

In slogan form, subjectivism regards probabilities as *degrees of belief,* or *credences.* But what are credences? Subjectivists since Ramsey ([1926] 1980) have insisted that they must be intimately tied to the behavioral dispositions of suitable agents. In one influential account, advocated by de Finetti ([1937] 1980):

> An agent's credence in *e* is *p* iff *p* units of utility is the price at which the agent would buy or sell a bet that pays 1 unit of utility if *e*, 0 if $\neg e$.

This is at best a first approximation to an analysis of credence. One surely should allow the buying and selling prices of at least some bets to come apart. And even when they agree, there are problems. How does one separate the agent's epistemic attitude to *e* from his or her attitude (favorable, unfavorable, or neutral) to gambling? Indeed, one may insist on separating epistemic attitudes from desire-based attitudes altogether; one can imagine, for example, a chronic apathetic who has opinions but lacks corresponding desires (for bets or for anything). Moreover, the very placement of the bet may change the world in ways that affect the agent's credences.

Be that as it may, there are famous arguments that credences must conform to the probability calculus, at least if one demands that the agent be in some sense *ideally rational.* For example, if one's credences do *not* so conform, one is susceptible to a **Dutch Book**, a sequence of bets that one regards as acceptable taken individually but that collectively guarantee one's loss, however the world turns out. Conversely, if one's credences *do* so conform, one is immune to a Dutch Book. Rationality, it is concluded, requires obedience to the probability calculus (see Dutch Book Argument).

Utilities (desirabilities) of outcomes, their probabilities, and rational preferences are all intimately linked. The *Port Royal Logic* (Arnauld [1662] 1964) showed how utilities and probabilities together determine rational preferences; de Finetti's betting interpretation derives probabilities from utilities and rational preferences; von Neumann and Morgenstern (1944) derive utilities from probabilities and rational preferences. And most remarkably, Ramsey ([1926] 1980) (and later, Savage 1954 and Jeffrey 1983) derives *both* probabilities *and* utilities from rational preferences alone (see Ramsey, Frank Plumpton).

First, Ramsey defines a proposition to be *ethically neutral*—relative to an agent and an outcome—if the agent is indifferent between having that outcome when the proposition is true and when it is false. Suppose that the agent prefers *a* to *b.* Then an ethically neutral proposition *n* has probability $\frac{1}{2}$ iff the agent is indifferent between the gambles

a if *n*, *b* if not.
b if *n*, *a* if not.

One may assign arbitrarily to *a* and *b* any two real numbers $u(a)$ and $u(b)$ such that $u(a) > u(b)$, thought of as their respective desirabilities. Having done this for the one arbitrarily chosen pair *a* and *b*, the utilities of all other propositions are determined. Given various assumptions about the richness of the preference space, and certain "consistency assumptions," Ramsey can define a real-valued utility function of the outcomes *a, b*, etc.—in fact, various such functions will represent the agent's preferences. He is then able to define equality of differences in utility for any outcomes over which the agent has preferences. It turns out that ratios of utility differences are invariant—the same whichever representative utility function one chooses. This fact allows Ramsey to define degrees of belief as ratios of such differences. For example, suppose the agent is indifferent between *a* and the gamble "*b* if *x*, *c* otherwise." Then his or her degree of belief in *x*, $P(x)$, is given by:

$$P(x) = \frac{u(a) - u(c)}{u(b) - u(c)}.$$

Ramsey shows that degrees of belief so derived obey the probability calculus (with finite additivity). He calls what results "the logic of partial belief."

Ramsey avoids some of the objections to the betting interpretation, but not all of them. Notably, the essential appeal to gambles again raises the concern that the wrong quantities are being measured. And his account has new difficulties. It is unclear what facts about agents fix their preference rankings. It is also dubious that *consistency* alone requires one to have a set of preferences as rich as Ramsey requires, or that one can find ethically neutral propositions of probability $\frac{1}{2}$ This in turn casts some doubt on Ramsey's claim to assimilate probability theory to logic.

Savage (1954) likewise derives probabilities and utilities from preferences among options that are constrained by certain putative "rationality" principles. For a given set of such preferences, he generates a class of utility functions, each a positive linear transformation of the other (i.e., of the form $u_1 = au_2 + b$, where $a > 0$) and a unique probability function. Together these are said to "represent" the agent's preferences. Jeffrey (1983) refines the method further. The result is theory of decision according to which rational choice maximizes "expected utility," a certain probability-weighted average of utilities.

So far, this is a static picture of a rational agent. How should one update one's degrees of belief in the light of new evidence? The favored rule among subjectivists is *conditionalization:* Where e is the strongest proposition of which one becomes certain, one's new credence function is related to the old by:

$$\text{(Conditionalization)} \quad C_{new}(\bullet) = C_{old}(\bullet \mid e),$$

using $C(\bullet)$ here and in what follows to distinguish credence from other kinds of probability.

So-called *subjective Bayesianism* holds that an agent's epistemic trajectory is rational iff the agent's credences are representable at any moment by a probability function and the agent always updates by conditionalization. This is at once a highly demanding and highly permissive epistemology. It is demanding because conformity to probability theory is demanding. It is permissive because there is no requirement that degrees of belief in any way correspond to the way the world is. So someone who assigns probability 1 to the universe being ruled by a rubber chicken can meet the Bayesian standards for rationality—as long as the agent obeys the probability calculus in all other assignments and always updates by conditionalizing. Bayesians reply that various convergence theorems show roughly that in the long run, agents who do not give probability 0 to genuine possibilities, and whose stream of evidence is sufficiently rich, will eventually be arbitrarily close to being certain regarding the truth about the world in which they live. For skepticism about the value of these theorems, see Earman (1992).

In any case, there are numerous proposals for further constraints on priors. Some (e.g., Jeffreys and Jaynes) appeal to a version of the principle of indifference. Some can be regarded as instances of a certain schema, proposed by Gaifman (1988). He coins the term "expert probability" for a probability assignment that a given agent strives to track, codifying this idea as follows:

$$\text{(Expert)} \quad C(a \mid pr(a) = x) = x, \text{for all } x \text{ such that } C(pr(a) = x) > 0.$$

Here $pr(a)$ is the assignment that the agent regards as expert. For example, if one regards the local weather forecaster as an expert, and he or she assigns probability 0.1 to it raining tomorrow, then one may well follow suit:

$$C(\text{rain} \mid pr(\text{rain}) = 0.1) = 0.1.$$

More generally, one might speak of an entire probability function as being such a guide for an agent, over a specified set of propositions—so that (Expert) holds for any choice of A from that set. A *universal expert function* would guide *all* of the agent's probability assignments in this way. Van Fraassen (1995) argues that an agent's *future* probability functions are universal expert functions for that agent—his *reflection principle* is:

$$C_t(a \mid C_{t'}(a) = x) = x, \text{for all } a \text{ and for all } x \text{ such that } C_t(C_{t'}(a) = x) > 0,$$

where C_t is the agent's probability function at time t, and $C_{t'}$ his or her function at a later time t'. The principle encapsulates a certain demand for "diachronic coherence" imposed by rationality. Van Fraassen defends it with a diachronic Dutch Book argument (one that considers bets placed at different times) and by analogizing violations of it to the sort of pragmatic inconsistency that one finds in Moore's paradox. For example, suppose an agent is certain that he will tomorrow assign probability $\frac{1}{2}$ to it raining the day after but that he nonetheless assigns it probability $\frac{1}{3}$ now. While this is not logically inconsistent, it is surely puzzling.

One may go still further. There may be universal expert functions for all rational agents. The *principle of direct probability* regards the relative frequency function as a universal expert function. Let a be an event type, and let $relfreq(a)$ be the relative frequency of a (in some suitable reference class). Then for any rational agent:

$$C(a \mid relfreq(a) = x) = x, \text{for all } a \text{ and for all } x \text{ such that } C(relfreq(a) = x) > 0.$$

The next section takes up what many consider the most important such universal expert function.

Objective Chance

De Finetti (1990, x), the great probabilist, quipped that "probability does not exist." What he meant was that all probability is subjective. Yet there is a strong *prima facie* case for recognizing the existence of *objective chances:* probabilities that attach to physical systems and their behavior independently of anyone's mental state and that capture *contingent* facts about those systems, not merely quasi-logical relations among propositions concerning them. One wonders, for example, whether a certain coin is biased, and if so, to what degree and in what direction. Translation: One wonders what the *chance of heads* would be were the coin tossed fairly.

This example would not faze a committed subjectivist such as de Finetti or a frequentist like von Mises (1957), who denies that probability ever applies in the single case. But more serious examples from physics suggest that ultimately, resistance is futile. For starters, *statistical physics* says that the entire universe *could* evolve toward a state of lower entropy but that the chance of its doing so is vanishingly small. Such chances are seemingly compatible with determinism at the level of the fundamental dynamical laws (this is controversial; see Irreversibility; Statistical Mechanics), and following Bernoulli, various authors have for that reason doubted their credentials (however, see Levi 1990 for a valuable discussion of authors since Cournot and Venn who countenance such compatibilist chances). The compatibility is putatively secured by relativizing chances to a kind of trial. For example, a coin may have chance $\frac{1}{2}$ of landing heads relative to the specification of being tossed high above a flat surface, while it has a chance 1 of landing heads relative to a precise specification of the initial conditions of a particular toss in a deterministic world. Levi argues that the applicability of chance hypotheses and statistical techniques does not presuppose an underlying indeterminism, and so a theory of chance should remain neutral vis-à-vis determinism.

In any case, commitment to chances has a second source: collapse theories of quantum mechanics. These theories explicitly introduce indeterministic dynamical laws that not only specify what courses of evolution are possible for a given physical system with a given initial state, but also specify exact *probabilities* for each such trajectory (see Quantum Measurement Problem; Quantum Mechanics). The subjectivist or von Mises–style frequentist seems left only with the option of denying—from the armchair!—that the physical theories that postulate them are true or coherent. It is better simply to acknowledge that objective chances are or at least could be real, and then to go on to consider what sort of account one could give of them.

Reductionist accounts attempt to reduce facts about objective chances to the totality of nonmodal facts about a world. Nonreductionist accounts deny that chance even supervenes on the nonmodal facts. Actual frequentism is clearly a reductionist view. More sophisticated is Lewis's (1994) "best systems" approach, which sees the laws for a world *W* as, roughly, being theorems of that axiomatic system for describing nonmodal facts about *W* that achieves an optimal balance of simplicity and informativeness. In the case of probabilistic laws, Lewis invokes a third criterion: A system for *W* is "better" to the extent that it assigns a higher probability to the total history of *W*. The simplest form of nonreductionism is *primitivism*, which takes chances to be unanalyzable features of the world. Alternatively, one might try to explain one's nonreductionist chances by appeal to some other bit of metaphysical gadgetry, such as Armstrong's interpretation of chances as consisting in higher-order relations of "probabilification" that obtain between universals.

No mention has been made so far of "propensity" interpretations of probability. Everyone, reductionist and nonreductionist alike, can agree that in a chancy world, some physical systems will have propensities to exhibit certain behaviors under certain conditions, for all one need *mean* by that is that counterfactuals of the following form are true of these systems: "Were the system in conditions *C*, there would be a chance of *x* that it would manifest behavior *B*." So propensities—understood as tendencies, or variable-strength dispositions—can be analyzed straightforwardly in terms of subjunctive conditionals whose consequents make reference to objective chances.

Propensity interpretations of probability aim to reverse this order of analysis, explaining objective chances directly in terms of propensities. For some authors (Popper, Gillies), chances are dispositions for a chance setup to produce long-run relative frequencies; for others (Giere, Fetzer, Miller), they are dispositions for a chance setup to produce outcomes on single trials. Subtle variations can be found in the work of Hacking, Mellor, and Levi (see Gillies 2000 and Hájek 2003a for surveys.)

But there are some general problems that any propensity account faces. Suppose that system *S* has a certain tendency to manifest behavior *B* under conditions *C*. One must be able to attach numbers to such a tendency as a measure of its strength *without* appealing to the concept of chance; it is not clear how this is to be done or why the results should obey the probability calculus. Moreover, how do propensities for distinct systems yield propensities for the composite systems they make up? Here are two coins, each with a propensity of 0.5 of landing heads if tossed. Suppose both are tossed at once. If there is a chance that both will land heads, then there must be a propensity possessed by the *combined* two-coin system. If so, what guarantees that the marginal probabilities (for each coin considered separately) will be recovered correctly from this composite propensity? And one cannot stop here, but had better say that *the world as a whole* exhibits, at each moment, propensities to evolve in various different ways. Having gone thus far, one might as well simply say that instead of exhibiting propensities, the world exhibits *chances,* thus avoiding (by stipulation) the original problem of their conformity to the probability calculus—and thus arriving at primitivism about chance. If that is right, then it is not clear that propensity accounts offer a genuinely new option for understanding probability.

Although distinct, objective and subjective probability display an extremely important connection. Lewis (1980) formulates it in his principal principle:

$$\text{(PP)} \quad C_0(a \mid e \wedge ch_t(a) = x) = x.$$

Here C_0 is some reasonable "initial" (a priori) credence function; a an arbitrary proposition; $ch_t(a) = x$ the claim that the chance, at time t, of a is x; and e an "admissible" proposition—one that does not contain information relevant to a beyond that given by its chance at t (thus, e.g., a itself is inadmissible).

One can apply (PP) to a non-initial agent by modeling credence C as the result of conditionalizing some reasonable initial credence C_0 on some suitable evidence. Let h describe a complete possible course of history until time t. Let l describe some possible fundamental laws compatible with h, and assume that the way in which chances depend on history is underwritten by these laws. Then the conjunction $h \wedge l$ picks out a unique chance distribution $P(\bullet)$ for time t. Thus, if a proposition of the form $(h \wedge l \wedge ch_t(a) = x)$ is consistent, then the third conjunct is entailed by the first two. Assuming, as seems reasonable, that the conjunction $h \wedge l$ is admissible, it follows that

$$\text{(PP}^*\text{)} \quad C_0(a \mid h \wedge l) = P(a).$$

Much of the debate between reductionists and nonreductionists consists in a war of intuitions. For instance, the reductionist claims to find the nonreductionist's extra, irreducibly modal feature of metaphysical reality unintelligible, while the nonreductionist claims to "show" that distinct chances can give rise to exactly the same total histories of nonmodal fact—a draw, perhaps. But (PP) and (PP*) appear to open up new lines of argument.

The nonreductionist alleges that reductionism is *inconsistent* with (PP*). Typical reductionist views will allow that the chance laws can have some nonzero chance of failing to obtain, for the reductionist says that these laws are determined by the total history of nonmodal fact. But these laws issue in chance distributions over possible total histories of nonmodal fact. Thus, it may turn out that positive chances are assigned to total histories that would specify *different* laws—the "undermining" of the chance laws by themselves (see Lewis 1994). Example: A coin is about to be tossed exactly 10^{10} times. As it happens, exactly half the tosses will land heads. A reductionist might say that it *follows* that the chance of heads on each toss is 0.5, adding that the correct chance laws will treat the tosses as independent. So there is now a large chance that the frequency of heads will be *different* from what it actually is—and if so, the *laws* will be different as well.

The inconsistency with (PP*) is now manifest. Consider those consistent history–law conjunctions $h \wedge l$ that entail that $P(l) < 1$. Pick such a conjunction; by (PP*),

$$C_0(l \mid h \wedge l) = P(l) < 1.$$

But by the probability calculus,

$$C_0(l \mid h \wedge l) = 1.$$

Lewis responds by amending (PP*) to what he calls the "new principle":

$$\text{(NP)} \quad C_0(a \mid h \wedge l) = P(a \mid l),$$

thus avoiding the inconsistency. The consensus in the literature seems to be that unlike the original principal principle, (NP) is unintuitive and, in application, unwieldy.

The reductionist (e.g., Lewis 1994) retorts that nonreductionists are hard-pressed to show how chances, understood their way, constrain rational credences according to (PP). But can the *reductionist* meet this challenge? Presumably, in addition to his reductionist analysis of chance, he ought to provide a *derivation* of (PP) from constraints on rational credence to which he is *already* committed. The literature provides no such derivation. And while a nonreductionist may also be unable to supply such a derivation, it is not clear why it would be needed. Arguably, both the reductionist and the nonreductionist are committed to the existence of substantive constraints on rational credence; why can't the nonreductionist simply include (PP) as one of them? (See Hall 2003 for further discussion.) Perhaps, then, the debate between reductionism and nonreductionism remains a stalemate.

Finally, a "deflationary" account of chance, associated with de Finetti and his followers, has proved to be very influential. Consider an infinite exchangeable sequence of events with respect to a probability function P. De Finetti's representation theorem states that the probability according to P of exactly k of the events occurring in n trials is given by

$$\int_0^1 \binom{n}{k} p^k (1-p)^{n-k} f(p)\,dp,$$

for all n and k and for some density function f. The upshot is that any such probability distribution is representable as a "weighted average" of distributions. Each distribution corresponds to a hypothesis about the value of the probability p of an event occurring on a single trial; it gives the probability of k such events occurring in n independent, identically distributed trials, given that fixed value of p. One can then average these distributions using the probabilities of their corresponding hypotheses about the value of p as weights. The result is significant because it enables a subjectivist to "simulate" being an objectivist about chance when the exchangeability assumption holds, and for many situations this seems reasonable. If P is one's subjective probability function, then it is *as if* one spread probability over various hypotheses about the single-case objective chance of the event, which remains fixed across infinitely many independent trials of the experiment in question. (See Skyrms 1994 for an excellent discussion of generalizations of exchangeability and their use in formulating various Goodmanian theses about projectability.) Indeed, common sense often (but not invariably) seems to require that probabilities be exchangeable over "green"-like hypotheses but not "grue"-like hypotheses.

Conclusion

Feller (1957, 19) writes: "All possible definitions of probability fall short of the actual practice." Certainly, a lot is asked of the concept of probability. It is supposed at once to capture a quasi-logical notion, a subjective notion, and an objective notion instantiated in the mind-independent world. Perhaps one would do better to think of these as distinct *concepts* of probability. Each of the leading interpretations, then, attempts to illuminate one of these concepts, while leaving the others in the dark. In that sense, the interpretations might be regarded as complementary, although to be sure each may need some further refinement. Clearly, much work remains to be done on the philosophical foundations of probability. Equally clearly, the field has come a long way since the *Port Royal Logic*.

BRANDEN FITELSON
ALAN HÁJEK
NED HALL

References

Arnauld, Antoine ([1662] 1964), *Logic, or, The Art of Thinking ("The Port Royal Logic")*. Translated by J. Dickoff and P. James. Indianapolis, IN: Bobbs-Merrill.

Billingsley, Patrick (1995), *Probability and Measure*, 3rd ed. New York: John Wiley & Sons.

Carnap, Rudolf (1950), *Logical Foundations of Probability*. Chicago: University of Chicago Press.

——— (1952), *The Continuum of Inductive Methods*. Chicago: University of Chicago Press.

——— (1962), *Logical Foundations of Probability*, 2nd ed. Chicago: University of Chicago Press.

——— (1980), "*A Basic System of Inductive Logic II*," in Richard Jeffrey (ed.), *Studies in Inductive Logic and Probability, Volume II*. Berkeley and Los Angeles: University of California Press, 7–155.

de Finetti, Bruno ([1937] 1980), "*Foresight: Its Logical Laws, Its Subjective Sources*," in H. E. Kyburg Jr. and H. E. Smokler (eds.), *Studies in Subjective Probability*, 2nd ed. Huntington, NY: Robert E. Krieger, 53–118. Originally published as "*La prévision: Ses lois logiques, ses sources subjectives*," *Annales de l'Institut Henri Poincaré*, 7: 1–68.

——— (1990), *Theory of Probability*, vol. 1. Chichester, UK: Wiley Classics Library/Wiley & Sons.

Earman, John (1992), *Bayes or Bust?* Cambridge, MA: MIT Press.

Feller, William (1957), *An Introduction to Probability Theory and Its Applications*. New York: John Wiley & Sons, Inc.

Fine, Terrence (1973), *Theories of Probability*. New York: Academic Press.

Gaifman, Haim (1988), "*A Theory of Higher Order Probabilities*," in Brian Skyrms and William L. Harper (eds.), *Causation, Chance, and Credence*. Dordrecht, Holland: Kluwer Academic Publishers.

Gillies, Donald (2000), "*Varieties of Propensity*," *British Journal for the Philosophy of Science* 51: 807–835.

Hájek, Alan (2003a), "*Interpretations of Probability*," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/sum2003/entries/probability-interpret

——— (2003b), "*What Conditional Probability Could Not Be*," *Synthese* 137: 273–323.

Hall, Ned (2003), "*Two Mistakes about Credence and Chance*," *Australasian Journal of Philosophy*: 93–111.

Jeffrey, Richard (1983), *The Logic of Decision*, 2nd ed. Chicago: University of Chicago Press.

Johnson, William E. (1921), *Logic*. Cambridge: Cambridge University Press.

Keynes, John M. (1921), *A Treatise on Probability*. London: Macmillan.

Kolmogorov, Andrei. N. ([1933] 1950), *Foundations of Probability* [*Grundbegriffe der Wahrscheinlichkeitrechnung, Ergebnisse der Mathematik*]. New York: Chelsea Publishing.

Laplace, Pierre Simon ([1814] 1951), *A Philosophical Essay on Probabilities*. New York: Dover Publications.

Levi, Isaac (1990), "*Chance*," *Philosophical Topics* 18: 117–148.

Lewis, David (1986a), *Philosophical Papers: Volume II*. New York: Oxford University Press.

——— (1986b), "*A Subjectivist's Guide to Objective Chance*," in *Philosophical Papers: Volume II*. New York: Oxford University Press, 83–132.

——— (1994), "*Humean Supervenience Debugged*," *Mind* 103: 473–490.

Maher, Patrick (2001), "*Probabilities for Multiple Properties: The Models of Hesse and Carnap and Kemeny*," *Erkenntnis* 55: 183–216.

Mückenheim, W., G. Ludwig, C. Dewdney, P. Holland, A. Kyprianidis, J. Vigier, N. Petroni, M. Bartlett, and

E. Jaynes (1986), "*A Review of Extended Probability*," *Physics Reports* 133: 337–401.

Popper, Karl (1959), *The Logic of Scientific Discovery*. London: Hutchinson & Co.

Ramsey, Frank P. ([1926] 1980), "*Truth and Probability*," in H. E. Kyburg Jr. and H. E. Smokler (eds.), *Studies in Subjective Probability*, 2nd ed. Huntington, NY: R. E. Krieger, 23–52.

Renyi, Alfred (1970), *Foundations of Probability*. San Francisco: Holden-Day.

Roeper, Peter, and Hughes Leblanc (1999), *Probability Theory and Probability Logic*. Toronto: University of Toronto Press.

Savage, Leonard J. (1954), *The Foundations of Statistics*. New York: John Wiley.

Skyrms, Brian (1980), *Causal Necessity*. New Haven, CT: Yale University Press.

——— (1994), "*Bayesian Projectibility*," in D. Stalker (ed.), *Grue: Essays on the New Riddle of Induction*. Chicago: Open Court, 241–262.

van Fraassen, Bas (1995), "*Belief and the Problem of Ulysses and the Sirens*," *Philosophical Studies* 77: 7–37.

Venn, John (1866), *The Logic of Chance*. London and Cambridge: Macmillan.

von Mises, Richard (1957), *Probability, Statistics and Truth*. New York: Macmillan.

von Neumann, John, and Oskar Morgenstern (1944), *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

PROGRESS

See **Scientific Progress**

PROTOCOL SENTENCES

Protocol sentences are reports of an individual's experiences. The simplest and paradigmatic example is a report of "red here now." The term 'protocol sentence' was introduced by Rudolf Carnap (1932a, b) (and the example here is his). It reflects two chief aspects of logical empiricism (or positivism): (i) the importance of linguistic form and (ii) the role of experience as the source of acceptability and cognitive significance of scientific beliefs (see Logical Empiricism). The form, role, and status of protocol sentences became the topic of an important philosophical debate in the early 1930s involving logical empiricists such as Carnap, Moritz Schlick, Edgar Zilsel, Otto Neurath, Karl Popper, and a few others.

In *Der logische Afbau der Welt*, Carnap ([1928] 1967) investigated the logical "construction" of objects of intersubjective knowledge from several possible bases. These included a physicalist basis, but Carnap epistemically privileged a basis consisting of the autopsychological objects of private sense experience, termed "elementary experiences," which were supposed to provide the simplest and natural starting point for epistemological constructions. The main problem then became how objective knowledge is possible. To solve that problem required the existence of connecting statements or rules linking those other statements to these epistemically privileged experiential statements.

The *Aufbau*'s focus on an empiricist or phenomenalist model of knowledge in terms of the immediate experiential basis was interpreted by other members of the Vienna Circle as manifesting three philosophical positions: reductionism, atomism, and foundationalism. Reductionism took one set of terms to be fundamental or primitive; the rest would be logically derived from them. Atomism, especially in Neurath's reading, was manifest in the elementary structure of protocol sentences each as a single experiential report, such as "red here now." Foundationalism took autopsychological beliefs to be infallible and as epistemic warrant for all other beliefs.

Neurath rejected Carnap's subjectivism on the grounds that if the language and the system of statements that constitute scientific knowledge are intersubjective, then phenomenalist talk of immediate subjective, private experiences should have no place. To replace Carnap's phenomenalist language, Neurath (1932) introduced a physicalist language. Along with it came the thesis of *physicalism:* The unity, intelligibility, and objectivity of science rests on statements in a language of public things, events, and processes in space and time, including behavioral and physiological events. While inspired by materialism, for Neurath this was a methodological and linguistic rule, and not an ontological thesis.

Following Neurath, Carnap also explicitly drew a contrast between the language of experience and an intersubjective physicalist language. After the *Aufbau,* the unity of science rested on the universal possibility of translation of any scientific statement into this physicalist language. The physicalist language was intersubjective because it was also intersensual, that is, translatable into the private protocol language of any type of sensory experience of any subject, such as the atomic elementary statement reporting "red here now." For Carnap, a protocol language was a subjective language for each subject and, by the translation rules, also part of the physicalist language. However, it was still the epistemic point of departure for all scientific theorizing. The protocol language thus provided the tools for empirical verification in any language or knowledge system. It is the language of statements that have the function of control, support, or warrant. Even with its limitations, it is an indispensable source of knowledge:

> [A]n inferential connection between the protocol sentences and the singular physical statements must exist, for if, from the physical statements, nothing can be deduced as to the truth or falsity of the protocol statements there would be no connection between scientific knowledge and experience. Physical statements would float in a void disconnected, in principle, from all experience. (Carnap 1932a, 81)

Neurath (1932) responded to Carnap with a different account of protocol sentences that considered their distinctive linguistic form, contents, and methodological status. This account was supposed to explicate the concept of scientific evidence in an empiricist framework, by specifying the conditions for acceptance of a statement as empirical scientific evidence. The doctrine was meant to circumvent the pitfalls of the alleged subjectivism, atomism, reductionism, and foundationalism attributed to Carnap's earlier discussion in the *Aufbau.*

Unlike Carnap's ideal of a basic statement, whether as a protocol sentence or in the physicalist language, Neurath's protocol sentences were not clean, precise, or pure. For Neurath, the physical language, and hence science in turn, was inseparable from ordinary language of any time and place. In particular, it is muddled with imprecise, unanalyzed, cluster-like terms (*Ballungen*) that appear especially in the protocol sentences. They were often to be further analyzable into more precise terms or mathematical coordinations, but they often could not be eliminated. Even the empirical character of protocol sentences was not pure and primitive, as physicalism allowed the introduction of theoretical terms.

The empirical genealogy expressed by protocol sentences was the linguistic expression of empiricism that identified protocol sentences as epistemologically special units in the system of statements that would constitute objective scientific knowledge. This was both Carnap's and Neurath's attempt to develop a full-fledged empiricist position. Yet another difference from Carnap's account of protocol sentences appears in the second aspect of Neurath's account, in their role in providing empirical tests. Neurath's protocol sentences did not have the atomic structure or the atomic testing role that Carnap's did. Their methodological role reflected Duhem's holism (see Duhem Thesis). Hypotheses are not tested individually; only clusters of statements confront empirical data. Their methodological value in the testing of other statements did not make Neurath's protocol sentences unrevisable. Carnap (1932b and 1934) later adopted a similar conventionalist and pragmatic attitude. Such is the anti-Cartesian nonfoundationalism and fallibilism of Neurath's account. The system of knowledge is constrained by historically and theoretically accepted terms and beliefs and cannot be rebuilt on pure, secure, infallible empirical foundations. This was made only worse by the presence of *Ballung*-type terms. The method of testing could not be carried out in a logically precise, determinate, and conclusive manner. Protocol sentences, by virtue of perception terms, would provide a certain stability in the permanence of information.

But methodologically they could only bolster or shake confidence. Reasons underdetermine actions, and thus pragmatic extralogical factors were required to make decisions about what hypotheses to accept. A loose coherentist view of justification and unification is the only option available. To acknowledge these limitations is a mark of rationality, which Neurath opposed to "pseudorationalism." (see Neurath, Otto).

Carnap came to share Neurath's epistemological attitude (especially in *Logical Syntax*), but found the complexity of Neurath's protocol sentences excessive, raising too many practical difficulties. He also believed that their self-referential character raised logical difficulties. In reaction to Neurath's proposal, Carnap (1932b) adopted a distinct conventionalist and pragmatist attitude toward protocol sentences. He distinguished his early protocol theory from Neurath's (1932) proposal and another, unpublished one from Popper. He believed, under Popper's influence, that any concrete physicalist statement could be used as a protocol or basic sentence. The foundation or stopping-point of reduction or validation was achieved pragmatically. Neurath agreed only in that the equivalence between Carnap's physicalist basic statements and his own protocol sentences was logical or functional because both were revisable and that their acceptance was ultimately pragmatic. He lamented, however, that the conventionalist Carnap–Popper standpoint had cut itself loose from any references to perception, and thus amounted to abandoning empiricism.

In 1936, after adopting the semantic turn—in contrast to the syntactic approach of *Logical Syntax*—Carnap defended the distinctive character of protocol sentences in science provided by observational predicates. While defending a distinction between theoretical and observational language, he acknowledged that the distinction could not be made precise and fixed. The observational character was not a logical or syntactic property, but a psychological and instrumental one decided in the course of actual scientific practice.

By 1934, Popper had adopted an approach to scientific knowledge based on the logic of method, not on meaning. For Popper, any talk of individual experience could have no linguistic expression; a theory of scientific knowledge was to be not a subjective or descriptive account, but a normative logic of justification and demarcation (Popper [1935] 1959). He also criticized Neurath's antifoundationalism about protocol sentences as a form of anti-empiricism that merely opened the door to dogmatism or arbitrariness. Instead of protocol sentences, Popper proposed to speak about "basic statements"—a term more attuned to their logical and functional role. They are basic relative to a theory under test and are singular existential statements reporting observable facts. But acceptance of basic statements, much as in Neurath's account, could in principle be revoked.

Finally, the most radical empiricist attitude toward protocol sentences within the Vienna Circle came from Schlick, who endorsed a formal, structural notion of communicable, objective knowledge and meaning as well as a correspondence theory of truth. His realism opposed Neurath's coherentism, as well as the pragmatism and conventionalism of Carnap. In 1934, Schlick proposed to treat protocol sentences, left by Neurath with the status of little more than mere hypotheses, as key to the foundation of knowledge. They would be physicalist statements that, albeit fallible, could be subjectively linked to statements about immediate private experiences of reality, such as "red here now," which he called *affirmations* (Schlick 1934). Affirmations carried certainty and elucidated what could be shown but not said. They provided the elusive confrontation or correspondence between theoretical propositions and facts of reality. In this sense they afforded the fixed starting points and the foundation of all knowledge. But the foundation raised a psychological and semantic problem about the acceptance of a protocol sentence. Affirmations, as acts of verification or giving meaning, lacked logical inferential force; in Schlick's words, they "do not occur within science itself, and can neither be derived from scientific propositions, nor the latter from them" (Schlick 1934, 95). Schlick's empiricism regarding the role of protocol sentences suggests, but does not logically support, any strong epistemological foundationalism.

Neurath (1934 and 1935) replied to Schlick's and Popper's criticisms, respectively, emphasizing the role of extralogical factors in accepting theories and criticizing metaphysical talk of comparing knowledge and reality and the overidealized notion of a precise and conclusive logic of science. The discussion of the nature and value of protocol sentences was taken up subsequently by Hanson, Kuhn, Feyerabend, and others, all of whom stressed the theory-ladenness of observations (Feyerabend 1958 and 1962; Hanson 1961; Kuhn 1962). They denied that an absolute distinction between observation and theoretical predicates and statements was available and could ground the rationalism of scientific testing. Quine made famous the holistic idea of a *web of beliefs;* he and Davidson examined the role of experience by looking at the relation between empirical beliefs and reality (Quine 1963; Davidson 1987). The physicalist bridge they see is neural and causal, respectively, but not epistemological in an inferential sense. These bridges nevertheless play a crucial role in our attempts to fix meaning and belief.

JORDI CAT

References

Carnap, R. ([1928] 1967), *The Logical Structure of the World*. Berkeley and Los Angeles: University of California Press.

——— (1932a), "*Der physikalische sprache als Universalsprache der Wissenschaft*," *Erkenntnis* 2: 432–465.

——— (1932b), "*Ueber Protokollsaetze*," *Erkenntnis* 3: 215–228.

——— (1934), *The Logical Syntax of Language*. London: Kegan Paul, Trench, Trubner & Co.

Carnap, R. (1950), "*Empiricism, Semantics and Ontology*," *Revue Internationale de Philosophie* 4: 20–40.

Davidson, D. (1987), "*Empirical Content*," in E. Lepore (ed.), *Essays on Truth and Interpretation*. Oxford: Oxford University Press.

Feyerabend, P. K. (1958), "*An Attempt at a Realistic Interpretation of Experience*," *Proceedings of the Aristotelian Society* 58: 143–170.

——— (1962), "*Explanation, Reduction, and Empiricism*," in H. Feigl and G. Maxwell (eds.), *Scientific Explanation, Space, and Time*. Minneapolis: University of Minnesota Press, 28–97.

Galison, P. (1989), "*Aufbau/Bauhaus: Logical Positivism and Architectural Modernism*," *Critical Inquiry* 16: 709–752.

Hanson, R. N. (1961), *Patterns of Discovery*. Cambridge: Cambridge University Press.

Kuhn, T. S. (1962), *The Structure of Scientific Revolutions*. Chicago: Chicago University Press.

Neurath, O. (1931), "*Physikalismus*," *Scientia* 50: 297–303 (trans. in Neurath 1983).

——— (1932), "*Protokollsaetze*," *Erkenntnis* 3: 281–288.

——— (1934), "*Radikaler Physikalismus und 'wirkliche Welt,'*" *Erkenntnis* 4: 346–363.

——— (1935), "*Pseudorationalismus der Falsification*," *Erkenntnis* 5: 353–365.

——— (1983), *Philosophical Papers, 1913–1946*. Edited by M. Neurath and R. Cohen. Dordrecht, Netherlands: Reidel.

Popper, K.R. ([1935] 1959), *The Logic of Scientific Discovery*. London: Hutchinson.

Quine, Willard Van (1963), *From a Logical Point of View*. New York: Harper and Row.

Schlick, M. ([1918] 1925), *General Theory of Knowledge*. LaSalle, IL: Open Court.

Schlick, M. (1930), "*Die Wende in der Philosophie*," *Erkenntnis* 1: 4–11 (trans. in Schlick 1979).

Schlick, M. (1932), "*Positivismus und Realismus*," *Erkenntnis* 3: 1–31 (trans. in Schlick 1979).

Schlick, M. (1934), "*Ueber das Fundament der Erkenntnis*," *Erkenntnis* 4: 79–99 (trans. in Schlick 1979).

Schlick, M. (1979), *Philosophical Papers, Vol. 2 (1925–1936)*. Dordrecht, Netherlands: Reidel.

See also **Carnap, Rudolf; Cognitive Significance; Demarcation, Problem of; Feyerabend, Paul; Hanson, Norwood Russell; Kuhn, Thomas; Logical Empiricism; Neurath, Otto; Popper, Karl Raimund; Schlick, Moritz; Vienna Circle**

PSEUDOSCIENCE

See **Cognitive Significance; Demarcation, Problem of**

PHILOSOPHY OF PSYCHOLOGY

Traditionally, when general philosophy of science dominated the discipline, a simple division was often invoked to talk about philosophical issues specific to particular kinds of science: that between the natural sciences and the social sciences. Over the last twenty years, philosophical studies shaped around this dichotomy have given way to those organized by more fine-grained categories, corresponding to specific disciplines, as the literatures on the philosophy of physics, biology, economics, and psychology—to take the most prominent four examples—have blossomed. In general terms, work in each of these areas has become increasingly enmeshed with that in the corresponding science itself, and so increasingly naturalistic (in at least one sense of that term).

The philosophy of psychology, like psychology itself, is concerned with mind and cognition. When psychology cut itself loose—institutionally and professionally—from philosophy in the late nineteenth and early twentieth centuries, it was the discipline that predominantly studied mind and cognition. This has changed over the last thirty years. With the development of artificial intelligence, cognitive anthropology, linguistics, and neuroscience—perhaps, together with psychology, best referred to collectively as the "cognitive sciences"—philosophers of psychology have found themselves both drawing on and contributing to scientific work in this more interdisciplinary milieu. There are two consequences of this. The first is that the field has become increasingly entwined with the philosophical aspects of cognitive science. One view is that one does greater justice to the interdisciplinary motivations behind cognitive science by placing an emphasis on the cognitive sciences, rather than on foundational assumptions that constitute a single paradigm of cognitive science (see Cognitive Science). In this view, the philosophical aspects of the cognitive sciences occupy the greater part of the philosophy of psychology (cf. Wilson 1999). The second consequence is that the more lively areas or topics of contemporary discussion in the philosophy of psychology are quite diverse, including (for example) philosophical issues in neuroscience, the nature and physical bases of consciousness, the evolution of mind, and the ontogenetic and phylogenetic development of intentional states in human agents.

Despite the first of these points, and contributing to the second, the material that philosophers of psychology discuss also covers questions about the mind and areas of psychology that even a pluralistic conception of the cognitive sciences excludes. These issues include debates over the scientific status of psychoanalysis, questions about the foundations of the taxonomy of psychopathology, and discussions of the nature of social psychology, all of which concern areas of psychology other than cognitive psychology.

What further complicates any simple characterization of work in the philosophy of psychology, and to some extent what distinguishes it from the other "philosophy of *X*" studies within the philosophy of science, is its close relationship to a traditional area of philosophy—the philosophy of mind—that has not typically viewed itself as a part of the philosophy of science at all. Thus, many of the topics that philosophers of psychology discuss that arise from their reflection on the cognitive sciences have analogues in traditional philosophy of mind. For example, concerns about the causal role of semantic- or representational-level properties in computational theories of cognition echo the more general problem of mental causation; many of the issues about the nature of cognitive architecture that separate, for example, "classic" from connectionist approaches to cognitive architecture are also reflected in the historical debates between rationalists and empiricists. Perhaps because the nature of the mind has been one of the central issues in metaphysics and epistemology throughout the history of philosophy, the connections between the philosophy of psychology and philosophy more generally are more extensive than in any other disciplinarily specialized area of the philosophy of science. What follows will attempt to convey something of the flavor of three topics within the philosophy of psychology that have dominated the field over the last twenty years: intentionality, cognitive architecture, and consciousness. It will also briefly discuss another pair of more specific topic clusters that represent novel and perhaps trend-setting topics for future research.

Intentionality and Mental Representation

The postulation of mental representation has been central to the cognitive sciences throughout their history. Human agents do not simply or reflexively respond to their environments, but are equipped with some internal, mediating mental machinery, which is sensitive to what is in the environment but which has enough complexity to it to thwart any attempt (e.g., made by behaviorists) to exhaustively characterize it in terms of that environment (e.g., in terms of stimulus–response pairs) (see Behaviorism). Mental representations play precisely such a mediating role, both containing information about the world and combining to guide an individual's behavior in that world (see Intentionality).

The form that mental representation takes in commonsense, folk psychology is *propositional:* Agents have beliefs and desires, where each of these mental states can be thought of as an attitude to a proposition. Psychology has built on such folk psychological representations since its inception, from the Freudian extension of folk psychology from conscious to unconscious states, to work on stereotypes and schemata in social psychology, to classic artificial intelligence (AI) models of human problem solving or reasoning. Because of this link between folk and scientific psychology, propositional representation has been a focus of discussion within the philosophy of psychology. In fact, due to its prominence, many general discussions of mental representation have been cast exclusively in terms of propositional representation, or even its folk psychological guise. What follows constitutes three of the central issues in the literature and a sampling of positions that have been adopted with respect to them:

1. *How many kinds of mental representation are there?* Much of the debate over mental imagery (Pylyshyn, in press) has focused on the reality of mental images and their relationship to propositional representations. There has also been more recent discussion of the extent to which mental representations are "local" as opposed to "distributed" in their nature. The role of language in mental representation, and thus thought, has also structured a range of related debates, such as over the language of thought hypothesis and the question of the form that mental representation takes in nonlinguistic creatures, such as human infants and nonhuman animals.
2. *What determines a representation's content?* Three chief answers to this question have been entertained: conceptual role or procedural semantics, causal or informational theories, and teleological theories (see also Cognitive Science). The first of these is typically internalist in that mental content is determined entirely by intrinsic, physical properties of the agent or system. But the most pervasive views here are externalist; that is, they allow an individual's social or physical environment to be a determinant of the type of mental states the individual has, which reinforces externalist views of psychology and psychological explanation (Wilson 1995 and 2002). Both causal and teleological views allow an individual's historical, social, and physical location to partially determine what content its representations have. An alternative form of externalism that departs from the sort of realism about mental representation that has been taken for granted by the three chief views here is a conventionalism about the nature of representational content (see Horst 1996).
3. *Is mental representation dispensable within the cognitive sciences?* Stich (1983) was an early defender of the view that the cognitive sciences could be (indeed, should be) content free. Patricia S. Churchland (2002) has expressed an alternative, neuroscientifically inspired form of eliminativism about mental representation. Both of these forms of eliminativism about mental representation have pitched their critiques at the sorts of representations posited by folk psychology. Proponents of connectionist architectures and, more recently, of dynamic approaches to cognition have also often introduced their views as avoiding the postulation of mental representation. But as the descriptors 'distributed' and 'dynamic' suggest, such approaches do not necessarily imply the rejection of all forms of mental representation, and the place of mental representation within them remains a topic of continuing interest (Érdi 2000) (see Epistemology).

Cognitive Architecture and Processing

If debates over the nature of mental representation concern *what* it is that cognition ranges over, those over cognitive architecture and processing concern *how* it is that cognition proceeds. Part and parcel of the "cognitive revolution" of the late 1950s that formed the basis for the cognitive sciences was the conceptualization of cognitive processing as a form of *computation.* This view, computationalism, has received both general and somewhat vague characterizations ("cognition is computation") as well as more specific formulations ("cognition is explicit symbol manipulation") that are tied to particular research programs, the best known of which is the *physical symbol system hypothesis*, associated with Allen Newell and Herb Simon (1981): "A physical symbol system has the necessary and sufficient means for general intelligent action" (41). Central to any account of cognitive processing is a commitment to the nature of the basic design of the cognitive system, the *cognitive architecture* of that system, and hypotheses about cognitive architecture have usually been formulated as explicit computational models that generate behavior that approximates some aspect of (often human) cognitive behavior. In Newell and Simon's own view, production systems, which consist of chains of condition-action rules defined over data structures, form the heart of human cognitive architecture, and the types of behaviors to which their computational models were applied most extensively were problem solving and reasoning. Variations on this general view were predominant in much of AI and psychology until the 1980s, and the philosopher perhaps most firmly associated with this sort of "rules and representations" approach to cognitive architecture is Jerry Fodor (1981).

Over the past twenty years, connectionism has come to represent a general alternative to the rules and representation approach. The basic idea of connectionist architectures and the neural network models that correspond to them is that cognition involves the adjustment of weighted connections between many relatively simple processing units through a process of feedback from environmental inputs (learning). Although these basic units are often compared to neurons, the bulk of the psychological work to which philosophers appeal (e.g., in modeling the acquisition of the past tense in English) involves processing units that are on the wrong scale to be very neuron-like (see Connectionism).

The most fruitful work within the computational paradigm, broadly construed, involves models that appeal to aspects of both rules and representations and connectionist architectures. A common suggestion is that the former handles "higher" cognitive functions, such as problem solving, while the latter are applicable to "lower-level" cognition, such as pattern recognition. But more truly integrative models of cognitive architecture focus on the role that *probability* has within computational models; for example, Boltzmann machines, developed within the neural network paradigm, are essentially identical to Bayesian networks developed within traditional AI (Pearl 2000). The significance of such models is that they straddle the supposed divide between "classic" and "connectionist" architectures. Their rise within work on computational intelligence signals the next stage in cognitive modeling (see Jordan and Russell 1999).

Dynamic approaches to cognition attempt to pose a more radical challenge to these two views of cognitive architecture and their corresponding paradigms for the cognitive sciences. The chief idea of dynamicism is that cognitive systems are a form of dynamic system that exists in real time and whose movement over time is not governed by any special computational principles (Port and van Gelder 1995). On the dynamic conception of cognitive processing, internalized rules and symbols do not play any special role in cognition; rather, cognition proceeds through the settling of the cognitive system into an equilibrium state. The mathematical equations that govern such processes are not internalized within the cognizer any more than Newton's laws of motion are internalized in the objects whose behavior they govern. The dynamic approach has thus challenged both the representational and computational dimensions to standard cognitive science, and it also suggests that cognitive systems are fundamentally *embedded* or *embodied*, a point discussed further later.

The development of connectionist architectures has led many philosophers of psychology to rethink a range of issues concerning the nature of cognitive processing. Many of these concern the nature of mental representation, as noted above, but the rise of connectionism has also generated more general discussions, such as those over the nature of computation (including the relationship between computational models and computation) and the role that cognitive neuroscience has to play in addressing some of these questions about large-scale cognitive organization. Despite the fact that most of the neural network models of influence within the cognitive sciences are not neurally very realistic, connectionist architectures have redirected attention to the brain itself, particularly as noninvasive techniques of imaging, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), have allowed researchers to explore the activity of the brain in real time. One pair of related themes concerning cognitive architecture has been the modularity of cognitive design and the localization of mental processes. Fodor (1983), crystallizing and generalizing a view of the mind articulated within linguistics by Chomsky as part of his approach to generative grammar in linguistics, rekindled interest in a modular view of cognitive capacities of the sort introduced originally by Gall almost two hundred years earlier. According to Fodor's view, many such capacities are *domain specific* and *encapsulated:* roughly, cognition is structured so that particular mental organs are sensitive only to specific kinds of inputs and are insulated from the causal influence of the operation of other mental organs. Fodor's own view here was that such a view of the mind held only of input systems—the five senses, plus language, according to Fodor—and he cautioned against the extension of the view to "central systems." This caution has been largely ignored as developmental psychologists have postulated Fodorean modules for the domains of physics, number, biology, and psychology, and evolutionary psychologists have endorsed what has become known as the *massive modularity* thesis, the claim that the mind is overwhelmingly modular, with the number of modules running into the hundreds if not thousands (see Evolutionary Psychology). Philosophers have had much to say about these topics, particularly about the "theory of mind" within developmental psychology and evolutionary psychology in general.

It has typically been assumed that modules are *physically localized* in the brain, roughly in the way in which other bodily organs, such as the heart or the kidney, are so localized. As Fodor (1983) himself pointed out in his brief discussion of the "fixed neural architecture" (98–99) associated with modules, one might articulate this assumption in terms of broader systems that are somewhat distributed throughout the brain. But the basic idea is that functionally individuated modules have neural hardware specifically dedicated to the function they perform. As PET and fMRI have been increasingly used in experimental investigations of cognition, data on such localization assumptions have accumulated, though it is worth mentioning that these methods themselves have often been used in ways that presuppose a basically localistic view of cognitive function (Uttal 2001). Lloyd (2000) has presented a striking, even if preliminary, meta-analysis of the data across independent studies, arguing that these data support the claim that the brain is a distributed processor and refutes the stronger, "localistic," modularity hypotheses common in the field (see also Neurobiology).

Consciousness

Consciousness has been a buzz topic in the philosophy of psychology for the past ten years, returning to occupy center stage after a long absence, and commanding the attention both of philosophers of science (i.e., of psychology) and of traditional philosophers of mind (see Consciousness). Amongst the latter, there has been an explicit a priori strand, with a focus on the challenge that consciousness, phenomenal states, and qualia pose to views, such as physicalism and functionalism, that continue to operate as working assumptions for many within the cognitive sciences. A work that has galvanized such discussion is Chalmers' *The Conscious Mind* (1996), a book whose central conclusions echo the skepticism about physicalism associated with well-known, earlier papers by Thomas Nagel and Frank Jackson, and whose emphasis on conceivability arguments and what they putatively show about the limits to the scientific study of consciousness has fueled some interesting debate over the role of conceptual analysis within a naturalistic account of the mind. Philosophical work on consciousness has also reacted to attempts to eliminate qualia, developed representationalist views of the nature of phenomenal experience, and debated the idea that consciousness is just awareness, so that a conscious mental state is some sort of second- (or, in general, higher-) order mental state. A succinct overview of this work is provided by Levine (1997).

Within the cognitive sciences themselves during roughly the same period, consciousness also had a renaissance, with much of this literature focused on the phenomena of visual awareness and attention. There has been speculation in this literature about the function(s) and evolutionary origins of consciousness, as well as a variety of neural techniques to try to pinpoint the parts of the brain that are most directly causally responsible for conscious experience. In an influential paper, Crick and Koch (1990) advocated that the time was ripe for neural speculations about consciousness, and proposed (building on the work of von Malsburg and others) that 40-hertz oscillatory cycles in the brain, particularly in the visual cortex, were especially important to consciousness. A detailed, recent empirical account of consciousness has been offered by Rodney Cotterill (1998), which emphasizes the relationship between consciousness and movement and the importance of *timing* to consciousness. Cotterill offers an integrated psychological and neurological view of the bases for consciousness that posits a triangular neural circuit linking the posterior lobes, the premotor area of the frontal lobes, and the nucleus reticularis thalami between the thalamus and the medulla oblongata as the neural basis for conscious experience.

One obvious question concerns the relationship between such work on consciousness and that on the nonconscious, representational mental states that have been central to the cognitive sciences over the last thirty years. *Representationalism* about conscious states constitutes one sort of answer, for it holds that qualitative states just *are* representational states. Indeed, one of the motivations for representationalism is to deflate the commitments that one makes in admitting conscious mental states as well as intentional states to one's ontology. Another type of answer is provided by John Searle (1992), who has defended what he calls *the connection principle*, which says that unconscious mental states must be, in principle, accessible to consciousness (see Searle, John). This principle has one of two implications for traditional cognitive science: Either the states it posits do not exist or those states are, contrary to what most of those investigating them believe, accessible to consciousness. Given the disparate writings on consciousness, it is no surprise that some of these have become more explicitly self-reflective. Perhaps the best-known is by Block (1995), who introduced the distinction between "phenomenal consciousness," or *p*-consciousness (the what-it's-likeness of mental experience), and "access consciousness," or *a*-consciousness (the feature of mental experience that allows its reportability). One suggestion is that mental states such as pain and sensations are *p*-conscious, while those such as occurrent thoughts are *a*-conscious; another is that the former is really the subject of the literature inspired by Nagel, by Jackson, and by Chalmers, while the latter is what cognitive scientists investigate. Block himself introduced the distinction to critique claims, especially in the psychological literature, that were often made about the function of phenomenal consciousness that relied implicitly only on data about access consciousness. Philosophers remain divided over whether Block's distinction makes sense of much of the consciousness literature or constitutes a confusion about consciousness itself (see Consciousness).

Pain, Psychopathology, and Color

One of the concomitant products of the extended-consciousness fest has been work on topics concerning particular phenomenal states. *Color* is perhaps the most richly mined of these, beginning with C. L. Hardin's *Color for Philosophers* (1988), which significantly raised the bar regarding the level of empirical detail relevant to philosophical discussions of color. From its characterization as a secondary quality in seventeenth-century mechanical philosophy and science, color has constituted both an epistemic and an ontological puzzle: Just what is color, in the world, and does one's epistemic access to it constitute some sort of privileged knowledge? Some of the recent work on color processing in the cognitive sciences suggests that color is at least as much of an enigma in accounts of cognitive processing. For example, it now appears that there is no place or system in the central nervous system that is modularly dedicated to process color, and this has led some philosophers to rethink the evolutionary function of color perception and its role within the perceptual life of the individual (cf. Matthen 1999).

While psychopathology itself is not a new topic for philosophers, work here has taken a novel turn as a by-product of the focus on consciousness. Conscious experience sometimes deviates from its normal course. Philosophical issues abound here, whether it be in cases of blindsight in patients with severed corpus callosa, where subjects are causally influenced by phenomena of which they report no conscious awareness (Weiskrantz 1986), or in clinical breakdowns of the self, such as those involving "injected selves," or dissipated and disjoint mental lives (Graham and Stephens 1994). Clinical, medical, and cognitive psychologies have represented distinct traditions studying mental pathologies, and as they begin to share more common phenomena, data, and theoretical bases, there is an opportunity for philosophers of psychology not only to contribute to discussions of foundational questions about the nature of the self, rationality, and normative mental functioning, but also to bring together these discussions with those on each of the three topics with which this article began: mental representation, cognitive architecture, and consciousness. Pain is the third and newest of these topics within empirically attentive philosophy of psychology. The large community of researchers on pain have their home base in the medical sciences and have focused not so much on the theorization of pain as on its amelioration and treatment. Along with color, pain is the qualitative mental phenomenon most commonly invoked by philosophers discussing consciousness, and like color the empirical work on pain has exploded in recent years. There are sensory and affective dimensions to pain, where the former reflects the role of pain as a detector of bodily damage, and the latter, the phenomenal character of pain. Moreover, there turns out to be considerable interpersonal bodily variability for those experiencing pain. Conceptually, the sensory and affective dimensions of pain are distinct, and early empirical work offered support for the hypothesis that there are two separate pain systems. Dennett ([1978] 1998) used some of the complexities of the folk psychology commonsense conception of pain to argue for an eliminativist view of pain, and more recently philosophers have taken opposing views on whether pain is essentially perceptual or emotional in nature (see Aydede, Güzeldere, and Nakamura, forthcoming).

Embodied, Embedded, and Situated Cognition

A second general area in which there has been a hive of activity is that of *embedded* cognition, also referred to as *situated* or *embodied* cognition. In part as a reaction to the general character of traditional symbolic AI and connectionism, both of which have abstracted away from the nature of the environment in which cognition actually operates, this cluster of views emphasizes the organism/environment coupling in theorizing about cognition. While the "embeddedness movement" has sometimes represented itself as anticomputational (e.g., Brooks 1997), there has been a concerted effort within an overarching computational framework to capture the spirit of the movement, ranging from Cantwell Smith's (1996) reconceptualization of computation to Dennett's (2000) emphasis on the important role of out-of-the-head scaffolding in higher mental processes. Central to the embedded movement is the idea that cognizers are *agents* who act in the world, gathering information about the world in order to act. This agent-centered conception of cognition has become increasingly a part of mainstream artificial intelligence (e.g., Russell and Norvig 1995). Indeed, as stated at the outset, it is one of the motivating themes of folk psychology. In light of these points, this development within the philosophy of psychology is less a departure from traditional views than a return to one of the themes familiar to those in the field.

There is an obvious affinity between such approaches to cognition and the externalist views that have come to dominate philosophical reflection on intentionality and mental representation. There are a number of attempts (e.g., Clark 1997) to build some firmer bridges between the philosophical and scientific work. But there is much more to be done here.

Robert A. Wilson

References

Aydede, M., G. Güzeldere, and Y. Nakamura (forthcoming), *The Puzzle of Pain: Philosophical and Scientific Essays*. Cambridge, MA: MIT Press.

Block, N. (1995), "On a Confusion About a Function of Consciousness," *Behavioral and Brain Sciences* 18: 227–247.

Brooks, R. (1997), *Cambrian Intelligence*. Cambridge, MA: MIT Press.

Chalmers, D. (1996), *The Conscious Mind*. New York: Oxford University Press.

Churchland, P. S. (2002), *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press.

Clark, A. (1997), *Being There*. Cambridge, MA: MIT Press.

Cotterill, R. (1998), *Enchanted Looms*. New York: Cambridge University Press.

Crick, F., and C. Koch (1990), "Towards a Neurobiological Theory of Consciousness," *Seminars in the Neurosciences* 2: 263–275.

Dennett, D. C. ([1978] 1998), "Why You Can't Make a Computer That Feels Pain," in *Brainstorms: Philosophical Essays in Mind and Psychology*. Montgomery, VT: Bradford Books, 190–232.

——— (2000), "Making Tools for Thinking," in D. Sperber (ed.), *Metarepresentations*. New York: Oxford University Press.

Érdi, P. (2000), "On the 'Dynamic Brain' Metaphor," *Brain and Mind* 1: 119–145.

Fodor, J. A. (1981), *Representations*. Sussex, UK: Harvester Press.

——— (1983), *The Modularity of Mind*. Cambridge, MA: MIT Press.

Graham, G., and L. Stephens (eds.) (1994), *Philosophical Psychopathology*. Cambridge, MA: MIT Press.

Hardin, C. L. (1988), *Color for Philosophers*. Indianapolis: Hackett Publishing.

Horst, S. (1996), *Symbols, Computation, and Intentionality*. Berkeley and Los Angeles: University of California Press.

Jordan, M., and S. Russell (1999), "Computational Intelligence," in R. A. Wilson and F. C. Keil (eds), *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.

Levine, J. (1997), "Recent Work on Consciousness," *American Philosophical Quarterly* 34: 379–404.

Lloyd, D. (2000), "Terra Cognita: From Functional Neuroimaging to the Map of the Mind," *Brain and Mind* 1: 93–116.

Matthen, M. (1999), "The Disunity of Color," *Philosophical Review* 108: 47–84.

Newell, A., and Simon, H. (1981), "Computer Science as an Empiral Inquiry: Symbols and Search," in J. Haugeland (ed), *Mind Design: Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA: MIT Press, 35–66.

Pearl, J. (2000), *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Port, R., and T. van Gelder (eds.) (1995), *Mind as Motion*. Cambridge, MA: MIT Press.

Pylyshyn, Z. (in press), *Seeing and Visualizing: It's Not What You Think: An Essay on Vision and Imagination*. Cambridge, MA: MIT Press.

Russell, S., and P. Norvig (1995), *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall.

Searle, J. (1992), *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Smith, B. C. (1996), *On the Origin of Objects*. Cambridge, MA: MIT Press.

Stich, S. (1983), *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.

Uttal, W. (2001), *The New Phrenology*. Cambridge, MA: MIT Press.

Weiskrantz, L. (1986), *Blindsight*. Oxford: Oxford University Press.

Wilson, R. A. (1995), *Cartesian Psychology and Physical Minds*. New York: Cambridge University Press.

——— (1999), "Philosophy," in R. A. Wilson and F. C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.

——— (2002), "Individualism," in S. Stich and T. A. Warfield (eds.), *Blackwell Guide to Philosophy of Mind*. New York: Blackwell.

See also **Artificial Intelligence; Biology, Philosophy of; Cognitive Science; Connectionism; Consciousness; Evolutionary Psychology; Intentionality; Neurobiology; Reductionism; Social Sciences, Philosophy of**

HILARY PUTNAM

(31 July 1926–)

Hilary Putnam is a philosopher whose unmistakable originality has had a great impact on many areas of contemporary philosophical concern, including, but not restricted to, the philosophy of science, the philosophy of language, and the philosophy of mind. He is also commonly perceived as having changed his mind drastically on a myriad of interrelated issues over the years, notably in areas in which his ideas have been extremely influential, such as functionalism in the philosophy of mind. This perception, which at times Putnam himself appears to share, is, however, very much overstated. The better interpretation of his scholarship and contribution to philosophy emphasizes continuities in his thought. Indeed, careful attention to his writings reveals a remarkable unity of overall concern.

Quine and the later Wittgenstein have been unquestionably important influences on Putnam's thinking over the years. With each he shares, in different ways, a stance that shuns speculative metaphysics as First Philosophy. But Putnam's own position is distinctive. He has always rejected Quine's austere scientism, in particular Quine's tendency to regard philosophy as a branch of natural science—the branch of science that studies how human organisms can have contrived the language of science, the very language within which this investigation is pursued. And in contradistinction to the later Wittgenstein, Putnam has not abstained from offering novel theoretical solutions to philosophical perplexities at the level at which they were traditionally contemplated, concerns as fundamental as how it is possible to think and talk about the world at all. Putnam's lifelong suspicion of speculative metaphysics is evidenced in his earliest writings. In a classic paper on microreduction coauthored with Paul Oppenheim (Openheim and Putnam 1958), Putnam portrays the unity of science—understood as the microreduction of all of science (including social science) to the level of particle physics—as a working hypothesis within science (see Unity and Disunity of Science; Unity of Science Movement; Reductionism). One remarkable aspect of this early paper is its choice of main target: the view that acceptance of the microreducibility of all of science to a single lowest level is a mere act of faith. As against the speculative physicalist metaphysician, who professes belief in all facts being at bottom physical facts, Oppenheim and Putnam defend microreducibility as a hypothesis that is made credible both by empirical evidence and on general methodological grounds. In other words, in Putnam's hands, physicalism, which has become the metaphysics of choice for an entire generation of analytic metaphysicians, becomes a hypothesis on a par with other working hypotheses within science.

Among the main themes running throughout Putnam's corpus are:

- the rejection of an absolute analytic/synthetic distinction and a concomitant rejection of an absolute a priori (for some early statements, see Putnam 1962a, 1962b, and 1968);
- emphasis on the primacy of practice within the philosophical understanding of natural science (for an early statement, see Putnam 1974);
- realism and the attendant accommodation of the possibility of constancy of meaning across radical theory change (see, e.g., Putnam 1975a);
- anti-scientism and the insistence on the indispensability of nonscientific knowledge (see, e.g., Putnam 1978 and 1981);
- anti-Cartesianism in the philosophy of mind and language (see, e.g., Putnam 1988);
- interest-relativity of explanation and what Putnam terms "conceptual relativity" (see, e.g., Putnam 1987); and
- rejection of the fact/value dichotomy (see, e.g., Putnam 1981, Ch. 6; 1990, pt. 2; 2002).

In this article, only some of these themes can be elaborated. The overall aim will be to show that many of them are best viewed as elements in a sustained engagement with, and rejection of, a certain Cartesian outlook on people's cognitive rapport with the world.

It is worth noting in this context that other anti-Cartesian strains abound throughout Putnam's work. A vivid early illustration is provided by his anti-Cartesian rejection of absolute a priori. He argues (Putnam 1968) that even logic itself is empirical. While this may sound familiar to readers of Quine, Putnam's method of argumentation is characteristic of a distinct philosophical style. Unlike Quine (see Quine, Willard Van), whose attack on the analytic/synthetic distinction consists mainly of narrowly circumscribed illustrations of the circularity and futility of various attempts to draw it, Putnam's arguments are rich in historical-scientific precedents. In his view, just as Euclidean geometry is best viewed as falsified by general relativity, classical logic may best be viewed as falsified by quantum mechanics. Specifically, it may be reasonable to conclude that quantum logic, according to which conjunction does not distribute over disjunction, is the "correct" logic on empirical grounds. So much the worse, contends Putnam, for the alleged a priori of logic.

Realism and Reference

A natural point of entry into Putnam's work is through the question of realism (see Realism; Instrumentalism). Putnam has always aligned his position with one form or other of realism, even though some of these versions seem to his critics too remote from what they claim to understand by the term 'realism.' However, one of Putnam's main concerns in this area has been to inquire after what 'realism' might possibly mean. Part of the problem here, of which anyone who has thought about the realism issue is aware, is that it is far from obvious how to formulate realism as a distinctive, and therefore controversial, thesis.

As part of his ongoing engagement with realism, and alongside his insistent realist commitment to the distinction between what is the case and what only seems to be the case (even at the level of the entire species), Putnam has vehemently criticized a version of realism he terms "metaphysical realism." His main efforts in this area have been to find formulations of metaphysical realism that would actually render the position controversial, and then to proceed to controvert it. It is against this backdrop that his most famous argument against metaphysical realism, the model-theoretic argument, should be understood (Putnam 1976 and 1977; see Putnam 1989 for a discussion of the implications of naturalizing semantics). The remainder of this section will be devoted to summarizing the argument, and the next section will examine Putnam's own later reservations about it. This will pave the way for a consideration of Putnam's staunch anti-Cartesianism about content.

The model-theoretic argument is designed to demonstrate the inconsistency of the following set of claims presumed to encapsulate the metaphysical realist position:

The world is a totality of
mind-independent objects. (1)

Sentences are true or false by virtue of a
correspondence between words and
portions of this mind-independent totality.
(2)

A theory that is epistemically ideal might be false.
(3)

The upshot of this position is that truth might very well outrun what is epistemically best. Theories of the world are true or false by virtue of a correspondence between the terms employed and portions of the world, a correspondence to which agents have no direct cognitive access. Thus, even the very best theory that agents come up with might really be false for all they are able to tell.

Putnam's argument for the inconsistency of the set {(1),(2),(3)} is as follows. Let *T* be a formalization within first-order logic of an epistemically ideal theory meeting all theoretical and operational constraints. (The former restraints are requirements such as elegance, simplicity, and explanatory power; the latter restraints are requirements roughly of the form "If *T* implies that agent *A* sees object *O*, then it seems to *A* that *A* is seeing *O*.") Now, if *T* is epistemically ideal, then it is at least consistent. So, given certain minimal requirements on *T* and on the size of the world, *T* has a model *M* of exactly the same size as the world. (Suppose that the world is an infinite totality of things, that the size of *L*, the language of *T*, does not exceed that of the world, and that *T* has infinite models. Then, by one version of the Löwenheim-Skolem theorem, *T* has a model of exactly the same size as the world. If the world happens to be finite, enrich *L* to $L \cup \{\mathrm{c}_j \mid j{\in}J\}$, where the index set *J* is of the same finite size *n* as the world and for every $j{\in}J$, $\mathrm{c}_j{\notin}L$. It is then assumed that the extended theory $T \cup \{\forall j,\ k{\in}J,\ c_{\mathrm{j}} \neq c_{\mathrm{k}}\} \cup \{\forall x(x = c_1 \lor \ldots \lor x = c_n)\}$ is consistent, so that *T* has, once again, a model of the same size as the world. Either way, given minimal requirements on *T* and on the size of the world, *T* will have a model *M* of that size.) But if *T* has a model of the same size as the world, then *T* has a model M^W isomorphic to M that has the world itself as its domain. So, given (1) and (2), T is true of the world, that is, true *simpliciter*. In other words, the sentences of T are made true by a correspondence—the one supplied by M^W—between L and portions of the world. So it turns out that (3) is false: T is bound to be true (under the above minimal conditions). The moral of this argument is not the falsity of (3). After all, (1) and (2) are not supposed to be Putnam's premises but rather the metaphysical realist's. The moral, rather, is the inconsistency of the set {(1),(2),(3)}.

Metaphysical realist replies to the model-theoretic argument usually amplify the original set of theses in some way so as to block the inference from metaphysical realist premises to the negation of (3). For example, David Lewis's (1984) version of this move is to claim that metaphysical realists should insist on supplanting (1) and (2) with something more metaphysically robust along the lines of the following two theses:

> The world is a totality of mind-independent objects *that bear objective similarity relations to one another*. (1′)

> Sentences are true or false by virtue of a correspondence *that respects those objective similarity relations* between words and portions of this mind-independent totality. (2′)

By "objective similarity" in this context is to be understood a similarity among things that is prior to and independent of any explanatory interests investigators might have—a primitive similarity built into the world itself. With this understanding, endorsing (1′) and (2′) indeed allows the metaphysical realist to endorse (3) as well: Given that the world is already carved into objective similarity classes, there is no guarantee that T will have a model isomorphic to this structure. So T might turn out to be false of the world after all. But whether or not metaphysical realism is worth the price of commitment to such objective similarity relations is a question on which Lewis and Putnam strongly disagree.

Reference and Perception

In more recent work (Putnam 1993; 1994a, Lecture I), Putnam identifies the following difficulty in the model-theoretic argument's transition from epistemic ideality, through having a model with the world itself as its domain, to being true *simpliciter*. The transition rests on the tacit assumption that there can be no independent determination of what T's terms refer to other than through the theory's meeting the theoretical and operational constraints. In other words, it is assumed that the correspondence of T's terms to portions of the world is an interpretation in the formal sense, that is, a mapping that figures in a definition of satisfaction, where speakers' perceptions are treated as perceptual conditions ("appearances") to be interpreted as part of the operational constraints. Given the assumption, and given that T has model M^W, T's terms cannot but refer to portions of the world in such a way that T comes out true, as the model-theoretic argument illustrates. And, while the assumption is shared by many of the intended targets of the argument, Putnam himself finds it deeply problematic.

To fully appreciate why requires turning to Putnam's earlier work on the so-called new theory of reference. This will occupy the next section and will reveal an often-neglected continuity in Putnam's thought about meaning and mind—his pronounced anti-Cartesianism. But for now, a preliminary consideration of the implications of the above assumption for the general bearing of speakers' words on their environment is in order. These implications become most vivid by setting aside the subtleties of the model-theoretic argument and focusing on the worn example of the cat on the mat.

Consider a speaker who, in the presence of a certain salient cat on a certain salient mat, reports: "The cat is on the mat." Question: What makes it the case that it is the salient cat being on the salient mat that make the speaker's "The cat is on the mat" come out true? How is it that the speaker's words are about this cat and this mat, rather than about anything else? In a traditional epistemological view famously bolstered by Bertrand Russell (1910; [1912] 1952), the expression 'the cat' contributes a certain perceptual condition ϕ to what is said by 'The cat is on the mat,' the condition of being appeared to as if some specific cat is before one under the relevant conditions. Similarly, the expression 'the mat' contributes a certain perceptual condition ψ to what is said by 'The cat is on the mat,' the condition of being appeared to as if some specific mat is before one under the relevant conditions. So what is said overall by 'The cat is on the mat' is that the ϕ is on the ψ, which Russell analyzes as the following existence claim:

$$\exists x \exists y(\forall z(\phi z \leftrightarrow z = x) \wedge \forall z(\psi z \leftrightarrow z = y) \wedge (\mathrm{ON}xy)).$$

In short, what makes it the case that the speaker's 'the cat' is about the salient cat and the speaker's 'the mat' is about the salient mat is that the cat and mat in question uniquely satisfy conditions ϕ and ψ, respectively.

Suppose all perceptual judgments receive a similar treatment in terms of satisfaction of perceptual conditions. Terms employed for perceived objects are about them by virtue of the objects satisfying associated conditions in the manner illustrated above. Now consider the totality of sentences held true—call it the agent's "overall theory." According to the view of reference espoused by the model-theoretic argument, all terms are about whatever is assigned to them under an interpretation that makes the overall theory come out true. In other words, terms are about whatever satisfies the overall theory, including the perceptual part of the theory. For a term to be about a thing (or things) is for it to pose a certain condition that the thing satisfies (or things satisfy) so as to make the overall theory come out true. This is what reference comes to in this picture—a word-world relation that makes the overall theory come out true.

Even from this rough sketch of the view of reference presupposed by the model-theoretic argument, some lessons can be drawn. For example, if all that is required for a term to refer is that the referent be assigned to the term by an interpretation that makes the overall theory come out true, then reference is bound to be a radically underdetermined affair. Given any interpretation of the language that makes the overall theory come out true, it is a trivial matter to construct a systematic reinterpretation of the language, with compensatory adjustments accommodating all occurrences of the terms in the overall theory, that is truth preserving. To borrow a vivid example from Quine (1992, Ch. 2), any singular term for some spatiotemporally extended item under the old interpretation can be reinterpreted as referring to the item's mereological complement, that is, the whole universe minus the item, with compensatory adjustments to the assignments to all other terms such that truth-values of all whole sentences remain unaltered. One implication of the existence of such possibilities might be that reference is indeed radically underdetermined. This is the conclusion Quine draws (Quine 1969; 1992). For him, there is no fact of the matter as to what a given term refers to independently of some assignment under which the overall theory is true. But Putnam draws a different conclusion. For him, reference is not a matter of interpreting the language of a theory so as to render the theory true. It is not a mapping that figures in the satisfaction of theories. It should be clear that there is a certain traditional view of perception that accompanies the above view of reference-as-mapping (see Putnam 1993). It is the view that the objects of perceptual judgments are specified only through satisfying perceptual conditions that are entertained in the mind, that perceptual judgments are about items in the environment only to the extent that the latter satisfy perceptual conditions with which the agent is "acquainted." (A corollary is that ordinary perceptual beliefs are at bottom general, or *de dicto*, and are never genuinely singular, or *de re.*) This view of perception figures prominently in a general Cartesian picture of the relation between mind and world. A thing is experienced visually, say, to the extent that it "fits" an appearance present before the mind. So the mind comes into contact with the object itself only derivatively, via the mind's immediate contact with an appearance. In this picture, perceivers are confined in their cognitions to how things appear to them and are never in direct contact with the objects about which they cognize. Specifically, their only perceptual contact with the world is through portions of the world satisfying perceptual conditions. Reference and perception are inevitably indirect.

Much of Putnam's work has been devoted to uprooting this picture of the relation between mind and world. But perhaps none has been more influential than his celebrated work on meaning, which is the next topic of discussion.

Meaning and Mind

In considering accounts of meaning, or semantic content, it is useful to draw a distinction between two kinds of theories that are easily conflated (see Kaplan 1989; Stalnaker 2001). The first kind is a semantic theory. Such a theory specifies what the semantic contents of expressions are and their modes of composition. So, for example, the Millian thesis that the semantic content of a name is its bearer is a semantic thesis. The second kind of theory is a *meta-semantic* theory. A meta-semantic theory specifies how the semantic contents of expressions are determined—how, in other words, expressions come to possess the contents they do. So, for example, the historical-chain view of names, according to which the content of a name—its bearer—is determined by an initial act of naming and subsequent inheritance down the generations of users, is a meta-semantic thesis.

From the start, Putnam's work on meaning has been motivated by an effort to explain how meanings can remain fixed despite substantial changes in theory. During the heyday of logical positivism, philosophers of science were attracted to the view that terms gain their meanings from the theories in which they are couched, so that differences in theory implied differences in the meanings of the terms contained therein. This was a certain application of a widely received and traditional view about meaning—that the meaning of terms are general extension-fixing criteria that are shaped by the speakers' body of general beliefs. Thus, for example, it was thought that the meaning of a noun such as 'gold' is given by a set of severally necessary and jointly sufficient general conditions for being gold. And proficiency with the noun was thought to entail an implicit grasp of such an extension-fixing criterion.

Putnam's reaction to this picture, which grounds his overall attack on the positivists' construal of natural science, consists in making two central moves. The first is semantic. As against the tradition that saw meaning as an extension-fixing criterion, Putnam argues that extension is the crucial determinant of the meaning of a typical common noun and that an extension-fixing criterion is no part of the meaning. He concedes that there are some common nouns in the language, such as vixen, mare, and sow, that are associated with extension-fixing criteria in the sense that knowledge of meaning demands grasp of such a criterion. These are the so-called one-criterion nouns, but they are clearly the exception among the nouns. For the vast majority of nouns, no such extension-fixing criterion is forthcoming.

It would surely be an unreasonable cognitive burden on speakers of the language to have to grasp general extension-fixing criteria for nouns such as gold, water, aluminum, and elm, and it seems that no amount of loose talk about implicit grasp of such criteria could loosen this unreasonable requirement on proficiency. The point is vividly illustrated by Putnam in a series of provocative thought experiments involving a distant planet ("Twin Earth") that is indistinguishable from Earth but for a complete absence of H_2O and a complementary abundance of some superficially similar yet alien substance ("XYZ"). Putnam considers a community of English speakers on Earth and its counterpart community of Twin English speakers on Twin Earth before the rise of Daltonian chemistry. The two linguistic communities are stipulated to be exact replicas of one another but for the above difference between Earth and Twin Earth, and it is presumed that the state of knowledge in the two communities is such that it does not support any ability to distinguish H_2O from XYZ. Yet, when English speakers employ 'water,' they refer to portions of H_2O, whereas when Twin English speakers employ 'water,' they refer to portions of XYZ. Whatever general criterion individuals on one planet can be said to associate with 'water,' or to grasp collectively, no such criterion can fix the extension of the term to the exclusion of the substance on the other planet. But there is still a strong inclination to conclude that the extension of 'water' on Earth is H_2O, while the extension of 'water' on Twin Earth is XYZ.

In Putnam's account, the meaning of a typical common noun does not determine the noun's extension. Rather, the meaning of a typical common noun is determined by its extension, whereas an extension-fixing criterion is no part of the overall meaning of the noun. Of course, if the meaning, or semantic content, of the noun is determined by its extension rather than the other way around, and if an extension-fixing criterion is no part of its content, then there is no question of proficiency with the noun being a matter of speakers' grasp of an extension-fixing criterion as its meaning. Semantic content, in this view, turns out not to be in the head.

Putnam's second move against the traditional view of meaning is meta-semantic. If extensions of common nouns are components of their overall contents and are not fixed by grasped criteria, then how are they determined? To make the issue vivid, consider a speaker who has the terms 'elm' and 'beech' in his idiolect and also has some undifferentiating general beliefs about elms and beeches, but who cannot distinguish elms from beeches. (Putnam famously claims himself to be such a speaker.) In the old view, 'elm' and 'beech' for such a speaker would be synonymous and thus coextensive. Yet it seems that by 'elm,' this person can intend to refer to elms, and by 'beech' to refer to beeches. For example, such a person might easily wish to affirm the distinctness of elms from beeches. It would be a distortion of the situation to construe what such a person would be affirming as the contradictory claim that anything is an elm (a beech) just in case it is not an elm (a beech). Perhaps an even more vivid illustration is afforded by the noun 'gold.' Most speakers cannot distinguish gold from iron pyrite, yet, by employing 'gold,' they mean to speak of gold and not of either-gold-or-iron-pyrite. How can such an extension determination be achieved if the meaning of gold does not include an extension-fixing criterion?

Putnam's answer consists of a novel meta-semantic account with two importantly distinct components. The first is the claim that extension determination is achieved indexically. The second is the claim that extension determination is often a social matter. Consider these points in turn. As for the indexical or environmental aspect of extension determination, Putnam's view is that speakers employ a typical common noun N as if they are committed to the following extension-fixing stipulation:

$$N(x) \leftrightarrow \text{same}_{\text{r}}(x, \text{this}),$$

where same_{r} is a relevant similarity relation and the demonstrative 'this' refers to a paradigmatic instance of N. In other words, the stipulation specifies that x is N just in case it is relevantly similar to a paradigmatic instance of N demonstratively referred to. And even though Putnam himself does not fully unpack what it is to speak "as if" one is committed to such an extension-fixing stipulation, his account implies that speakers employ N with the intention to refer to anything bearing same_{r} to paradigm instances.

The second component of Putnam's meta-semantic account is social. For nouns such as 'gold,' 'elm,' and 'beech,' the average speaker cannot determine that something or other is relevantly similar to paradigmatic instances of the relevant kind. Yet 'gold' in the mouth of the average speaker refers determinately to gold. How is this determination achieved? It is achieved through a social cooperation that Putnam dubs 'division of linguistic labor.' In the case of gold, an average speaker intends to refer to anything bearing the relevant similarity relation, $\text{sameness}_{\text{metal}}$, to paradigmatic instances. But whether or not $\text{sameness}_{\text{metal}}$ obtains among samples of substance is not a determination that the average speaker can be presumed capable of making. It is a matter left to metallurgy to decide. Similarly for elm and beech. The extension-fixing stipulations for them would be: $\text{elm}(x) \leftrightarrow \text{same}_{\text{plant}}(x,\text{this})$ and $\text{beech}(x) \leftrightarrow \text{same}_{\text{plant}}(x,\text{this})$. But whether or not $\text{sameness}_{\text{plant}}$ obtains is not something that the average speaker can be presumed capable of deciding. It is a matter left to botany to decide. In other words, for nouns such as 'gold,' 'elm,' and 'beech,' speakers are linguistically deferential to a relevant expertise for fixing their extensions. In this way, there is a special sense in which extension determination for many nouns in language is a social matter. For such nouns, extension, and thus semantic content more generally, is determined both environmentally and socially.

The implications of Putnam's meta-semantic account for the philosophy of science, specifically for the possibility of constancy of meaning across radical theory change, are far-reaching. If extension determines content rather than being determined by it, and is itself fixed in the manner sketched above, then there is little mystery as to how, for example, the ancient Greek term for gold (*chrysos*) and the English term 'gold' can share in meaning. Generally speaking, what is relevant for determinations of sameness of content are the contents themselves and not the ways in which they are determined.

Consider any mass noun N_1 as it is used in linguistic community C_1, and any mass noun N_2 as it is used in a remote linguistic community C_2. Suppose that the two terms are linguistically deferential, that the dominant theory about N_1 in C_1 is T_1 and that the dominant theory about N_2 in C_2 is T_2. Finally, suppose that T_1 informs procedure P_1 for detecting N_1, whereas T_2 informs procedure P_2 for detecting N_2. Then an important consequence of Putnam's meta-semantic account is that N_1 and N_2 can share in meaning as long as P_1 and P_2 overlap sufficiently in their outcomes, that is, as long as there is sufficient overlap in what P_1 identifies as N_1 and what P_2 identifies as N_2.

So regarding *chrysos* and 'gold,' the following explanation emerges. It is highly likely that an average ancient Greek speaker employing *chrysos* back then was linguistically deferential to a certain gold-expertise of that time. And the average English speaker employing 'gold' today is linguistically deferential to a certain contemporary gold-expertise. And it so happens that the two procedures for detecting the presence of the relevant substance are largely (although surely not entirely) consonant. This is the crucial detail in what warrants translating their term *chrysos* into the term 'gold.'

Further Implications

Some broader implications of Putnam's views about meaning are noteworthy. One is a general anti-Cartesian outlook on speakers' cognitive rapport with the environment. If Putnam's semantic and meta-semantic views are on the right track, then portions of the world that speakers think and talk about are not determined via satisfying conditions, perceptual or other, that are entertained in their minds. In Putnam's view, reference, and thus content more generally, is a world-involving affair. It is no longer thought to be a criterion-involving affair internal to the mental lives of the agents, as the Cartesian tradition would have it.

The importance of this point for general epistemology is difficult to exaggerate. It makes agents' cognitive contact with the environment direct, as it were, rather than mediated by criteria entertained in their minds. If one says "Water is wet" and one's twin on Twin Earth utters the same words, then the content of one's 'water' involves H_2O directly, whereas the content of one's twin's 'water' involves XYZ directly. And this is so even if neither one can distinguish the two contents from one another. Generally speaking, an important implication of Putnam's views is that one cannot identify what one says or thinks in abstraction from the global context in which the saying or thinking is conducted, that is, in abstraction from one's relations to others and to the environment. This immediately raises questions about first-person epistemic access to one's own thoughts. If what one says or thinks is shaped by one's relations to the world, then knowledge of these contents depends on knowledge of the world more generally. (For recent work on this topic, see Wright, Smith, and Macdonald 1998.)

Another interesting implication of Putnam's work on meaning regards the prospects of naturalizing reference within some branch or other of cognitive science. If Putnam's views are correct, then it seems hopeless to look to cognitive science for explaining reference. Consider, for example, the phenomenon of division of linguistic labor. One of the striking features of Putnam's meta-semantic story is the hypothesis that linguistic practices are organized within elaborate authority structures—social networks in which novices are deferential to experts by virtue of the latter's possession of expert doctrines in which the former place their trust. If this is correct, then it is highly implausible that reference can be naturalized within cognitive science to the extent that it is highly implausible that such notions as authority can be captured within cognitive science. (For a different take, see Fodor 1994, especially Ch. 2.) If reference implicates an elaborate social structure in the way emphasized by Putnam, then there is good reason for thinking that reference will resist naturalization in the way that is often envisioned by those who think that cognitive science holds the key to the most philosophically fundamental issues concerning the relation between the mind and the world.

Finally, there is an important implication of Putnam's views on meaning for the question of realism. As mentioned at the outset, Putnam has always aligned his views with one form or other of realism, even though many philosophers (including, at times, himself) have viewed his take on the realism question as having undergone dramatic shifts over the years. But one thing that Putnam's work on meaning illustrates is that a realist commitment to the idea that sentences are made true or false by virtue of something external to speakers is already packed into the very notion of semantic content. Consider, again, the term 'gold' and the meta-semantic question of how it gains its content. Putnam's view is that content determination proceeds by way of extension determination, and that extension determination proceeds via referential intentions to pick out anything relevantly similar to paradigmatic instances of the kind in the speakers' environment, where the relevant similarity in this case is left to metallurgy to decide. One thing that emerges from this account is that in order for the term to gain a determinate extension and thus become semantically significant, there has to be an objective standard of relevant similarity. Otherwise, whatever will *seem* to speakers to be relevantly similar to paradigmatic instances will thereby *be* relevantly similar, in which case no determinate extension will be secured for the term after all, and sentences involving 'gold' will not be truth-evaluable. In short, in order to think of words as significant, the relevant seem/is distinction has to be in place. And the latter can be facilitated only by a standard that is external to speakers in the sense of opening up the possibility that what seems to be the case falls short of what is the case. This is clearly a realist commitment in any reasonable understanding of the term 'realism,' and it is one that Putnam has never abandoned. To be sure, it is a far cry from the metaphysical realist picture according to which a theory that is epistemically ideal may nevertheless be false of the world as it is "in itself" because what the terms of the theory refer to is beyond the cognitive access of the theorist. Indeed, Putnam's emphatic rejection of the latter picture is inseparable from his insistence on a person's cognitive being-in-the-world, an insistence he sometimes refers to as 'direct' or 'natural' realism.

Ori Simchen

References

Fodor, Jerry A. (1994), *The Elm and the Expert: Mentalese and Its Semantics*. Cambridge, MA: MIT Press.

Kaplan, David (1989), "Afterthoughts," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford: Oxford UP, 565–614.

Lewis, David (1984), "Putnam's Paradox," *Australasian Journal of Philosophy* 62: 221–236.

Oppenheim, Paul, and Hilary Putnam (1958), "Unity of Science as a Working Hypothesis," *Minnesota Studies in the Philosophy of Science* (Vol. II), 3–36.

Putnam, Hilary (1962a), "It Ain't Necessarily So," in Putnam 1975b, 237–249.

——— (1962b), "The Analytic and the Synthetic," in Putnam 1975c, 33–69.

——— (1968), "The Logic of Quantum Mechanics," in Putnam 1975b, 174–197.

——— (1974), "The 'Corroboration' of Theories," in Putnam 1975c, 250–269.

——— (1975a), "The Meaning of 'Meaning,'" in Putnam 1975c, 215–271.

——— (1975b), *Mathematics, Matter and Method: Philosophical Papers* (Vol. 1). Cambridge: Cambridge UP.

——— (1975c), *Mind, Language and Reality: Philosophical Papers* (Vol. 2), Cambridge: Cambridge UP.

——— (1976), "Realism and Reason," in Putnam 1978, 123–140.

——— (1977), "Models and Reality," in Putnam 1983, 1–25.

——— (1978), *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.

——— (1981), *Reason, Truth and History*. Cambridge: Cambridge UP.

——— (1983), Realism and Reason: Philosophical Papers (Vol. 3), Cambridge: Cambridge UP.

——— (1987), *The Many Faces of Realism*. LaSalle, IL: Open Court.

——— (1988), *Representation and Reality*. Cambridge, MA: MIT Press.

——— (1989), "Model Theory and the 'Factuality' of Semantics," in Putnam 1994b, 351–375.

——— (1990), *Realism with a Human Face*. Cambridge, MA: Harvard UP.

——— (1993), "Realism without Absolutes," in Putnam 1994b, 279–294.

——— (1994a), "Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind," *Journal of Philosophy* 91: 445–517.

——— (1994b), *Words and Life*. Cambridge, MA: Harvard UP.

——— (2002), *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard UP.

Quine, Willard Van (1969), "Ontological Relativity," in *Ontological Relativity and Other Essays*. New York: Columbia UP.

——— (1992), *Pursuit of Truth*. Cambridge, MA: Harvard UP.

Russell, Bertrand (1910), "Knowledge by Acquaintance and Knowledge by Description," *Proceedings of the Aristotelian Society* 11: 108–128.

——— ([1912] 1952), *The Problems of Philosophy*. Oxford: Oxford UP.

Stalnaker, Robert (2001), "On Considering a Possible World as Actual," *Proceedings of the Aristotelian Society* 75(suppl): S141–S156.

Wright, Crispin, Barry Smith, and Cynthia Macdonald (eds.) (1998), *Knowing Our Own Minds*. Oxford: Oxford UP.

See also **Realism: Reductionism**

Q

QUANTUM FIELD THEORY

Quantum field theory (QFT) provides the mathematical framework of modern fundamental elementary particle physics. It grew out of relativistic quantum mechanics conjoined with the demand for a variable particle number (which is required because particles may be created or annihilated during interactions) (see Particle Physics).

Although some of QFT's predictions, notably in quantum electrodynamics (QED), show a highly remarkable agreement with experimental results, its mathematical foundations still suffer from consistency problems. From a philosophical point of view, QFT raises such questions as whether a field or particle ontology applies, how to understand the lack of individuality of field quanta, what the meaning of the vacuum state and virtual particles is, whether and how quantized fields will be measurable and observable, and what quantization in general means.

Field Theory and Canonical Quantization

The need for a QFT in the literal sense of a "quantized field" theory can be seen from the fact that a field is a physical system with infinitely many degrees of freedom, which cannot be captured by ordinary quantum mechanics of finite systems (see Quantum Mechanics). A chain of connected simple harmonic oscillators, for instance, will—in the continuum limit of infinitely many oscillators—allow for infinitely many modes of oscillation. Since fields in physics generally refer to quantities with values associated with space-time points, the continuous character of space-time (as usually presumed in classical space-time physics) is the reason why fields represent systems with infinitely many degrees of freedom. A mathematical expression representing this situation is given by the fact that the field value at a point most generally requires a superposition in terms of a continuous Fourier expansion $\psi(x) = \int d^4k/(2\pi)^4 \tilde{\psi}(k) e^{ikx}$ where $\psi(x)$ is the field amplitude and $\tilde{\psi}(k)$ is its Fourier transform in momentum space, with arbitrarily high four-momenta k representing an infinite number of modes.

With the advent of quantum mechanics in the years between 1925 and 1927, the question of the quantization of the electromagnetic field arose. Following previous work by Born, Heisenberg, and Jordan, Paul Dirac (1927) in his seminal paper first arrived at a full-fledged version of a QFT in terms of the canonical quantization of the free electromagnetic field. The general scheme of canonical quantization is to formulate the theory within the Hamiltonian framework and then to impose commutation relations between the canonical variables. Generally, for a field $\varphi(x)$ with Lagrangian $L(x)$ the canonical variables are $\varphi(x)$ and the canonical momentum $\pi(x) = \partial L(x)/\partial\dot{\varphi}(x)$. These variables are analogous to classical positions and momenta in quantum mechanics, and by making them operator valued, one gets canonical commutation relations (CCRs) analogous to the ordinary Heisenberg relations between $\hat{x}_i$ and $\hat{p}_i$ in quantum mechanics (where the hats ^ denote operators).

In more technical terms (see Ryder 1985) and for the simple case of a scalar field, one starts from the Fourier expansion

$$\hat{\varphi}(\vec{x}, t) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega}} \Big(\hat{a}(\vec{k})e^{-i(\vec{k}\vec{x}-\omega t)} \\ \hat{a}+(\vec{k})e^{i(\vec{k}\vec{x}-\omega t)}\Big).$$

Here, the Fourier coefficients $\hat{a}^{+}(\vec{k})$, $\hat{a}(\vec{k})$ can be considered as creation and annihilation operators of momentum states $|\vec{k}\rangle$ (written in abstract Dirac notation) with the CCRs

$$[\hat{a}(\vec{k}), \hat{a}(\vec{k}\,')] = [\hat{a}^{+}(\vec{k}), \hat{a}^{+}(\vec{k}\,')] = 0,$$
$$[\hat{a}(\vec{k}), \hat{a}^{+}(\vec{k}\,')] = \delta(\vec{k} - \vec{k}\,').$$

The Hamiltonian can be written as $\hat{H} = \int d^3k\omega(\vec{k})\big(\hat{n}(\vec{k}) + 1/2\big)$, where the number of states $\hat{n}(\vec{k}) = \hat{a}^{+}(\vec{k})\hat{a}(\vec{k})$ depends on creation and annihilation operations.

The structure of the CCRs determines the structure of the many-particle Hilbert space, the so-called Fock space, which is defined over a vacuum state $|0\rangle$ with the property $\hat{a}(\vec{k})\,|\,0\rangle = 0$. Applying creation operators to the vacuum produces field quanta states $|\,\vec{k}\rangle = \hat{a}^{+}(\vec{k})\,|\,0\rangle$, which are characterized by the number $\hat{n}(\vec{k})$ only. Therefore, the Fock space allows for an occupation number representation (see below) and can be understood as a direct sum $H = H_1 \oplus H_2 \oplus H_3 \oplus \ldots$ of n-particle Hilbert spaces $H_n = \bigotimes_n H_1$ (where H_1 denotes the one-particle state space).

Two Routes to Quantum Field Theory

The canonical quantization of a classical field theory is not the only route to QFT proper. It can be shown that the quantum mechanics of many particles is indeed mathematically equivalent to the description given above of a QFT in Fock space. Starting from classical particle mechanics *CM*, one may go over to a classical theory with infinitely many particles being equivalent to a classical field theory *CFT*. Applying now the techniques of canonical quantization (basically by imposing CCRs) directly leads to QFT. This route chiefly consists of two procedures: first, the transition from one to (infinitely) many degrees of freedom ($n \rightarrow \infty$) and, second, the introduction of commutation relations (CCRs). Reversing the order of these procedures leads in a first step from *CM* to ordinary quantum mechanics *QM* (of a finite number of particles) by canonical quantization. In a second step, the transition to a theory with infinitely many particles yields a theory equivalent to *QFT*. One may therefore draw the following diagram:

$$\begin{tikzcd}
QM \arrow[r, "n \longrightarrow \infty"] & QFT \\
CM \arrow[u, "CCRs"] \arrow[r, "n \longrightarrow \infty"'] & CFT \arrow[u, "CCRs"']
\end{tikzcd}$$

Puzzles arise once the quantum mechanical wave function is formally treated as a classical field (e.g., the "Schrödinger" or "Klein-Gordon field"). Since this tacitly assumes the equivalence *QM* ⇔ *CFT*, the two routes become confusingly intertwined. In his 1927 paper, Dirac introduced this procedure under the heading "second quantization." The usual stance among physicists on this is best captured in a quote from Ryder (1985, 126):

> What we shall do first is to consider the equation as describing a field $\phi(x)$. Since the equation has no classical analogue, $\phi(x)$ is a strictly quantum field, but nevertheless we shall begin by treating it as a classical field. . . . We shall then take seriously the fact that $\phi(x)$ is a quantum field by recognising that it should be treated as an operator, which is subject to various commutation relations analogous to those in ordinary quantum mechanics. This process is often referred to as 'second quantisation', but I prefer not to use this term. There is, after all, only one quantum theory, not two; what we are doing is quantising a field, rather than the motion of a single particle, as we do in quantum mechanics. It turns out that the field quantisation has an obvious interpretation as a many-particle theory, which is just what we want.

But the conceptual and philosophical problems surely lie deeper (see Cao 1997, Sec. 7.3). One must at least carefully distinguish between the quantization of real-valued fields (such as the electromagnetic field, which allows for a classical interpretation of waves in space-time) and complex-valued fields (such as probability wave functions defined on a configuration rather than position space). For the latter the field operators are non-Hermitean operators and, hence, not observable, whereas for the first the fields—classical and quantized—are observable. In these cases, however, there exist no conserved (particle) currents, and, thus, as Pauli (1933) put it in an early paper, for real fields "the notion of a local particle density [e.g., the photon density] … does not meaningfully exist."

Particles or Fields?

Roughly stated, the two routes to QFT indicate the two primary ontological aspects of quantum fields—the wave or field aspect on the "right-up-route" *CM* → *CFT* → *QFT*, as opposed to the particle aspect on the "up-left-route" *CM* → *QM* → *QFT*. As philosophical debates have shown, both views do contain certain, seemingly ineliminatable problems. Teller (1995), for instance, has argued that quantum fields may be described by space-time–indexed operators but that those operators do not, as classical fields do, represent *specific values* of physical quantities associated with space-time points, but rather represent the quantities ("determinables") themselves—by possibly taking any of a continuum of values. A simple particle interpretation, on the other hand, is also blocked. As a technical result, Malament (1996) has proved a no-go theorem, which excludes the possibility of localizable particles in relativistic quantum theory (see Locality). The other important attack on the particle view stems from the indistinguishability of field quanta.

Indistinguishability

As already mentioned, Fock space allows for an occupation number representation, more formally $|n_1, n_2, n_3 \ldots\rangle = \frac{1}{\sqrt{n_1!n_2!n_3!\ldots}}(\hat{a}_1^+)^{n_1}(\hat{a}_2^+)^{n_2}(\hat{a}_3^+)^{n_3} \ldots |0\rangle$ with $\hat{a}_i^+ \equiv \hat{a}^+(\vec{k}_i)$, where n_i denotes the number of the *i*-th state. The combinatorial factor in the state representation, here written for bosonic fields, indicates that quanta are observationally indistinguishable. (For fermionic fields one has to use totally antisymmetrized representations instead due to the spin-statistics theorem (see Particle Physics). This holds true already on the level of quantum mechanics and becomes a generic feature of QFT.

Ontologically speaking, the indistinguishability means that quanta possess all intrinsic properties in common. Nevertheless, quanta do possess a cardinality. In this sense, then, Leibniz's principle of the identity of indiscernibles fails as a principle of individuation. One way of dealing with this is to consider quanta as nonindividuals lacking even such basal substance-ontological categories as "transcendental individuality," "primitive thisness," or "haecceity." Obviously the indistinguishability of quanta poses a serious problem for a straightforward particle ontology of QFT (cf. French and Redhead 1988; Teller 1995).

Alternative Approaches to Quantum Field Theory

The interpretive issues of QFT become even more complicated due to the fact that there is more than one approach to QFT. Over and above the canonical Hilbert space formulation, mention must be made at least of path integrals and algebraic quantum field theory (AQFT).

Path integrals are based on Feynman's representation of transition amplitudes in nonrelativistic quantum mechanics as sums over path histories weighted by the classical action (Feynman 1948). Due to the principle of least action, the classical trajectory of a particle corresponds to an extremal action functional. The quantum-mechanical interpretation consists of allowing all the possible trajectories "at once." This leads to a transition amplitude $K(x_f, t_f; x_i, t_i) = \int_i^f D[x(t)]e^{iS[x(t)]}$ between an initial state $|x_i, t_i\rangle$ and a final state $|x_f, t_f\rangle$, where $S[x(t)]$ is the action functional (with a suitable measure $D[x(t)]$ in the function space of all trajectories $x(t)$). This formalism can straightforwardly be extended to infinite degrees of freedom—and also provides a convenient method of performing perturbation theory visualized in terms of Feynman diagrams and needed for the calculation of interactions (see below). However, a literal interpretation of Feynman diagrams in terms of spatiotemporal particle trajectories is untenable and does not support a particle ontology.

AQFT, the second approach, was invented by Haag and others in the 1960s (see Haag 1992) as an attempt to construct QFT on a more rigorous and solid mathematical basis. The main idea is that the algebra of observables represents the core physical structure of quantum theory. Given such an algebra A, it is possible to identify the physical states with the linear forms ω over A. It can be shown that each ω defines a Hilbert space H_ω and a representation π_ω of A by linear operators acting on H_ω (Gelfand-Naimark-Segal [GNS] construction).

One considers in particular the C^*-algebra $A(O)$ of all bounded operators associated with a space-time region O; thus, AQFT replaces the correspondence $x \rightarrow \psi(x)$ (where x is the position and $\psi(x)$ is a field amplitude) by the more operationally defined correspondence $O \rightarrow A(O)$. In AQFT, the description of systems with an infinite number of degrees of freedom leads to unitarily inequivalent representations. Different representation classes, so-called *sectors*, must therefore be connected to each other by superselection rules, which include the different, empirically known charges. It is of particular conceptual and philosophical interest to explore the meaning of inequivalent representations.

An important result of AQFT is the Reeh-Schlieder theorem, which, roughly, asserts that any physical state may be created from the vacuum. Due to this theorem and in spite of AQFT's superficial spirit of locality, the notorious quantum nonlocalities (such as EPR-Bell correlations) reappear and can be demonstrated also within the framework of a relativistic QFT (see Redhead 1995) (see Locality).

Interacting Fields

For real applications of QFT, one has to consider interacting fields. Here the field equations become nonlinear (they contain coupling and self-interaction terms with two or more fields) and can be solved only by means of perturbation theory (i.e., as approximations in terms of infinite power series in the coupling constants). Higher-order expansions will then include divergent terms that are due to the integration over internal loops. In order to cope with the infinities, one introduces renormalization procedures, where the various coupling constants and masses are treated first as "bare" infinite parameters that, in a next step, become interaction modified. In QED, for instance, the electric bare charge is screened by vacuum polarization, which means that, for instance, the naked electron charge attracts positrons and repels electrons from the surrounding cloud of virtual electron-positron pairs leading to an effective finite electron charge. It can be shown that the interaction theories in the standard model (see Particle Physics) can be made finite at all orders through the renormalization procedure. From a meta-theoretical point of view, the very idea of renormalization touches the deep question, whether and on what grounds one should expect a unified and rigorous fundamental theory.

Moreover, the fundamental standard model of today distinguishes between quantized matter and interaction fields and describes their interconnection in terms of gauge theories. The latter are characterized by a local symmetry requirement; more precisely, the postulate of local gauge covariance of the free matter field theory leads to the introduction of a covariant derivative where the inhomogeneous connection term can be interpreted as a gauge potential field (cf. Auyang 1995). However, since gauge transformations refer to unphysical degrees of freedom, only gauge invariant quantities can be observable. This raises questions about the status of the seemingly redundant excess or "surplus structure" in gauge theories. Apparently, modern physics QFTs show in many ways the unresolved puzzles of the notorious intertwining between mathematics and physics.

Good starting points for further studies can be found in Huggett (2000) and Redhead (1982), as well as in the collection of papers in Brown and Harré (1988), Cao (1999), Clifton (1996), and Kuhlmann, Lyre, and Wayne (2001).

HOLGER LYRE

References

Auyang, Sunny Y. (1995), *How Is Quantum Field Theory Possible?* New York: Oxford University Press.

Brown, Harvey R., and Rom Harré (eds.) (1988), *Philosophical Foundations of Quantum Field Theory*. Oxford: Clarendon Press.

Cao, Tian Yu (1997), *Conceptual Developments of 20th Century Field Theories*. Cambridge: Cambridge University Press.

——— (ed.) (1999), *Conceptual Foundations of Quantum Field Theory*. Cambridge: Cambridge University Press.

Clifton, Robert (ed.) (1996), *Perspectives on Quantum Reality*. Dordrecht, Holland: Kluwer.

Dirac, Paul A. M. (1927), "The Quantum Theory of Emission and Absorption of Radiation," *Proceedings of the Royal Society of London* A 114: 243–256.

Feynman, Richard P. (1948), "Space-time Approach to Non-Relativistic Mechanics," *Reviews of Modern Physics* 20: 367–385.

French, Steven, and Michael Redhead (1988), "Quantum Physics and the Identity of Indiscernibles," *British Journal for the Philosophy of Science* 39: 233–246.

Haag, Rudolf (1992), *Local Quantum Physics*. Berlin: Springer.

Huggett, Nick (2000), "Philosophical Foundations of Quantum Field Theory," *British Journal for the Philosophy of Science* 51: 617–637.

Kuhlmann, Meinard, Holger Lyre, and Andrew Wayne (eds.) (2002), *Ontological Aspects of Quantum Field Theory*. Singapore: World Scientific.

Malament, David (1996), "In Defense of Dogma: Why There Cannot Be a Relativistic Quantum Mechanics of (Localizable) Particles," in R. Clifton (ed.), *Perspectives on Quantum Reality*. Dordrecht, Holland: Kluwer.

Pauli, Wolfgang (1933), "Einige die Quantenmechanik betreffende Erkundigungsfragen,“ *Zeitschrift für Physik* 80: 573–586.

Redhead, Michael (1982), "Quantum Field Theory for Philosophers," in P. D. Asquith and T. Nickles (eds.), *PSA 1982*. East Lansing, MI: Philosophy of Science Association, 57–99.

——— (1995), "More Ado about Nothing," *Foundations of Physics* 25: 123–137.

Ryder, Lewis H. (1985), *Quantum Field Theory*. Cambridge: Cambridge University Press.

Teller, Paul (1995), *An Interpretive Introduction to Quantum Field Theory*. Princeton, NJ: Princeton University Press.

See also **Quantum Mechanics; Particle Physics; Locality**

QUANTUM LOGIC

Much work done under the rubric 'quantum logic' concerns various mathematical structures connected with quantum mechanics (for reviews, see Beltrametti and Casinelli 1981; Pták and Pulmannová 1991). It includes important elucidations of the physical content of the formalism; in particular, "to what extent is the Hilbert space description of quantum systems coded into the order structure of propositions?" (Beltrametti and Casinelli 1981, 89). But this entry is concerned with a different issue: the possibility—or necessity—of using nonstandard logical systems in the interpretation of quantum-mechanical formalism. In keeping with most literature on the topic, discussion is confined, on the physical side, to nonrelativistic quantum mechanics; and, on the logical side, to first-order propositional calculus and first-order logic—also called first-order functional or predicate calculus (see, e.g., Stoll [1963] 1979). Certain needed topics in classic mechanics (see "Classical Ensembles and Probability Densities") and quantum mechanics (see "Probability Amplitudes and Feynman Paths") not treated elsewhere in this volume are also discussed.

In assessing various 'quantum logics,' a distinction is made between "*the* logic *of*" approaches, which claim that some unique nonstandard logic is *necessary* for the correct interpretation of quantum mechanics, and 'logic(s) *for*' approaches, which discuss the possible utility of one or more alternate logic(s) in the interpretation of the formalism (cf. Gibbins [1987, 142] distinction between *activist* and *quietist* interpretations of quantum lattice logic).

History

Von Neumann ([1932] 1955, 247–254) and Birkhoff and von Neumann (1936) claimed that a nonstandard propositional calculus based on the non-Boolean lattice structure of the closed linear subspaces of a quantum-mechanical Hilbert space (see Quantum Mechanics) should replace the classical propositional calculus based on the Boolean lattice of subspaces of a classical-mechanical phase space. Strauss ([1936] 1972) used a propositional calculus based on the partial Boolean algebra of projection operators acting on Hilbert space to formalize Bohr's concept of complementarity (see Complementarity). Both of these approaches preserved the traditional bivalent valuation of propositions as either true or false, but Février (1937) (later Destouches-Février) and Reichenbach (1944) introduced indeterminacy as a third truth value; and von Weiszsäcker (1958) advocated a many-valued "complementarity logic," in which a probability rather than a definite truth value is associated with each proposition. Recent discussions of quantum logic devote little attention to nonbivalent logics (but see Haack 1996 on Reichenbach's three-valued logic); nor has there been extensive discussion of operational or dialogic approaches to quantum logic or of a probabilistic variant of intuitionist logic for the consistent-histories interpretation of quantum mechanics, both of which are briefly discussed at the end of this entry. The rest of this entry is devoted to the discussion of quantum logics based on either non-Boolean lattices or partial Boolean algebras (see Jammer 1974, 340–416, for a history of quantum logic; Hooker 1975 for a selection of the relevant papers).

Classical Mechanics and Classical Logic

Phase space is the space of all possible states of a classical mechanical system. This state is determined by particular values of the system's generalized position q and momentum p (hereafter, in a system with n degrees of freedom, a single symbol such as q or p will stand for the entire set of n components of the corresponding quantity). Physically, phase space coordinates may represent energies, angles, angular momenta, etc. (see Classical Mechanics). The Boolean lattice of all subsets (or just measurable subsets if desired) of its phase space may be put in correspondence with the standard ("classical") logic of a corresponding set of predicates of, or propositions about, the system, which form an isomorphic Boolean lattice. That is, each such subset can be interpreted semantically as corresponding to either a predicate of the system or a proposition about the system—or even to a "yes-no" question about the system in operationalist interpretations of this logic. For brevity, 'proposition' will be used hereafter to stand for any of these possible interpretations. A singleton set, that is, one containing a single point (q, p) of phase space, corresponds to an "elementary proposition," representing the maximal possible specification of the properties of the system at a given time: "The system has position q and momentum p."

Throughout this entry, all propositions are interpreted extensionally. For example, if a set of canonical variables q and p are related to a new set Q and P by a canonical transformation, then the propositions 'The system has values Q, P' and 'The system has values q, p' are equivalent. With this identification of equivalent propositions, the Boolean algebra associated with the lattice of subspaces discussed above becomes a Lindenbaum algebra (see, e.g., Stoll [1963] 1979, chap. 6).

The point in phase space occupied by the system at any time t is often referred to as 'the state of the system at t.' A classical system that is closed in the interval between the times t_1 and t_2 is assumed to be deterministic during that interval: given its state at some intermediate time t, the equations of motion of the system (usually, but not necessarily, assumed to be derivable from a Hamiltonian) may be used to *predict* its future state up to t_2, or *retrodict* its past state back to t_1. Consequently, each *elementary* proposition, 'The system has position q and momentum p' is either true or false (with exclusive 'or') at any time t; and for each t, the proposition is true for one and only one pair of values, let us say q_0 and p_0.

There is an isomorphism between the Boolean algebra of subspaces of phase space and the Boolean algebra of elementary and compound propositions about the system. Any *compound* proposition about the system is equivalent to one of the form 'The system's position and momentum lie within some subset S of the phase space' and also has a determinate truth value for each t: it is true if the point (q_0, p_0) lies in the subset S at that time, false if it does not. Such compound propositions are generated by logical operations on elementary or other compound propositions. (Note that propositions like 'The system has position p,' which might appear simple, are actually *compound*.) The logical operations correspond to set-theoretical operations on the subsets corresponding to the propositions, the basic correspondences being given in the following table:

Logical operation	Set-theoretical operation
Negation ('not,' '$\neg$')	Complementation
Conjunction ('and,' '$\wedge$')	Intersection ('$\cap$')
Disjunction ('or,' '$\vee$')	Union ('$\cup$')

All other logical operations may be defined in terms of these. In particular, the (material) conditional ('if ... then') '$a \rightarrow b$' is defined as '$\neg a \vee b$'; and the biconditional ('if and only if') '$a \leftrightarrow b$' is defined as '$(a \rightarrow b) \wedge (b \rightarrow a)$.' All of these operations are *truth-functional;* that is, the truth or falsity of any compound proposition is uniquely determined by a valuation, that is, the assignment of bivalent (true or false) truth values to each of the propositions entering into the compound.

Various other logical concepts can be defined in terms of such sets or relations between them. For example:

Logical concept	Set or relation between sets
Tautology ('T')	The universal set ('I')
Contradiction ('F')	The empty set ('□')
Implication $\Rightarrow$	Inclusion ('$\subseteq$')

This means that a proposition corresponding to the universal set—the entire phase space in our case—is a tautology; a proposition corresponding to the empty set is a contradiction; and if the set corresponding to a proposition is included in the set corresponding to a second one, then the first proposition implies the second.

Since the classical propositional calculus is truth-functional, the semantic interpretation of all classical logical operations and concepts may be defined in terms of the above set-theoretical operations and relations (see 'Semantic Problems' below). Nevertheless, it has recently been shown (Pavičić and Megill 1999) that there exist complete, non-Boolean lattice models of the axioms of the classical propositional calculus.

Classical Ensembles and Probability Densities

When introduced into classical mechanics, probabilities are always symptomatic of ignorance, i.e., renunciation of full information about a classical system that, in principle, could be obtained. Classical ensembles in configuration space (Schiller 1962; Berry and Mount 1972) provide an example that is important for comparison with the situation in quantum mechanics (see "Probability Amplitudes and Feynman Paths" below). Consider a complete solution $W(q, \alpha, t)$ to the Hamilton-Jacobi equation for a system with Hamiltonian H. The solution is based on stipulation of "half" the total number of variables needed to specify an individual trajectory in phase space, i.e., each pair (q, α) specifies such a trajectory. The complete solution corresponds to a real or virtual ensemble of such trajectories, one for each value of α. The density $\rho(q, \alpha, t)$ of these trajectories in the extended configuration space with coordinates (q, t) is given by the Van Vleck determinant $|\partial^2 W/\partial q\, \partial\alpha|$; and, hence, if a trajectory is chosen at random, the probability $P(q, \alpha, t)$ of its having the values (q, α) at time t is proportional to $\rho(q, \alpha, t)$.

But, whether one *knows* their values or not, at any time each individual system in a classical ensemble always *possesses* some *definite* position and momentum, and it is only renunciation of a possible measurement of half this information in preparing the ensemble that leads to the probabilities of values for individual systems in the ensemble. In other words, a Hamilton-Jacobi ensemble can always be subdivided into subensembles, for which more narrowly defined probabilities can be calculated; and it is ultimately divisible into individual trajectories, with each of which definite values are associated.

For later comparison with quantum ensembles, consider a complete solution of the form $W(q_i, t_i; q, t)$—a Hamilton's principal function—and let t be later than t_i. It follows from the previous discussion that the conditional probability for a system initially prepared with position q_i at time t_i to be found in an interval dq around q at time t is given by:

$$P(q_i, t_i; q, t)dq,$$

where the probability density $P(q_i, t_i; q, t)$ is proportional to the Van Vleck determinant of Hamilton's principal function. This probability may be given a propensity interpretation for a virtual ensemble associated with a single system, or a frequency interpretation for an ensemble of identically prepared systems (see Probability).

As noted above, the entire ensemble of trajectories associated with a principal function may be decomposed into subensembles of trajectories, with each of which a probability density is associated. A collection of such subensembles is *complete* if

- no trajectory belongs to more than one subensemble; and
- every trajectory belongs to some member of the collection.

It follows that the robability density for the entire ensemble is the sum of the densities for all the members of the complete collection.

It also follows from the properties of the principal function and its Van Vleck determinant that probabilities are cumulative. That is, for $t_i \leq t \leq t_f$:

$$P(q_i, t_i; q_f, t_f) = P(q_i, t_i; q, t)P(q, t; q_f, t_f).$$

The main difference between classical and quantum-mechanical conditional probabilities is that probability densities must be replaced by probability amplitudes in the assertions of the last two paragraphs (see "Probability Amplitudes and Feynman Paths"; but see "Consistent-Histories Approach" below for an attempt to associate probabilities with quantum histories).

First-Order Functional Calculus

Standard first-order functional calculus deals with statements involving one or more propositional or predicate functions and the universal and existential quantifiers. A typical function is symbolized by $F(x)$, where x ranges over some class C of individuals. As x ranges over its values, the truth values of $F(x)$ range over true and false. Propositional functions are compounded in the same ways as propositions. The universal quantifier $(\forall x)[F(x)]$ asserts that $F(x)$ is true for all values of x in C. The existential quantifier $(\exists x)[G(x)]$ asserts that there is at least one value of x in C for which $G(x)$ is true.

In the case of classical mechanics, the class C is the Boolean (or rather Lindenbaum) algebra of subspaces of the phase space of the system being considered. Then the *universal* quantifier is equivalent to the assertion of the *conjunction* of all propositions $F(x)$ as x ranges over the points (or better, the singleton subsets) of phase space; while the *existential* quantifier is equivalent to the assertion of the *disjunction* of all these propositions.

A semantics for the standard propositional and first-order functional calculi is easily developed such that given the antecedent of any conditional propositional tautology, the consequent can be proven using standard rules of inference (basically *modus ponens*).

What Is a Quantum-Mechanical Proposition?

Much of the debate over quantum logic(s) hinges (implicitly if not explicitly) on the question of the quantum-mechanical analogue of the classical-mechanical propositions discussed above. Classically, propositions refer to only the properties of the system itself, considered as closed (that is, not interacting with anything outside the system). But, as Bohr emphasized, the existence of the quantum of action h prevents such a complete separation between a quantum-mechanical system and its macroscopic surroundings. Two major consequences are:

1. A full description of a quantum-mechanical *phenomenon* (Bohr 1958) or *process* (Feynman 1968; the word "process" will be used hereafter) must include a specification of the result of an initial preparation of the system, an account of the type of interactions it undergoes subsequently, and the result of some act of *registration* ("measurement") to which the system is finally subjected (see Quantum Measurement Problem);
2. A maximal quantum-mechanical preparation or registration specifies only half the data about a system that would be specifiable classically. For example, while one could in principle prepare or register a classical-mechanical system with both a determinate position *and* momentum, one can prepare or register a quantum-mechanical system only with *either* a determinate position *or* momentum. (Such quantum-mechanical quantities are often referred to, in a somewhat misleading fashion, as *observables.*)

As a consequence, a typical proposition about a process involving an electron might read: "At time t_1 the electron was prepared with momentum p_0, subsequently passed through a certain electric field E, and at (a later) time t_2 was registered at position q_0." Quantum mechanics assigns a probability to such a proposition as explained in the next section.

Probability Amplitudes and Feynman Paths

Because of consequence (2) above, rather than being analogous to preparation of an individual classical system, a quantum-mechanical preparation is analogous to the preparation of a classical ensemble (see "Classical Ensembles and Probability Densities" above). Given such an ensemble, only the probability for a definite value of that half of the final data chosen for final registration (measurement) can be calculated. In quantum mechanics too, only the *probability* of a quantum-mechanical process leading from an initial preparation to a final measurement can be defined—in limiting cases, this probability may be 1 (certainty) or 0 (impossibility). The central difference in quantum mechanics is that rather than a probability (or probability density in the continuous case), as in the classical system, one computes a *probability amplitude*—a complex number of amplitude ≤ 1—for each process. This amplitude must then be "squared" to get the corresponding probability. *Partial amplitudes* are calculated for all paths leading from the initial preparation to the final measurement that are indistinguishable within the given process, and (rather than partial probabilities as in the classical case) these partial amplitudes must be summed to get the total amplitude for the process. Again, it is the probability amplitudes that are cumulative, rather than the probabilities.

Using Dirac's bra-ket notation, one may write $<b|a>$ as the total amplitude for some process connecting an initially prepared value a and a finally registered reading b (note that a and b may be the values of *different* observables). The probability of this process is then equal to the square of the absolute value of the amplitude:.

$$P(a \rightarrow b) = |< b\,|\,a>|^2.$$

The two rules mentioned above are summarized in the following formula. Let $|b_i>$ be a complete collection of indistinguishable alternatives at some time between preparation $|a>$ and registration $|c>$. This means that the paths $<c\,|\,b_i>\ <b_i\,|\,a>$ are all mutually exclusive and between them exhaust the possibilities (cf. the definition of "complete" for classical ensembles). Then:

$$< c \mid a > = \sum_i < c \mid b_i > < b_i \mid a > .$$

If a registration of one of the alternatives were to take place at some intermediate point, thus distinguishing between these alternatives, then the probabilities $P(a \rightarrow b_i\ ) = |<b_i|\,a>|^2$ would have to be computed and used in all calculations of later probabilities.

Note that, as in the classical case, all the quantum-mechanical probabilities are *conditional*; one does not ask for the probability of a final b-value *tout court* but given an initial prepared value $a.$ Once computed, these conditional probabilities obey the laws of the classical probability calculus based on classical logic. In contrast to the classical case, quantum-mechanical conditional probabilities cannot be attributed to ignorance but are fundamental. Without altering the physical conditions defining the process under consideration (thus producing a different process), it is impossible to further subdivide an initially prepared quantum-mechanical ensemble into subensembles, let alone into individual trajectories, for which probabilities (as opposed to probability amplitudes) can be defined. For example, a sample of initially undecayed radioactive nuclei all with the same average half-life cannot be subdivided into subensembles each with different predicted average lifetimes. (Of course, retroactively, *afte*r some or all have decayed, it is easy to do so.)

To complete the analogy to classical ensembles, note that the probability amplitude $<q_i, t_i \mid q_f, t_f>$, like any complex number of amplitude ≤ 1, can be written:

$$<q_i, t_i \mid q_f, t_f> = [P(q_i, t_i; q, t)]^{1/2} \exp[h/iW(q_i, t_i; q, t)],$$

where P is a real number between 0 and 1. To a first, semiclassical approximation in the classically allowed region of configuration space (see Berry and Mount 1972 for the semiclassical treatment of turning points and the classically forbidden region), W is a Hamilton principal function, and P is its Van Vleck determinant. In other words, a classical ensemble can be used to produce a semiclassical wave function for a system described by some Hamiltonian.

Prediction, Retrodiction, and State Functions

Preparation of a system usually involves a choice between one of several initial input channels a_i (e.g., the choice of that beam of atoms with given z-component of spin in a Stern-Gerlach type of experiment); and registration usually involves one of several possible outcome channels b_j (e.g., the result of a subsequent measurement of the x-component of the spin of the atoms in the chosen beam); so that one may calculate a set of amplitudes $<a_i \mid b_j>$ for these related processes. An ensemble of (real or virtual) copies of the system may then be prepared by *preselection,* based on the choice of one of an a_i. In this case the probability amplitudes will be used for *prediction* of the probabilities of each of the subsequent possible *registration* results b_j. As an aid in calculating these probability amplitudes for prediction, a *state* of the system $|c>$ is often defined for the time interval between preparation and registration. (In the Schrödinger picture, the state changes with time; in the Heisenberg picture it does not.) The possible states of a quantum-mechanical system are assumed to form a Hilbert space, and most standard accounts of quantum mechanics begin with consideration of the abstract state space or one of its representations (see Quantum Mechanics).

Usually, only preselected ensembles are considered, but an ensemble may also be *postselected* based on the choice of a unique final registration result b_j, and amplitudes introduced as aids in calculating the conditional probabilities for *retrodiction* of the various possible values of initial a_i, given a particular final value of b_j. The limited physical significance of the concept of state in quantum mechanics is indicated by the fact that at any intermediate time between preparation and selection, *different* states will be assigned to preselected and postselected ensembles for the same physical system. In contrast, the same state of a classical ensemble can be used for purposes of either prediction or retrodiction.

Classical or Quantum Logic?

As noted in the section "What Is a Quantum-Mechanical Proposition?," classical propositional logic is all that is needed to handle propositions describing a complete quantum-mechanical process. Logical problems begin when such propositions are truncated by omission of reference to preparation and/or registration. Since classical-mechanical propositions are correlated with states of the system in phase space without reference to anything outside the system, an attempt is made to correlate quantum mechanical propositions with states of the system in Hilbert space (see below). Usually, only preselected ensembles are considered, and explicit reference to both reparation and registration is dropped. Attention is focused on "the state $|c>$ of the system at time t," to which one tries to attach a significance similar to that of the state in the classical case.

The maximal goal of such an approach is to attribute a *complete* set of classical properties, e.g., *both* position *and* momentum, to a quantum system in a given state. Note that, in contrast to the classical case, here propositions, such as "The system has position q" and "The system has momentum p," are *elementary.* They are then compounded by suitably defined operations of conjunction and disjunction in an attempt to give meaning to the resulting compound propositions, even when the properties referred to in such compound propositions are incompatible quantum mechanically. The compound propositions are assumed to have a definite truth value at any time *t*, and the inability to actually determine these truth values by experiments performed at this time is regarded as a purely epistemic problem. An attempt to carry out such a program while maintaining classical propositional logic would clearly fail. So some alternate, nonstandard logic is introduced, with the aid of which the program can be carried out. The earliest and most widespread such logic is based on the (non-Boolean) *orthomodular lattice of linear subspaces* of the Hilbert space associated with the system, with the claim that this results in *the* logic of quantum mechanics, that is, of the reality underlying the formalism (see "Orthomodular Lattice Logic" below).

More modestly, it may be agreed that one must restrict simultaneous attribution of properties to a system, and hence the attribution of simultaneous truth values to the corresponding sets of elementary propositions, to quantum-mechanically *compatible* sets of properties. Compounding of such properties and/or propositions is similarly restricted to compatible propositions. There are a number of ways of developing such an approach into a formal logical system, with the claim that use of (one or more of) the resulting nonstandard logic(s) aids the understanding of the quantum-mechanical formalism. The earliest and most widespread of such logic(s) *for* quantum mechanics is based on the *partial Boolean algebra of projection operators* acting on the Hilbert space associated with the system (see "Partial Boolean Algebra Logic" below).

Orthomodular Lattice Logic

As traditionally formulated since von Neumann (1927), the state space of a quantum-mechanical system is a Hilbert space (see Quantum Mechanics). In the lattice-logical approach, elementary quantum propositions are taken to correspond to one-dimensional linear subspaces (rays) of the Hilbert space, compound propositions corresponding to higher-dimensional closed linear subspaces. No reference to the circumstances of preparation or registration of the system is supposed to be included in the semantical interpretation of these propositions, which are assumed to refer exclusively to the state of the system at some time *t*. But, as noted earlier, an elementary proposition about a quantum system can involve only half the canonical variables of the corresponding classical system. For example, the statements 'The system has position *q*' and 'The system has momentum *p*' are each (mutually exclusive) elementary quantum propositions, and no further information about the system may be included in such elementary propositions. As a consequence (and in contrast to the classical case), the truth of an elementary quantum proposition is *not* sufficient to determine the truth value of all other (elementary or compound) propositions. If an elementary quantum proposition is true at some time, for example, it does not follow that all other elementary quantum propositions are false at that time.

To give them a determinate truth value, (most) such propositions would have to be rephrased probabilistically—for example, 'The probability that the system has a position *q* is *x*,' where *x* is a real number between 0 and 1—or more accurately, as conditional probability statements, such as 'If the system has momentum *p*, then the probability that the system has a position *q* is *x*,' and this would actually describe a process (preparation of *p*, registration of *x*). In contrast to the classical situation, the conditional probabilities occurring in such propositions cannot be eliminated by any additional knowledge about the system.

How to interpret such quantum-mechanical probabilities has been a subject of controversy since the earliest days of the theory. In particular, how is "probability" in such a proposition to be interpreted semantically? It is certainly not absolute, but conditional on the truth of some other given (usually elementary) quantum proposition (this circumstance is usually expressed in terms of a certain "preparation" of the system, or in equivalent "state vector" language). Even assuming their conditional character, are such probabilistic assertions meaningful for a single system, as propensities? Or do they hold only for an ensemble of systems, for each of which the original antecedent proposition is true ("identically prepared systems" or "systems having the same state vector"), as in ensemble interpretations of quantum mechanics?

How to treat the logic of such quantum propositions became a subject of considerable controversy with the proposal to "just read the logic off from [the atomic orthomodular lattice of closed linear subspaces of] the Hilbert space" (Putnam 1969). The atoms of this lattice are the one-dimensional subspaces of the Hilbert space, and they correspond to the elementary propositions about the quantum system discussed in the previous section. Logical operations on (elementary or compound) propositions are then defined in terms of operations on the lattice subspaces corresponding to the propositions, as indicated in the following table:

Logical operation	Closed linear subspace operation
Negation ('not,' '¬')	Orthogonal subspace ('⊥')
Conjunction ('and,' '∧')	Intersection ('∩')
Disjunction ('or,' '∨')	Span of the two linear subspaces ('Sp')

Note that the definition of "conjunction" does not differ from the classical one in the sense that applied to closed linear subspaces, set-theoretical intersection does not take us outside that class. On the other hand, the definitions of "disjunction" and "negation" cannot employ set-theoretical complementation and union: When applied to linear subspaces, each takes us outside this class. The chosen correspondence of "disjunction" with "span" (i.e., the smallest closed linear subspace containing the sums of all vectors lying in either of the two spanned linear subspaces) and of "negation" with "orthocomplement" are mutually coherent, in the sense that they enable the preservation of the form (if not its semantical content, as shall be seen below) of the *tertium non datur,* the law of the excluded middle: 'A or not-A' is a tautology, or equivalently its negation 'A and not-A' is a contradiction (see below).

As a consequence of the non–set-theoretical definitions of negation and disjunction, the resulting operations are not truth-functional (Gibbins 1987, 135–136). Indeed, no consistent assignment of truth values to all three connectives is possible that preserves the relation between tautologies and inferences (Malament 2002, 16–18) (for the semantic implications, see "Semantic Problems" below). Recently, it has also been shown (see Pavičić and Megill 1999) that there are nonorthomodular lattice models that obey all the axioms of this quantum logic.

Various other logical concepts can now be defined by closed linear subspaces or relations between them. For example:

Logical concept	Closed linear subspace or relation between them
Tautology ('T')	The entire Hilbert space ('H')
Contradiction ('F')	The zero-dim linear subspace ('{0}')
Implication ('⇒')	Inclusion ('⊆')

Notice that these definitions do not differ from the classical ones in the sense that, as sets, the entire Hilbert space and the zero-dimensional linear subspace are both linear subspaces, and as applied to closed linear subspaces, set-theoretical inclusion is still meaningful. In the resulting lattice logic, both distributive laws relating conjunction and disjunction fail. Sometimes this has been taken to be the distinctive feature of quantum logic; but it is easily seen that lattice logics can be constructed for certain classical physical systems, in which the distributive law also fails (see Stachel 1986, 236–264).

It has been shown (see Megill 2005, where it is misnamed "implication") that there are five possible definitions of a lattice quantum-logical conditional as the closest possible approximation to the classical (Boolean) material conditional. In a certain sense, it does not matter which one is chosen: Any of these conditionals, together with negation, may be chosen as the only primitive operations in setting up lattice quantum logic. But no matter which is chosen, it does not have the desired relation to implication (see "Semantic Problems" below).

Partial Boolean Algebra Logic

As mentioned above, Strauss ([1936] 1972) introduced partial Boolean algebra logics in an attempt to formalize Bohr's concept of complementarity. The form of the elementary propositions is the same as in the lattice logic approach, but such propositions are now put into correspondence not with one-dimensional subspaces of the Hilbert space, but with the one-dimensional projection operators onto these spaces, with which they are in one-one correspondence. (A Hermitean operator P is a projection operator if and only if (*iff*) it is idempotent, i.e., if $P^2 = \mathrm{P}$.) The set of all these operators forms a partial Boolean algebra (defined below); and the logic is now read off from the structure of this algebra. Only those logical operations on propositions are allowed that correspond to operations on the corresponding projection operators that do not lead out of this algebra. While the sum of two projection operators P_1, P_2 (taken in either order) is always a projection operator and their addition is associative, their product is a projection operator *iff* they commute; and they commute only if

- The corresponding subspaces are orthogonal, in which case $P_1 \cdot P_2 = P_0$, where P_0 is the operator that projects onto the zero-dimensional linear subspace $\{0\}$; or
- One subspace is contained within the other: ($P_1 \subseteq P_2$), in which case $\mathrm{P}_1 \cdot \mathrm{P}_2 = P_1$.

Conjunctions and disjunctions of commuting projection operators P_1 and P_2 are defined as follows:

$$P_1 \wedge P_2 = P_1 \cdot P_2 \quad P_1 \vee P_2 = P_1 + P_2 - P_1 \cdot P_2.$$

(Conjunctions and disjunctions of noncommuting projection operators are excluded; they could be defined only "unnaturally," by first going to the corresponding linear subspaces, using the lattice definitions for these subspaces [see previous section], and then going back to the corresponding projection operators.) The negation of a projection operator $\neg P$ is defined by the complementary projection operator $I - P$, where I is the identity operator that projects onto the entire Hilbert space.

With these definitions, the set of projection operators forms a partial Boolean algebra (*pBa*). A *pBa* is a family of Boolean algebras B_i that obeys the following conditions:

1. If B_i, B_j are two members of the family, their set-theoretical intersection $B_i \cap B_j$ is also a member of the family. This implies that there are unique common zero (P_0) and unit (I) elements that belong to each of the B_i values.
2. Given three elements, such that any pair of them belong to some Boolean algebra (possibly different ones for different pairs) in the family, then there is a Boolean algebra in the family such that all three elements belong to it.

In the resulting *pBa* quantum logic, commuting operators correspond to compatible propositions, and each set of all mutually compatible propositions forms one of the Boolean algebras in the family. Since conjunction and disjunction are defined only for compatible propositions, one cannot make simultaneous assertions about incompatible propositions. Thus a sentence like "The particle has position x and momentum p" is not a well-formed proposition, so it cannot even be called false!

Partial Boolean algebras are closely related to orthoalgebras (see Wilce 2002), which play an important role in logics based on the consistent histories approach to quantum mechanics (see below).

Semantic Problems

In the previous two sections, syntactical characterizations of two quantum logics have been given, with some hints about the semantic interpretation of their propositions. But for them to be regarded as true logics, a full semantic interpretation must be given. The results of Gleason (Hooker 1975, 123–133; Dvurečenskij 1993), Specker (Hooker 1975; 135–140), and Kochen and Specker (Hooker 1975, 263–292) present formidable obstacles to a monotonic semantic interpretation of any bivalent lattice quantum logic. In the standard propositional calculus, a truth function exists for any Boolean algebra of propositions; i.e., the propositions can all be mapped onto the set $\{T, F\}$ in such a way that once truth values have been assigned to the elementary propositions (a valuation), the usual truth tables hold for all propositional connectives. In other words, the values of truth and falsity can be consistently applied to all the propositions in a Boolean lattice.

Gleason showed that no such truth function can exist for the quantum-mechanical lattice of subspaces of a Hilbert space of more than two dimensions. So a truth-valued semantics for lattice quantum logic is out. The problem arises from the existence of "incompatible" propositions in lattice quantum logic. Classical logic assumes that any proposition that has been asserted once in an argument can be asserted again at any later stage of that argument. Yet, even in many ordinary (i.e., non–quantum-mechanical) situations, this may fail to be the case for propositions that do not include an explicit specification of the implied background conditions. For example, the various states of matter (solid, liquid, gaseous, and plasma) are incompatible with each other. So a statement such as 'The hardness of a certain sample of some metal has such and such a value' may be asserted as a proposition (correctly or incorrectly) only as long as the metal is in the solid state; if the sample is later melted or vaporized, such a statement can be regarded as either false or meaningless. One way out of such difficulties is to consider such statements as propositions only when sufficient background conditions are added to make them "eternally" true or false—for example, 'The hardness of a certain sample of some metal has such-and-such a value under such-and-such conditions of temperature and pressure.' Classical logic is perfectly adequate to treat such propositions. Alternatively, one might continue to treat the incomplete statements as propositions, and modify one's logic in a way that admits incompatible propositions, such as 'The hardness of a certain sample of some metal has such-and-such a value' and 'The viscosity of the same sample has such-and-such a value'—each of the two statements having a possibility of being true or false only when the sample is in a solid or liquid state, respectively (Stachel 1986, 249–264). While a perfectly respectable way of handling such propositions, such a modification of logic hardly sheds new light on the existence of several incompatible states of matter, or on their respective properties of hardness and viscosity.

What Can Quantum Logic Do?

As suggested earlier (see "What Is a Quantum-Mechanical Proposition?" above), a rather similar situation is obtained in quantum mechanics. As Bohr (1958) emphasized, as long as the only statements admitted as propositions concern what he calls "phenomena," there is no need to go beyond classical logic in the treatment of such propositions. If one chooses to drop the description of the experimental arrangement and admit such "bare" statements as 'The position of the particle is x' as a proposition, then one must have recourse to a nonclassical logic in the treatment of such propositions. But, as in the classical case, it is hard to see how such recourse leads to a deeper understanding of those features peculiar to quantum mechanics often considered paradoxical.

The following example has been cited (see e.g., Putnam 1969) to show how the failure of the distributive law in lattice quantum logic leads to a deeper understanding of quantum mechanics: In this logic, as in classical logic, any statement of the form "The particle has position x and momentum p_i" (which shall be abbreviated "$X \wedge P_i$") is meaningful; but in lattice logic it is always false ($X \wedge P_i = \phi$, the absurd proposition), because $X \cap P_i = \{0\}$, the zero-dimensional linear subspace. Hence, the conjunction of such statements for all possible values of the momentum is false ($S = \sum_i (X \cap P_i) = \phi$) because the union of any number of exemplars of the zero-dimensional linear subspace is still $\{0\}$. On the other hand, the statemen: 'The particle has momentum p_1 or the particle has momentum p_2 or …. or the particle has momentum p_N,' where the list runs over all possible momenta (complications arising from a denumerably infinite or continuous set of values of quantities are ignored in the case of $\sum_i \vee P_i$), is always true ($\sum_i \vee P_i = I$), because $\sum_i Sp(P_i) = H$, the entire Hilbert space. So the statement S', "The particle has position x and (the particle has momentum p_1 or the particle has momentum p_2 or…. or the particle has momentum p_N)" ($S' = X \wedge (\sum_i P_i) = X \wedge I$) is true iff the statement "The particle has position x" is true ($S' = X$). If the distributive law held, S (which is always false) would be identical to S', which may be true or false—a clear contradiction. The failure of the distributive law means that there is no contradiction involved if S and S' fail to have the same truth value.

These formal manipulations are certainly correct. But what insight is gained in terms of the *meaning* of the relevant propositions? S' seems to assert that the particle has a definite position and *some* momentum. But this is a false appearance, due to an implicitly classical reading of the disjunction of momenta proposition. 'The particle has momentum p_1 or the particle has momentum p_2 or … or the particle has momentum p_N' is true only because of the quantum-logical definition of "or" as the span of the relevant linear subspaces: The span of all the one-dimensional orthogonal linear subspaces representing possible values of the momentum is the entire Hilbert space; so the assertion of this disjunction proposition amounts to no more than the trivial assertion 'The system exists.' It will continue to hold true for states in which the momentum fails to have a definite value—in particular, for the state under consideration, in which the position *does* have a definite value! Thus the compound proposition S asserts no more than the simple proposition "The particle has position x," and really has nothing to do with its momentum.

An even simpler example of how the seemingly classical nature of a quantum-logical law conceals a quite perverse meaning of its quantum-logical equivalent is provided by the law of the excluded middle, $A \vee (\neg A) = T$, which corresponds to the linear subspace $A \cup \mathrm{Sp}(\perp A)$. Classically, the law asserts that either a certain property or propositional holds true of a system or it does not. But lattice-logically, since $A \cup \mathrm{Sp}(\perp A) = H$, the entire Hilbert space, which corresponds to the trivial proposition ('The system exists'), the law has nothing to do with whether one may always assert one of the two alternatives, 'The system has property A' or 'The system does not have property A.' The most it does (via its negation) is show that *if* one is able to assert one of the alternatives, then one cannot assert the other.

Finally, note that quantum logics are always weaker than classical logic, which manifests a tendency to maximal definiteness (see Waismann 1968). Indeed, Gödel (1930) proves the completeness of the classical first-order predicate calculus: Every tautology of this calculus can be proven; and addition of a further (extensionally inequivalent) proposition to the set of tautologies would lead to contradictions. In other words, classical first-order logic cannot be strengthened; any nonclassical first-order calculus can represent only a *weakening* of the classical one; that is, in such a calculus, certain classical tautologies will be false for some valuation(s) of the primary propositions. Kochen and Specker (Hooker 1975, 263–276) showed this explicitly by construction for certain *pBa* values arising in quantum mechanics. As discussed above for the case of quantum-mechanical propositions, classical logic will always be valid if propositions are stated with sufficient attendant conditions to ensure that they will always be either true or false; and the possibility of nonstandard logics that weaken the first-order predicate calculus arises from the admission as propositions of statements that omit certain of these attendant conditions.

Operational or Dialogic Approaches

Various operational or dialogic logics, based on the temporal sequence of possibly incompatible assertions, have been proposed as aids in the interpretation of quantum mechanics. One gives up the assumption that once a proposition is asserted in an argument, it can always be reasserted at any later stage. Although there seems no reason why this need be the case, this observation has often (e.g., Lorentzen 1965) been associated with the dialogic approach to logic, which attempts to found logic on the formulation of a successful strategy to defend a proposition in dialogue with an opponent. This approach has been applied to quantum mechanics (Mittelstaedt 1978). Hintikka (2002) bases his approach on an extension of first-order logic that he argues is needed for many cases in which mutually interdependent properties are present. All of these are "logic(s) for" approaches to quantum logic:

> Putnam proposes changing logic for physics' sake; my starting point is a change in (or, rather, an extension of) basic logic for logic's sake, or more explicitly, for the sake of the expressive power of a logical language. ... [O]nce this extension has been carried out, we do not need any separate new quantum logic. (Hintikka 2002, 208)

Consistent-Histories Approach

The quantum logics considered so far start from the concept of the *state* of a system at each instant of time—or at least the state of the system's properties, or propositions about its properties, at each instant. An alternative approach to quantum mechanics (see "Probability Amplitudes and Feynman Paths" above) is based on the concept of *process*, a conditional probability being associated with the entire development of a system between two instants of time (preparation and registration). In nonrelativistic (i.e., Galilei-invariant) quantum mechanics, a unique link between the state and process approaches exists: The time interval of any process is uniquely divided into instants of absolute time, and a state is associated with each instant. But in the case of relativistic" (i.e., Lorentz-invariant) quantum mechanics or quantum field theory, such a unique division does not exist (indeed, the initial and final instants need not be parallel timelike hyperplanes, or even hyperplanes at all) in Minkowski space-time. In non–flat background space-times, timelike hyperplanes generally do not exist, so the process approach is much more natural in relativistic cases. Even in nonrelativistic quantum mechanics, the process approach has much to recommend it. As noted earlier, the basic task of any quantum-mechanical formalism is to compute a probability amplitude for a process; and the state function is an auxiliary computational device (see "Prediction, Retrodiction, and State Functions" above).

In contrast to the Feyman approach (see "Probability Amplitudes and Feynman Paths" above), the recently developed consistent-histories approach (Isham 1994, 1995, and 1997) treats quantum systems as closed; it asserts that under certain conditions, it is possible to assign *probabilities* to collections of individual paths associated with a given process, or at least to collections of subensembles of such paths, or *histories,* even if these paths cannot be distinguished within the experimental arrangement defining the process. For any two histories α, β a decoherence function $d(\alpha, \beta)$—a complex number of modulus ≤ 1—may be calculated using standard quantum mechanics as if intermediate measurements had been made. It is then assumed that this function is meaningful even in the absence of such intermediate measurements.

A set of such histories is said to be *complete* if the physical realization of any one history would exclude that of all the others; and, if this set of alternatives is chosen for testing, then one of them must be realized. The set is *d-consistent* if $d(\alpha, \beta)$ vanishes for any distinct pair in the set. If these conditions hold, the probability that any history α in the set will be realized is $d(\alpha, \alpha)$, and the sum $\sum d(\alpha, \alpha) = 1$.

There are many complete, mutually incompatible *d*-consistent sets of histories; and since the system is closed, one cannot adopt the usual Feynman interpretation and say that the choice of a preparation and registration apparatus fixes one of them. The coherent-histories proposal is to consider all of these sets together, in an interpretation of "many world-views" (Isham 1997) of quantum mechanics.

Such a collection of sets has the structure of an *orthoalgebra* (see e.g., Wilce 2002; Dvurečenskij et al. 2002), a collection of Boolean algebras pasted together into a structure similar to a partial Boolean algebra (see "Partial Boolean Algebra Logic" above). A probabilistic logic can be associated with this structure by associating a proposition with each history α, i.e., "The system followed history α during the process." The set of propositions form a nonbivalent logic that falls within the general class of intuitionistic logics, for which the distribution law holds, but the law of the excluded middle fails. Rather than valuations in which propositions are either true or false, consistent-history propositions are associated with probabilities, that is, valuations in the interval [01]. Isham proposed this "logics for" approach in the hope of developing a structure of sufficiently broad scope to be applicable to background-independent formulations of quantum gravity.

Conclusion

The final word on quantum logic may be left to Bohr (1939): "The question of the logical forms which are best adapted to quantum theory is in fact a practical problem, concerned with the most convenient manner in which to express the new situation that arises in this domain."

JOHN STACHEL

References

Beltrametti, E., and G. Casinelli (1981), *The Logic of Quantum Mechanics*. Reading, MA: Addison-Wesley.

Berry, Michael V., and K. E. Mount (1972), "Semiclassical Approximations in Wave Mechanics," *Reports on Progress in Physics* 35: 315–397, esp. 371–393.

Birkoff, G. and von Neumann, J. (1936), "The Logic of Quantum Mechanics," 37: 823–843.

Bohr, Niels (1939), "The Causality Problem in Atomic Physics," in *New Theories in Physics*. Warsaw: International Institute of Intellectual Cooperation.

——— (1958), "Quantum Physics and Philosophy," in R. Klibansky (ed.), *Philosophy in the Mid-Century*, vol. 1. Florence: La Nuova Italia, 308–314.

Coecke, Bob, David Moore, and Alexander Wilce (2000), *Current Research in Operational Quantum Logic: Algebras, Categories, Languages*. Dordrecht, Netherlands: Kluwer.

Dalla Chiara, Maria Luisa, and Roberto Giuntini (1998), "Quantum Logic, Quantum Histories and Quantum Turing Machines." http://www.illc.uva.nl/j50/contribs/chiara/chiara.pdf.

——— (2001), "Quantum Logic." In *Handbook Of Philosophical Logic*, 2nd ed., vol. E2: *Alternatives to Classical Logics 2*. http:///www.dcs.kcl.ac.uk/research/groups/logic/philo/toc-all.html.

Dvurečenskij, Anatolij (1993), *Gleason's Theorem and its Applications*. Dordrecht and Boston: Kluwer.

Dvurečenskij, Anatolij, Sylvia Pulmannová, and Karl Svozil (2002), *Partition Logics, Orthoalgebras and Automata*. http://tph.tuwien.ac.at/~svozil/publ/dvur.htm.

Février, Paulette (1937), "Les relations d'incertitude de Heisenberg et la logique," *Académie des Sciences (Paris), Comptes Rendus* 204: 481–483.

Feynman, Richard P. (1968) (with R. B. Leighton and M. Sands), *The Feynman Lectures on Physics*, vol. 3. Reading, MA: Addison-Wesley.

Gibbins, Peter (1987), *Particles and Paradoxes: The Limits of Quantum Logic*. Cambridge: Cambridge University Press.

Gödel, Kurt (1930), "Die Vollständigkeit der Axiome des logischen Funktionkalküls," *Monatshefte für Mathematik und Physik* 37: 349–360.

Haack, Susan (1996), "Chapter 8. Quantum Mechanics," in *Deviant Logic, Fuzzy Logic*, 2nd ed. Chicago: University of Chicago Press, 148–167.

Hardgree, Gary M. (1979), "The Conditional in Abstract and Concrete Quantum Logic," in C. Hooker (ed.), *The Logico-Algebraic Approach to Quantum Mechanics*. Dordrecht and Boston: D. Reidel, 49–108.

Hellman, Geoffrey (1981), "Quantum Logic and Meaning," *PSA 1980/ Proceedings of the 1980 Biennial Meeting of the Philosophy of Science Association*, vol. 2. East Lansing, MI: Philosophy of Science Association, 493–451.

Hintikka, Jaakko (2002), "Quantum Logic as a Fragment of Independence-Friendly Logic," *Journal of Philosophical Logic* 31: 197–209.

Hooker, C. A. (ed.) (1975), *The Logico-Algebraic Approach to Quantum Mechanics*, vol. 1, *Historical Evolution*. Dordrecht and Boston: D. Reidel.

——— (1979), *The Logico-Algebraic Approach to Quantum Mechanics*, vol. 2, *Contemporary Consolidation*. Dordrecht and Boston: D. Reidel.

Isham, C. J. (1994), "Quantum Logic and the Histories Approach to Quantum Theory," *Journal of Mathematical Physics* 35: 2157–2185.

——— (1995), "Quantum Logic and Decohering Histories," talk given at conference "Theories of Fundamental Interactions," Maynooth, Ireland, 24–26 May. Published in *Fundamental Interactions 1995*, 30–44.

——— (1997), "Topos Theory and Consistent Histories: The Internal Logic of the Set of All Consistent Sets," *International Journal of Theoretical Physics* 36: 785–814.

Jammer, Max (1974), *The Philosophy of Quantum Mechanics*. New York: Wiley-Interscience.

Lorentzen, Paul (1965), *Formal Logic*. Dordrecht, Netherlands: Reidel, 1965.

Malament, David (2002), "Notes on Quantum Logic." June 6. hypatia.ss.uci.edu/lps/home/fac-staff/faculty/malament/prob-determ/handouts/QuantumLogic.pdf.

Megill, Norman D. (2005), *Quantum Logic Explorer Home Page*. http://us.metamath.org/qlegif/mmql.html.

Mittelstaedt, Peter (1978), *Quantum Logic*. Dordrecht, Netherlands: Reidel.

Pavičić, Mladen, and Norman D. Megill (1999), "Non-Orthomodular Models for Both Standard Quantum Logic and Standard Classical Logic: Repercussions for Quantum Computers," *Helvetica Physics Acta 72*: 189–210.

Pták, Pavel, and Sylvia Pulmannová (1991), *Orthomodular Structures as Quantum Logics*, Dordrecht, Boston, and London: Kluwer.

Putnam, Hilary (1969), "Is Logic Empirical?" in Robert S. Cohen and Marx W. Wartofsky (eds.), *Boston Studies in the Philosophy of Science*, vol. 5. Dordrecht and Boston: D. Reidel, 216–241.

Reichenbach, H. (1944), *Philosophic Foundations of Quantum Mechanics*. Los Angeles, CA: University of California Press.

Schiller, Ralph (1962), "Quasi-Classical Theory of the Nonspinning Electron," *Physical Review* 125: 1100–1108.

Stachel, John (1986), "Do Quanta Need a New Logic?" in Robert Colodny (ed.), *From Quarks to Quasars: Philosophical Problems of Modern Physics*. Pittsburgh: University of Pittsburgh Press, 229–347.

Stoll, Robert R. ([1963] 1979), *Set Theory and Logic*. New York: Dover.

Strauss, Martin ([1936] 1972), "The Logic of Complementarity and the Foundation of Quantum Theory," in *Modern Physics and its Philosophy*. Dordrecht and Boston: D. Reidel, 186–199. Originally published as "Zur Begründung der statistischen Transformationstheorie der Quantenphysik," *Berliner Berichte 1936*: 382–398.

Svozil, Kurt (1998) *Quantum Logic*. Singapore: Springer.

von Neumann, John (1927), "Mathematische Begründung der Quantenmechanik," *Gesellschaft der Wissenschaften zu Göttingen. Nachrichten*: 1–57.

——— ([1932] 1955), *Mathematical Foundations of Quantum Mechanics* [*Mathematische Grundlagen der Quantenmechanik*]. Translated by Robert T. Beyer. Princeton, NJ: Princeton University Press.

von Weiszäcker, Carl (1958), "Die Quantentheorie der einfachen Alternative," *Zeitschrift für Naturforschung* 13a: 245–253.

Waismann, Friedrich (1968), "Are There Alternative Logics?" in Rom Harré (ed.), *How I See Philosophy*. London: Macmillan, 67–90.

Wilce, Alexander (2002), "Quantum Logic and Probability Theory," *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/qt-quantlog.

QUANTUM MEASUREMENT PROBLEM

While there are many problems surrounding the account of measurement in quantum theory, the so-called problem of measurement in quantum theory refers to a specific difficulty that has plagued its interpretation from early in its history and that continues to defy a widely accepted resolution. This entry addresses the problem primarily in terms of the relevant features of the formalism of standard quantum theory, and the usual account of measurement in quantum theory. However, as will be emphasized throughout, the problem of measurement is more generic than the terms in which it will be primarily described here, for it arises in every known quantum theory, in some form or other, regardless of the details of the formalism or the account of measurement (see Quantum Mechanics).

In briefest form, the problem is this: At the end of measurement-like interactions (between a measured system and a measuring apparatus), quantum theory sometimes (indeed, generically) assigns the "wrong" state to physical systems (henceforth "systems"), including systems that are apparently the objects of everyday observation. The wrong state of a system refers to a state whose standard interpretation fails to assign properties to the system that are apparently directly observed (for example, the property apparently enjoyed by many everyday objects of having a very well defined location in space). While the details to follow are important for a clear understanding, this simple statement of the problem should be kept firmly in mind.

Quantum States

States as Probability Measures

In quantum theory, the "state" of a system, S, is a probability measure, a function from the possible values of all physical quantities (such as position and energy, henceforth "observables") to numbers between 0 and 1, whose intended interpretation is the "chance" that S has the given value for the given observable. (How one understands these chances is a matter of interpretation, discussed below.) For each observable, the probability measure is "complete" in the sense that the sum of the probabilities for each possible value of a given observable is 1. The classical ("ignorance") interpretation of this fact would be that the system does in fact have one of the possible values, but which value it has is unknown. For example, in classical statistical mechanics, the state of a gas is a probability measure over all of the possible configurations of molecules in the gas. The usual understanding of this measure is that the gas has just one of these possible configurations, the probabilities arising purely from our ignorance about which configuration it is. For now, however, do not adopt this interpretation, nor any other.

Mixed versus Pure States

To get started on the task of interpretation, it helps to distinguish between "mixed" and "pure" states. Given two states, σ and σ', which, recall, are probability measures, one can form their "mixture," $\sigma'' = p\sigma + (1 - p)\sigma'$, with $0 < p < 1$. The mixture σ'' is a "mixed state." (Mixtures can contain an arbitrary number of states, but two is enough for the purpose of exposition.)

One appropriate use of mixed states is to describe a system whose actual state is unknown. For example, if the system is chosen randomly from an "urn" of systems, some in the state σ and some in the state σ'', in the proportion p to $1 - p$, then one should assign the system the mixed state σ''. In the standard parlance, mixed states that arise in this way are called "proper mixtures," as opposed to "improper mixtures," considered below.

Pure states are states that cannot be written as mixtures. Reflection on this definition produces a more positive characterization of pure states. Pure states are, intuitively, states of maximal information. For present purposes, the simplest account of that notion is this: A state is pure if and only if it assigns a definite value to some maximally specific (henceforth "maximal") observable A in the sense that it assigns probability 1 to one of A's possible values, and probability 0 to the rest.

Here and below, one should not presume that the qualifier "maximal" is essential to the statement of the problem of measurement. Its presence is for technical convenience. An observable A is maximal if it is not the coarse-graining of any other observable; that is, there is no other observable B such that there is a many-to-one map, m, from the possible values of B onto the possible values of A, where for any possible value b of B, every state that assigns probability 1 to b also assigns probability 1 to $m(b)$.

In quantum theory, any state that assigns a definite value (in the sense defined above) to a maximal observable is indeed a state of maximal information (a notion that can be made mathematically precise). In classical physics, states of maximal information assign definite values to *all* observables. In quantum theory, there are, provably, no such states. Instead, every pure state assigns a definite value to some maximal observable, and assigns nontrivial probabilities (that is, probabilities that are neither 0 nor 1) to the values of other observables.

The Standard Interpretation of Pure States

For the moment, it suffices to adopt the interpretation of mixed states mentioned above, *viz.,* that they represent ignorance about the actual, pure state of a system. (This interpretation is controversial, though standard in the case of so-called proper mixtures.) Note, however, that an interpretation of the probabilities in a mixed state does not necessarily carry over to the probabilities generated by a pure state (henceforth "quantum probabilities")—indeed, there are serious obstacles to a straightforward ignorance interpretation of quantum probabilities, a point discussed below.

It is helpful to begin with a minimal interpretation of quantum probabilities, one that is often adopted in applications, the so-called "eigenstate-eigenvalue link," which is a system in the state σ has the value a for the observable A if and only if σ assigns probability 1 to a and (therefore) probability 0 to the other possible values of A. (Such a state is necessarily an eigenstate of A corresponding to the eigenvalue a—hence the name *eigenstate-eigenvalue link,* coined by Fine [1969].) The measurement problem, in a very naive form, amounts to the fact that some quantum states fail to assign definite values to the results of a measurement, given the eigenstate-eigenvalue link. This fact is intimately connected with the principle of superposition in quantum theory.

Superpositions

Consider some maximal observable A and two possible values for it, a and a'. Let the state σ assign probability 1 to a, and let σ' assign probability 1 to a'. (Such states always exist. The requirement that A be maximal is only for ease of exposition.) The principle of superposition says that one can form a third pure state, σ'', with the following properties:

1. σ'' assigns a probability to the value a for A strictly between 0 and 1, and similarly for the value a';
2. σ'' assigns probability 0 to all values of A except the values a and σ'.

In the context of the eigenstate-eigenvalue link, the principle of superposition is curious at best, for it implies the existence of a state (σ'') in which the system has neither the value a nor the value a' for the observable A, and yet the probability that it has some other value is 0.

In any case, it is crucial to keep in mind that the superposition of two pure states is itself *pure*, not mixed. Classically, too, there are states having properties 1 and 2 given above; however, such states are necessarily mixed, and in this case, one could adopt an ignorance interpretation. On the other hand, superpositions are pure, and (at this stage of the discussion) their interpretation is given by the eigenstate-eigenvalue link. The existence of superpositions ultimately gives rise to the problem of measurement.

Measurements

Constraints on Measurement

The problem of measurement is perhaps misnamed, for it arises generically in quantum theory, not only in the context of a measurement. However, the problem is particularly clear in the context of measurements, which makes it a useful starting point.

A measurement establishes a correlation between some observable for the measured system (typically called the "measured observable") and some observable for the measuring apparatus (the "pointer-observable"—one can imagine, literally, a needle that points to a number on a scale). In the ideal case, the pointer-observable is perfectly correlated with the measured observable.

Noting that a measurement is a dynamical process, it is natural to make the following (extremely!) minimal requirement for a measurement. Let U represent the time-evolution of states during the measurement of a measured observable A by a pointer-observable B, so that if σ is the state of a system at the start of the measurement, then $U(\sigma)$ is the state after the measurement. We require: For any two states, Σ and Σ', of the compound system (the measured system plus the apparatus), if Σ and Σ' differ in their probabilities for at least one value of A, then $U(\Sigma)$ and $U(\Sigma')$ differ in their probabilities for at least one value of B. Intuitively, U renders the pointer-observable "somehow sensitive to" the measured observable B.

The Quantum-Theoretic Representation of Measurements

There are models of measurements satisfying this weak constraint. Indeed, there are models satisfying the following stronger, though still natural, constraint: Let Σ and Σ' each assign probability 1 to the (distinct) values a and a' of A, respectively; then $U(\Sigma)$ and $U(\Sigma')$ must assign probability 1 to two distinct values of B. That is, under these conditions, U establishes a perfect correlation between the values of A and B. (If the compound system's state continues to assign probability 1 to the value a for A, then the measurement is said to be "of the first kind," or "nondisturbing." In general, measurements are *not* nondisturbing.)

In general, given any time-evolution describing any sort of interaction between two systems, there will be many observables in the two systems that become correlated (though generally not perfectly) as a result of the interaction. In other words, interactions between systems will, most of the time, establish the sort of correlation that is sufficient to satisfy the weak constraint on measurements. Hence whatever problems flow from meeting this weak constraint will be quite generic to quantum theory—and the measurement problem does flow more or less directly from the imposition of this weak constraint.

On the other hand, the problem follows very directly from the stronger constraint, and it is therefore helpful to consider that more restricted version of the problem first. One should keep in mind, moreover, that there are interactions, which apparently occur in the real world, that satisfy the strong constraint. Actual measurements are not an example of such interactions, if one measures success as establishing a perfect correlation between the *intended* pointer-observable and the *intended* measured observable. However, the lack of a perfect correlation between these two observables does not imply that there are not two *other* observables (one for the measured system and one for the apparatus) between which the measurement *does* establish a perfect correlation. Indeed, one can prove that after just about any interaction between two systems, there almost always is a pair of (nontrivial) observables—one for each system—that is perfectly correlated. (The proof is by way of the biorthogonal decomposition theorem, and the point was recognized by Schrödinger [1935].)

The Problem

The State at the End of a Measurement

It is a straightforward consequence of the considerations above and the formalism of quantum theory (specifically, the linearity of the equation of motion) that if U satisfies the strong constraint, then there are initial states for the measured system (specifically, states that are superpositions of two or more eigenstates of the measured observable A) such that, at the end of the measurement, the state of the compound system is a superposition of eigenstates of B, so that, according to the eigenstate-eigenvalue link, the measuring apparatus has no definite value for the pointer-observable.

(Strictly speaking, one should say not "the observable *B*," but "the observable $B \times I$," which is an observable on the compound system whose value is determined solely by the value of *B* on the apparatus, so it "ignores the measured system.") One can conceive the reason as follows: In the context of the formalism of quantum theory, the strong constraint on measurement essentially requires that the state of a measuring apparatus must "formally mimic" the state of the system that it is measuring. When the system that it is measuring has a definite value for the measured quantity, then mimicking is exactly what one wants, of course. The problem is that when the measured system is in a superposition of definite-valued states for the measured quantity, the same requirement implies that the apparatus will, again, mimic the state of the measured system, but, alas, in this case "mimicking" means that the apparatus is itself in a superposition of definite states of the pointer-observable.

It is also worth noticing that as a result of the measurement, the measured system becomes entangled with the apparatus. While this fact is perhaps not, strictly speaking, a problem, some find it counterintuitive that merely as a result of measurement, the measured system (and apparatus) come to lack their own pure states.

Proofs of the More General Problem

Perhaps the first notable "proof" of the problem more or less as we have just considered it was von Neumann's ([1932] 1955) derivation, which demonstrated, roughly, that in a nondisturbing measurement (see above) satisfying the strong constraint, the final state of the apparatus must be a superposition of eigenstates of the pointer-observable. A series of increasingly more general proofs followed, for example from Wigner (1963), d'Espagnat (1966), Earman and Shimony (1968), Fine (1969), and Shimony (1974). A review of this history, together with arguably the simplest proof of the theorem formulated by Fine and later reconsidered by Shimony, was given by Brown (1986).

Two features of these theorems (particularly the later ones) are notable here. First, they do *not* adopt the eigenstate-eigenvalue link, but a much weaker condition on when observables have values. Second, they adopt (more or less) the weak, rather than the strong, constraint on measurements.

The interpretive condition that these theorems adopt is, roughly, that the final state of the apparatus assigns a definite value to the pointer-observable for the apparatus just in case the final state of the compound system is a mixture of states each of which has a definite value for the pointer-observable according to the eigenstate-eigenvalue link. In other words, they allow the adoption of an ignorance interpretation of mixed states of the compound system, accepting that when the compound system has the sort of mixed state just described, it is actually in one of the pure states appearing in the mixture, and therefore, according to the eigenstate-eigenvalue link, the apparatus has a definite value for the pointer-observable.

Note that the condition applies to the *compound* system, *not* to the apparatus on its own. The difference is subtle, but crucial. For example, in an ideal measurement, the apparatus by itself will *always* be in a mixture of (pure) eigenstates of the pointer-observable. It does not follow, however, that this mixture can be given an ignorance interpretation.

Why not? The short answer is that the mixture, in this case, is improper, meaning that it does not arise in the way that proper mixtures do (described above), but rather from ignoring the rest of the universe, in this case, the measured system. (Formally, the improper mixture is obtained by tracing out the degrees of freedom corresponding to the measured system.) One cannot straightforwardly adopt an ignorance interpretation of this improper mixture. The following considerations illustrate the difficulty.

Let *I* represent the logically trivial property possessed by all systems. For any two properties, *P* and *Q*, of two arbitrary systems, 1 and 2, let the compound property "system 1 has *P* and system 2 has *Q*" be denoted $P \wedge Q$ and the compound system be denoted $1 \wedge 2$. Consider the following, seemingly innocuous, principle: If system 1 has the property *P,* then $1 \wedge 2$ has the property $P \wedge I$.

This principle is violated by an interpretation that supplements the eigenstate-eigenvalue link with an ignorance interpretation of improper mixtures, for let the compound system be in a superposition of two states, each of which assigns (according to the eigenstate-eigenvalue link) a definite value for *A* to system 1, and a definite value for *B* to system 2. This superposition is, of course, a pure state, but the subsystem 2 is in an improper mixture of eigenstates of *B.* According to the proposed interpretation, then, subsystem 2 has a definite value for *B.* (Similar remarks hold for subsystem 1.) However, according to the eigenstate-eigenvalue link, the compound system does *not* have any value for the compound observable $A \times B$, thus leading to a violation of the seemingly innocuous principle above. (So-called "modal" interpretations of quantum theory attempt, in various ways, to dance around this issue and adopt an ignorance interpretation of [at least some] improper mixtures. The literature on this topic is vast, but see Bacciagaluppi (2003) for a lengthy discussion, with plenty of references.)

Hence, if one prefers to hold on to the apparently innocuous principle, then it is advisable to allow the ignorance interpretation only for (proper) mixtures of the total, compound system, and not for the improper mixed states of the components. Brown and his predecessors took this path. Doing so at least allowed them to consider the case—more general than that considered in many expositions of the measurement problem—in which the initial state of the compound system is mixed. The point is that in this case, allowing an ignorance interpretation of mixed states for the compound system, and insisting on only a very weak notion of what constitutes a measurement, one can still show that in quantum theory, measurements can lead to states for the compound system in which the apparatus does not have any definite value for the pointer-observable.

False Starts and Discredited Answers

There are many initially appealing, but ultimately insufficient, responses to the problem of measurement. A few of the most important or prevalent are considered here.

The Ignorance Interpretation of Quantum Probabilities

Probably the most obvious response to the problem of measurement is to question the proferred interpretation of quantum probabilities. Why not adopt a straightforward ignorance interpretation, according to which each observable in fact has a definite value all of the time, the quantum probability representing ignorance about the actual value? Adopting such an interpretation would certainly resolve the problem of measurement, because all observables, and *a fortiori* all pointer-observables, always have a value.

However, various theorems, most notably that of Kochen and Specker (1967), pose serious challenges to an ignorance interpretation. Kochen and Specker derive a contradiction from a few presuppositions, including that every quantum-theoretic observable has a definite value. The presuppositions of their derivation can be understood in various ways, but no matter how they are understood, violating them is a radical move. For example, one way to violate them is to deny the logical law of distributivity. While some have taken this route, clearly it is controversial. In any case, the main point is that the theorem of Kochen and Specker renders any naive ignorance interpretation of quantum probabilities untenable.

The Collapse Postulate

The practical response to the problem is the collapse postulate: At the end of a measurement, if the compound system is in a superposition of eigenstates of the pointer-observable, the state of the compound system "collapses" onto just one element in the superposition, corresponding to the actual result of the measurement. Such a collapse, which is a discontinuous change of state, contradicts the (continuous) equation of motion in quantum theory. Hence one would like a systematic principle, internal to the theory, dictating when such collapses occur. While there are adequate rules of thumb about when to collapse the state, there is no such agreed-upon principle. (A further discussion of the collapse postulate, and reasons for rejecting it as an interpretive principle, is in Dickson 1998, Chs. 1–2, with references therein.)

An additional problem for the collapse postulate is that it introduces a fairly explicit form of nonlocality into the theory. Given a pair of entangled systems that are well separated in space, suppose that a measurement is made on one of them. The result, in this view, is a collapse of the state. However, such a collapse will effect the state of the distant system as well. While there are arguments about why such a collapse does not violate the letter of the law of relativity, and while quantum theory is in general nonlocal independently of the collapse postulate, nonetheless that postulate does introduce a form of nonlocality that some find objectionable.

Decoherence

An increasingly common response to the problem of measurement appeals to decoherence. Very briefly: If the description of a measurement is made to take into account the interaction of the measuring apparatus with its environment (for example, the electromagnetic radiation in the environment), one can show (given careful though admittedly idealized modeling of this interaction) that the compound system composed of the apparatus and its environment itself undergoes an approximately nondisturbing measurement-like interaction in which the pointer-observable becomes the observable measured, by the environment. But the discussion above already entails that in this case (to a high degree of approximation), the state of the apparatus at the end of the interaction is a mixture of eigenstates of the pointer-observable. Hence, the suggestion goes, simply adopt an ignorance interpretation of this mixture, and conclude that the pointer-observable does, in fact, have a definite value.

Given the discussion above, however, the problem with this response should be clear. The mixture in question is improper, and so cannot straightforwardly be given an ignorance interpretation. Something more by way of interpretation *must* be said to justify this interpretation.

Decoherence does, however, help with (though it does not resolve) the most general form of the problem of measurement. In its most general form, the problem is that quantum theory apparently fails to assign definite values to observables that appear, on the basis of everyday observation, to have definite values. In the special case of a measurement, at least the state of the apparatus is a mixture, albeit improper, of "desired" states. A task for interpreters of the theory, then, is somehow to parlay this fact to secure the definiteness of the pointer-observable. But what about interactions that are not explicitly measurements? What can one say about the states of the subsystems involved in the interaction? Decoherence promises to secure the following: If the subsystem is "relatively large" (e.g., $>10^{-5}$cm across), its interaction with the environment will entail that (very quickly) its state will become (very close to) a mixture, albeit improper, of desired states. The qualifiers "very quickly" and "very close to" should not be ignored, but at the same time, one must admit that decoherence does seem to reduce the most general form of the problem of measurement to the more specific form that arises in explicit measurements. (An excellent discussion of decoherence, focused on explaining why it is not sufficient to resolve the problem of measurement, is in Bub 2000, 207–218.) Even so, however, this strategy has merely reduced the general problem to one that continues to elude a satisfactory solution.

Michael Dickson

References

Bacciagaluppi, Guido (2003), *Modal Interpretations of Quantum Mechanics*. Cambridge: Cambridge University Press.

Brown, Harvey (1986), "The Insolubility Proof of the Quantum Measurement Problem," *Foundations of Physics* 16: 857–870.

Bub, Jeffrey (2000), *Interpreting the Quantum World* Cambridge: Cambridge University Press.

d'Espagnat, Bernard (1966), "Two Remarks on the Theory of Measurement," *Nuovo Cimento Supplement* 4: 828–838.

Dickson (1998), *Quantum Chance and Nonlocality* Cambridge: Cambridge University Press.

Earman, John, and Abner Shimony (1968), "A Note on Measurement," *Nuovo Cimento B* 54: 332–334.

Fine, Arthur (1969), "On the General Quantum Theory of Measurement," *Proceedings of the Cambridge Philosophical Society* 65: 111–121.

Kochen, Simon, and E. P. Specker (1967), "On the Problem of Hidden Variables in Quantum Mechanics," *Journal of Mathematics and Mechanics* 17: 59–87.

Schrödinger, Erwin (1935), "Discussion of Probability Relations Between Separated Systems," *Proceedings of the Cambridge Philosophical Society* 31: 555–563.

Shimony, Abner (1974), "Approximate Measurement in Quantum Mechanics," *Physical Review D* 9: 2321–2323.

von Neumann, John ([1932] 1955), *Mathematical Foundations of Quantum Mechanics* Princeton, NJ: Princeton University Press.

Wigner, Eugene (1963), "The Problem of Measurement," *American Journal of Physics* 31: 6–15.

See also **Quantum Field Theory; Quantum Mechanics**

QUANTUM MECHANICS

Quantum mechanics has been a subject of intense philosophical discussion for well over 75 years, and research on the philosophy of quantum mechanics continues apace. Undoubtedly, much progress has been made—especially in answering questions about the mathematical structures used in quantum mechanics. However, when it comes to spelling out the philosophical implications of quantum mechanics, there continue to be pronounced disagreements. At one extreme, some claim that the moral of quantum mechanics is a negative one about the limited applicability of human concepts; that is, the concepts humans have developed to cope with everyday experience do not apply without restriction to the phenomena encountered in quantum mechanics. So, the upshot is that one must pay special attention to whether the proper conditions are in place for applying these concepts. At the other extreme, some claim that quantum mechanics supplies positive evidence for the existence of multiple universes, and that each time a measurement occurs, the universe divides itself again. Between these two extremes there is a wide variety of competing interpretations of quantum mechanics, and the range of philosophical options is both intimidating and exhilirating.

The Old Quantum Theory

It is often said that the introduction of quantum mechanics has brought about a radical change in our physical worldview. In particular, it is claimed, quantum mechanics is inconsistent with the intuitive picture supplied by classical mechanics. Thus, in order to understand what is supposed to be so revolutionary in quantum mechanics, it is necessary first to review the main features of classical mechanics.

Modern classical physics reached its definitive form in the "Hamiltonian" formulation of mechanics introduced in the mid-nineteenth century (see Classical Mechanics). In this formalism, the instantaneous state of a physical system is described by its generalized positions (q_i) and momenta (p_i). For example, the instantaneous state of a system of k particles is specified by $6k$ real numbers ($q_1,\ldots, q_{3k}, p_1,\ldots, p_{3k}$). The set of all such instantaneous states is called the phase space of the system. Moreover, the dynamics of systems in classical Hamiltonian mechanics is fully deterministic; that is, for any initial state $s(0)$, there is a unique state $s(t)$ for any future time t.

Classical statistical mechanics often investigates situations in which there is no certain knowledge of the precise microstate of the system. In this case, the instantaneous state of the system is represented by a probability distribution ρ over phase space. Furthermore, if the initial state is ρ, then the future state of the system can be computed by applying the deterministic dynamical laws to the individual points in phase space, and then by reapplying ρ. Thus, there is no difficulty in seeing apparent randomness at the macroscopic level as arising from a strictly deterministic dynamics of microsystems (see Statistical Mechanics).

Quantum Statistics

During the nineteenth century, classical statistical mechanics achieved many successes in explaining the macroscopic properties of bodies (e.g., the temperature of a gas) in terms of motions of their constituent atoms (which are subject to the laws of Newtonian particle mechanics). However, not all physicists were convinced of the atomic hypothesis. In particular, many believed that electromagnetism, rather than mechanics, would serve as the proper foundation for physics. As a result, efforts were being made to explain macroscopic quantities in terms of more fundamental electromagnetic quantities.

One crucial test case for this new statistics of radiation was the emission of radiation by a black-body. A black-body is the perfect mixer of electromagnetic radiation: No matter what frequency of radiation it absorbs (and it absorbs any frequency), it re-emits a range of frequencies of radiation whose intensities are a function only of the temperature of the body. The theoretical explanation for this fact is that the atoms in a black-body behave as small harmonic oscillators. When radiation is absorbed, these oscillators vibrate, each with some characteristic frequency.

Wilhelm Wien proposed that the energy density per unit frequency depends on frequency f and temperature T according to the formula $u(f, T) = af^3 \exp(-bf/T)$, where a and b are empirically determined constants. Although Wien's formula agreed with the available data (which were restricted to fairly high frequencies), its theoretical basis was suspect.

At roughly the same time, Lord Rayleigh and James Jeans showed that if classical statistical mechanics is applied to the oscillators in a black-body, the resulting distribution is given by the formula $u(f, T) = 8\pi f^2 kT$, where k is the Boltzmann constant. Clearly, the Rayleigh-Jeans formula diverges from Wien's in the range of high frequencies ($f \gg 1$), and so it disagrees with the empiricially observed values for the radiation distribution. But things get worse: the equipartition theorem of classical statistical mechanics entails that the distribution of energy over frequencies would tend toward a uniform distribution over the various degrees of freedom. But the higher frequencies are associated (classically) with higher energies, and therefore the system would tend toward an "ultraviolet catastrophe."

Experiments performed in 1900 by Lummer and Pringsheim showed that Wien's formula was inaccurate for radiation at low frequencies. In order to take account of this new data, Max Planck proposed the distribution $u(f, T) = af^3[\exp(bf/T) - 1]^{-1}$ that agrees approximately with Wien's formula in the case of high frequencies ($f \gg 1$), and with the Rayleigh-Jeans formula in the case of low frequencies ($f \ll 1$). Planck then supplied a theoretical derivation of his formula by assuming that an oscillator of frequency f can have an energy value only in the discrete set $E = hnf$, where h is a positive constant (now known as *Planck's constant*) and n ranges over the positive integers. (The constant b in Planck's formula turns out to equal h.) Moreover, since there are fewer ways to distribute energy among higher frequencies than among lower frequencies, statistical considerations entail that it is unlikely the energy would be concentrated among the higher frequencies. Thus, the ultraviolet catastrophe is avoided.

While Planck quantized the energy of the oscillators in a black-body, he did not quantize the energy of the electromagnetic field itself. It was Einstein, in 1905, who took this more radical step. (Kuhn [1978] argues that Planck did not intend to introduce quantum discontinuity, and its origin is properly traced to Einstein's 1905 paper.) Einstein noticed that for high frequencies, the formula for the entropy of black-body radiation has the same logarithmic volume dependence as the entropy of an ideal gas. The latter dependence results directly from the fact that the energy in an ideal gas is localized in molecules. So, Einstein inferred the same for high-frequency radiation, whose energy, he proposed, could be treated thermodynamically as if it is localized in units of hf (i.e., "light quanta"). He then went on to show that the light quantum hypothesis explains phenomena such as the photoelectric effect.

Bohr's Atomic Theory

As of 1912, the best available model of the atom was Rutherford's solar system model, according to which an atom consists of negatively charged particles (electrons) orbiting a dense nucleus of positive charge. However, Rutherford's model had a fatal defect: According to classical electrodynamics, an orbiting electron will radiate at a constant rate (proportional to the square of the magnitude of its acceleration) and will thereby quickly decelerate. In fact, an orbiting electron would spiral into the nucleus on the order of 10^{-10} seconds, which would entail a radical instability of matter.

Meanwhile, it was also known that atoms emitted characteristic frequencies of light when heated. When the emitted light is passed through a prism, the resulting frequencies of electromagnetic radiation satisfy some rather simple arithmetical relations. For example, the visible spectrum of hydrogen is given by the Balmer series:

$$f(m) = R\left(\frac{1}{2^2} - \frac{1}{m^2}\right), \tag{1}$$

where R is Rydberg's constant (approximately 109,737.3 cm^{-1}), and m is an integer greater than 2.

In a trilogy of papers in 1913, Niels Bohr proposed a "quantized" version of Rutherford's model of the hydrogen atom. Bohr's model not only solved the stability problem of Rutherford's model, but also explained the observed spectrum. Bohr's model is based on two fundamental postulates:

1. An atom has a discrete set of 'stationary states' that correspond to constant values of its total energy. In these stationary states, the mechanical properties of the electrons, (e.g., the relations between the radius, orbital angular momentum, and kinetic energy) are governed by classical mechanics. However, an atom does not emit radiation while in a stationary state (in violation of classical electrodynamics).
2. In certain circumstances, an atom will make a transition from one stationary state to another. But these transitions cannot be described by classical mechanics. Rather, during the transition the atom emits radiation of a single frequency (i.e., corresponding to one spectral line). If the energy of the initial state is $E(n)$, and the energy of the final state is $E(m)$, then the emitted radiation has frequency

$$f(n,m) = h^{-1}[E(n) - E(m)], \tag{2}$$

where h is Planck's constant.

More concretely, Bohr determined that the energy of the nth stationary state is $2\pi^2 m_e e^4 h^{-2} n^{-2}$, where m_e is the mass of the electron, and e is the charge of the electron. It follows then from equation 2 that

$$f(2,m) = 2\pi^2 m_e e h^{-3}\left(\frac{1}{2^2} - \frac{1}{m^2}\right). \tag{3}$$

If the known values for m_e, e, and h are plugged in, then the leading coefficient $2\pi^2 m_e e h^{-3}$ agrees with the empirically ascertained value for the Rydberg constant R. Thus, Bohr's frequency formula (equation 2) supplied a theoretical derivation of the Balmer series. Bohr's model also supplied an accurate value for the ionization energy $|E(1)|$ of hydrogen (i.e., the energy needed to remove the electron to an infinite distance from the nucleus), and for the diameter of the hydrogen atom in its ground state. In summary, Bohr's proposal was hugely successful, and it placed the idea of quantization at the forefront of theoretical physics.

Quantum Conditions

Following Bohr's quantization of the hydrogen atom, several research groups began to pursue similar lines of inquiry. Among these groups, Arnold Sommerfeld and his collaborators in Munich are especially noteworthy for their attempts to provide a systematic and mathematically rigorous framework for quantum theory.

Sommerfeld's (1919) quantization recipe involves two steps: First, describe the classically permissible motions of the system (i.e., those that satisfy the appropriate dynamical laws). Then impose "quantum conditions" to pick out a proper subclass of permissible motions. In particular, since the laws of classical Hamiltonian mechanics are deterministic, each point in phase space lies on a unique phase orbit—i.e., the set of points that can be reached from that point by a classical trajectory. The quantum conditions rule out all but a certain subclass of trajectories. In particular, suppose that a system is described by the canonical variables ($q_1, \ldots, q_k, p_1, \ldots, p_k$). Then the points on a permissible phase orbit satisfy the k equations $\oint p_i(q_i)d_{qi} = n_i h$, where the n_i are positive integers. [The variables' q_i values are assumed to be periodic, and the integral is taken over one period. For a system with one degree of freedom, the function $\oint p(q)dq$ evaluated at point x gives the area enclosed by the phase orbit through x and has the units of action (=energy × time).] The number n_i is called the "quantum number" of the corresponding phase orbit. Thus, each stationary state of a system with k degrees of freedom is specified by a k-tuple of quantum numbers.

For example, for an electron in a circular orbit around a nucleus, the standard choice of canonical variables are the angle $q = \theta$, and the orbital angular momentum $p = mr^2(d\theta/dt)$. In this case, $\oint p(q)dq = (2\pi)p$. Thus, the quantum condition requires that the orbital angular momentum is an integral multiple of $h/2\pi$. Combining this with the classical-mechanical equation for the kinetic energy of the orbiting electron entails that the energy of the system is $E(n) = 2\pi^2 m_e e^4 n^{-2} h^{-2}$, where n is a positive integer, in agreement with Bohr's derivation of the energy levels of the hydrogen atom.

In his original treatment of the hydrogen atom, Bohr made the simplifying assumption that the orbiting electron has one degree of freedom (corresponding to the radius of its orbit). In the more general treatment, a hydrogen atom is a system with three degrees of freedom, which can be given in spherical polar coordinates (r, θ, ϕ) for the orbiting electron. In this case, a stationary state of the atom is specified by three quantum numbers nr, $n\theta$, nϕ. Alternatively, a stationary state can also be specified by three numbers (n, k, m), where the principal quantum number $n = nr + n\theta + n\phi$ determines the energy; the azimuthal quantum number $k = n\theta + n\phi$ determines the total angular momentum; and the magnetic quantum number $m = n\phi$ determines the z-component of the angular momentum.

Sommerfeld's algorithm for quantization was succesfully applied to atoms in which a single electron orbits a positively charged shell (i.e., "hydrogenic" atoms). For example, by introducing the magnetic quantum number, Sommerfeld and Debye were able to explain the Zeeman effect, in which an atom's spectral lines separate into multiple lines ("multiplets") when it is placed in a magnetic field. However, the old quantum theory proved inadequate in application to multielectron atoms, even of the most simple sort (e.g., helium, in which two electrons orbit the nucleus). In particular, the old quantum theory gave incorrect predictions for the higher frequencies in the spectrum of helium, and for the ionization potential of helium. There were also a number of phenomena for which the old quantum theory simply could not provide any account—including the anomalous Zeeman effect (in which multiplet spectral lines recombine when the intensity of the magnetic field is increased), and the appearance of doublets in the spectrum of the alkali atoms. What is more, since there is no mechanically stable configuration for two (or more) electrons orbiting a single positively charged nucleus, the idea that the mechanical properties of stationary states can be treated classically could not be maintained. Indeed, it seemed that a more radical revision of classical physics was needed in order to supply a unified and empirically adequate account of atomic phenomena.

Quantum Mechanics

In the mid-1920s, two powerful new formalisms for quantum theory were introduced by Erwin Schrödinger and Werner Heisenberg, respectively. On the one hand, Schrödinger hoped to eliminate discontinuous transitions between stationary states by formulating a theory of "wave mechanics" in which a chargefield propagates through space according to a deterministic equation of motion. On the other hand, Heisenberg wanted to reformulate quantum theory on the basis of observable quantities (e.g., frequencies of spectral lines) and to rid it of unobservables (e.g., electron trajectories). The resulting formalism of noncommuting observable quantities was called matrix mechanics. For a short period of time, the existence of two competing quantum theories caused significant tension within the physics community. However, it was soon shown by von Neumann that there is a mathematical isomorphism between the theories of Schrödinger and Heisenberg. In fact, both theories can be subsumed under the more general Hilbert space formalism. In this formalism, states (or wavefunctions) are represented by vectors in a Hilbert space, and physical quantities (or observables) are represented by operators in this space.

The salient features of the Hilbert space formalism can be illustrated by looking at the space V of vectors in three-dimensional Euclidean space, i.e., arrows of variable length and direction with tails fixed at the origin (0, 0, 0). (In actuality, Hilbert spaces make use of complex numbers, but that detail is not important here.) The space V is equipped with two operations. First, for every vector $v \in V$ and real number a, there is a vector av that results from "stretching" v by the length a. (If a is negative, then av points in the opposite direction than v.) Second, if u and v are vectors in V, then the vector $u + v$ is defined by the following operation: Move the tail of the vector v along u, keeping their relative angles fixed, until the tail of v lies on the head of u; then, $u + v$ denotes the vector whose head coincides with the head of the transported vector. Let $\|v\|$ denote the length of the vector v.

(Thus, if the head of v lies at the point (a, b, c), then $\|v\| = (a^2 + b^2 + c^2)^{\frac{1}{2}}$.) If u and v are vectors, then their inner product is defined by

$$\langle u, v\rangle := \|u\| \cdot \|v\| \cdot cos\theta, \tag{4}$$

where θ is the angle between u and v.

More generally, a Hilbert space $\mathcal{H}$ is a vector space over the field of complex numbers C, with an inner product $\langle \cdot\ , \cdot\rangle$, and limit points for all of its Cauchy sequences (i.e., any sequence $\{vn\}$ of vectors that become arbitrarily "close" for large n). Vectors $u, v \in \mathcal{H}$ are said to be orthogonal if $\langle u, v\rangle = 0$. A subset W of vectors in $\mathcal{H}$ is called a subspace if when $u, v \in W$ and $a \in C$, then $av \in W$ and $u + v \in W$. In other words, a subspace of a Hilbert space is a subset that is itself a Hilbert space. In the case of R^3, there are two sorts of proper subspaces: Lines passing through the origin (one-dimensional), and planes passing through the origin (two-dimensional).

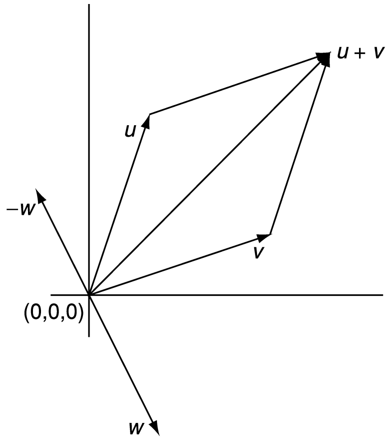


Fig. 1. Vector space operations.

An operator O on a Hilbert space $\mathcal{H}$ is a mapping that transforms vectors to vectors and that preserves linear relations. That is, $O(v + w) = Ov + Ow$ and $O(cv) = c(Ov)$ for any vectors v, w and for any complex number c. There is one special type of operator on Hilbert space that is particularly important for quantum mechanics. Let W be a fixed subspace of $\mathcal{H}$. Then for any vector $v \in V$, there is a unique closest vector $E_W\ v$ to v in W. (That is, $\|E_W v - v\| \leq \|w - v\|$ for all $w \in W$.) The vector $E_W\ v$ is called the projection of v onto W; and the mapping that takes v to $E_W\ v$ is called the projection operator onto W.

With this terminology in place, the "statistical algorithm" of quantum mechanics can be formulated as follows:

1. Assumption 1. Each quantum-mechanical system is associated with a Hilbert space $\mathcal{H}$, and its pure states are represented by rays (i.e., one-dimensional subspaces) in $\mathcal{H}$.
2. Assumption 2. Experimental yes-no questions (e.g., "Is the second particle located in the region Δ?") are represented by projection operators on $\mathcal{H}$.
3. Assumption 3. If the system is in state v, then the probability that an experimental question E will receive an affirmative answer is $\text{Prob}^v\ (E) = \|Ev\|^2$.

The projection operators that correspond to a single quantity are pairwise compatible—that is, their operator product is commutative—which is typically taken to mean that they can be measured simultaneously. On the other hand, projections corresponding to conjugate quantities (e.g., position and momentum) do not commute with each other.

Interpretations of Quantum Mechanics

Quantum mechanics and its descendant, quantum field theory, have yielded empirical predictions of unprecedented accuracy. However, quantum mechanics by itself does not supply any unconditional statements about how things are when no measurement is being performed. (As some philosophers have put it, quantum mechanics supplies only probabilities conditional upon certain measurements being performed.) Thus, quantum mechanics is analogous to an uninterpreted formal system, and an interpretation of quantum mechanics would supply a class of models that entail the conditional probabilities of quantum mechanics (Stein 1972).

It is often said that the primary obstacle to interpreting quantum mechanics is the superposition principle, which says that any two quantum states can be superposed to form another quantum state. In particular, if u and v are state vectors, then $w = c_1 u + c_2 v$ is a state vector (when $|c_1|^2 + |c_2|^2 = 1$), called a superposition of u and v. The state w shows different faces, depending on what measurements are performed. On the one hand, if E_u and E_v are measured, then w looks like an ignorance mixture of u and v. In particular, in a long run of experiments performed on systems prepared in w, a measurement of E_u will get a positive response in $(100 \times |c_1|^2)\%$ of the trials, and a measurement of E_v will get a positive response in $(100 \times |c_2|^2)\%$ of the trials. On the other hand, a system prepared in state w always gives a positive response to E_w, and this is a feature that no ignorance mixture of u and v can have. In particular, systems whose states are either u or v will occasionallly give a negative response to E_w, and so an ensemble of systems in states u and v is empirically discriminable from a system in state w.

The superposition principle is also the source of the nonlocality in quantum mechanics (see Locality). In particular, suppose that $v \otimes w$ is a "conjunctive" state of a pair of systems—that is, $v \otimes w$ is the state in which the first system is in state v and the second system is in state w. Suppose that x is another state of the first system, and y is another state of the second system. Then the superposition principle entails that $2^{-\frac{1}{2}}(v_1 \otimes w_1 + v_2 \otimes w_2)$ is a state of the composite system. Since, however, this state is not a simple conjunction of states of the component systems, it is called an "entangled" state.

According to the orthodox (Dirac–von Neumann) interpretation of quantum mechanics, the property represented by Ev is objectively indeterminate (i.e., neither possessed nor not possessed) when the system is in a state w such that $0 < \mathrm{Prob}^w(E_v) < 1$. More generally, the orthodox interpretation says that a system possesses a property E only if its state assigns probability 1 to E; and E is not possessed only if the state assigns probability 0 to E. (This pair of conditions is often called the eigenstate-eigenvalue link.) When neither of these conditions hold, E is objectively indeterminate.

However, this objective indeterminacy never manifests itself directly to observers. That is, in any single case, an observer who checks for the property E will find that it either is or is not possessed by the system. Thus, although E may initially be indeterminate, it becomes determinate whenever it is measured. In order to accommodate this fact, the orthodox interpretation claims that a measurement of E forces the quantum state to "collapse" onto a state that is either completely in the range or completely in the null space of E.

- Projection Postulate. If E is measured in state v, and the outcome is positive, then the final state is $\|E_v\|^{-1}\,E_v$. If the outcome is negative, then the final state is $\|v - E_v\|^{-1}(v - E_v)$.

(See Dirac 1958, 36; von Neumann 1932, Ch. 6). In the absence of a measurement interaction, the state of a quantum system changes deterministically, as dictated by the Schrödinger equation. However, when $0 < \mathrm{Prob}^v(E) < 1$, the projection postulate permits two possible postmeasurement quantum states. So, according to the orthodox interpretation, there are two types of processes in quantum mechanics: the deterministic process that occurs in the absence of measurement, and the indeterministic process that occurs as the result of a measurement (see Quantum Measurement Problem). Since quantum mechanics itself is sometimes (mistakenly) identified with its orthodox interpretation, it has been claimed that quantum mechanics is an inherently indeterministic theory (see Determinism).

The orthodox interpretation of quantum mechanics has been criticized for many reasons, but most severely for its claim that the standard rule for dynamical evolution is suspended in measurement processes. For one, there is no precise criterion for distinguishing "measurement" interactions from other physical interactions, and so invocations of the projection postulate have to rely on imprecise judgments (Bell 1990).

There are essentially three lines of response to the standard criticisms of the orthodox interpretation. First, some claim that measurements are in fact distinguished from ordinary interactions in that they involve a nonphysical object, *viz.,* the mind of a conscious being (Wigner 1962). However, this solution has been rightly criticized as conceding the incompleteness of the quantum-mechanical description of measurement. Second, some claim that the standard dynamical law of quantum mechanics—which is linear and deterministic—never holds exactly, but is just an approximation of the true dynamical law, which is stochastic and nonlinear (Shimony 1993, I.4). It is hoped that this new dynamical law would be able to explain the appearance of a determinate macroworld by showing that when systems similar to a human observer (e.g., large, heavy) interact with microsystems, then the former end up in states that can count as having registered a definite outcome. This idea has been taken quite seriously, and as a result, there are now a number of worked-out theories of dynamical wavefunction collapse (see e.g., Ghirardi, Rimini, and Weber 1986). However, one should be clear that these theories are replacements for, rather than interpretations of, quantum mechanics. In particular, these theories do not attempt to interpret or explain microscopic superpositions. Furthermore, while these theories are in principle empirically distinguishable from standard quantum mechanics, the experiments that could decide between them would be extremely difficult to carry out in practice.

Finally, the third response to the problems with the orthodox interpretation is to reject the projection postulate altogether. Thus, according to "no collapse" interpretations, quantum mechanics applies universally to all physical systems, and since measurements do have outcomes, the eigenstate-eigenvalue link must fail. There are essentially two types of no-collapse interpretation. On the one hand, some no-collapse interpretations claim that measurements do not have unique outcomes, appearances notwithstanding. For example, Everett's relative state interpretation (Barrett 2001) and Kochen's relational interpretation (Kochen 1985) both deny that a successful measurement has a single outcome. On the other hand, those no-collapse interpretations that maintain the uniqueness of observed measurement outcomes postulate "hidden variables" in order to account for these outcomes.

Hidden Variables

As noted previously, quantum mechanics predicts that there are measurements in which the state of the system does not uniquely fix an outcome. Thus, if the quantum-mechanical state provides a complete physical description, then there are some facts, *viz.,* that one measurement outcome obtains rather than another, that simply have no physical explanation.

Einstein maintained throughout his career that the quantum-mechanical description should be considered incomplete and that physicists should be searching for a more fundamental theory that will underwrite its statistical predictions. In contrast to Einstein, Bohr maintained that the quantum-mechanical description is complete. Einstein and Bohr had a long series of debates over this question, leading eventually to Einstein's famous proposal of the Einstein-Podolsky-Rosen thought experiment (Bohr 1949).

The question of whether quantum mechanics could be supplemented with hidden variables was given a mathematical formulation by von Neumann (1932). In particular, he asked whether it is mathematically possible to enlarge the space of quantum-mechanical states so as to include states (analogous to points in phase space) that determine the answers to all experimental questions. Of course, if nothing further is required of these hypothetical completed states than that they map the projection operators into $\{0, 1\}$, then it is trivially true that such completed states exist. However, in order for the completed states to have some explanatory power, the values they assign to operators should reflect the algebraic relations of these operators (since these commutation relations are supposed to have some physical significance). For example, for any pair of orthogonal projections, a state should assign 1 to at most one of the two.

Von Neumann formulated his requirement on hidden variables in terms of expectation values. He claimed that if A and B are operators (representing physical observables), then the hidden state should assign the operator $A + B$ the sum of the expectation values that it assigns to A and B. (Of course, the expectation values of quantum states satisfy this requirement.) He then proved a representation theorem that shows that any such state (at least in the pure case) is represented by a ray in Hilbert space. Thus, there are no "hidden" states. Von Neumann (1932) concluded that

> it is therefore not, as is often assumed, a question of a reinterpretation of quantum mechanics,—the present system of quantum mechanics would have to be objectively false in order that another description of the elementary process than the statistical one be possible. (325)

John S. Bell (1966) was the first to point out explicitly that von Neumann's linearity requirement may be too stringent. Nonetheless, Bell himself went on to show that von Neumann's result goes through—aside from the exceptional case where the state space is two-dimensional—even with a weakened version of the linearity assumption (the result was proved independently by Kochen and Specker [1967]). In particular, a hidden state is said to be partial-linear just in case it is linear with respect to simultaneously measurable observables. Bell then pointed out that it follows from Gleason's representation theorem, according to which all pure probability measures on the lattice of subspaces of a Hilbert space $\mathcal{H}$ are represented by rays in $\mathcal{H}$, that all partial-linear states are represented by rays in $\mathcal{H}$. Thus, if physical states must be represented by partial-linear functionals, then there are no hidden states.

Bell pointed out, however, that even the strengthened result has a loophole, because it assumes that hidden states are noncontextual—i.e., that propositions about systems have a truth value *simpliciter,* independent of contextual features (such as which measurements are being performed on distant systems). For example, suppose that E, F_1, and F_2 are projections such that E is jointly measurable with either F_1 or F_2, while F_1 and F_2 cannot be measured simultaneously. Now, the Bell-Kochen-Specker theorem assumes that a hidden state assigns a value (either 0 or 1) to E without any reference to whether F_1 or F_2 is measured along with E. But this can be denied; indeed, it is precisely by employing contextual value assignments that Bohm's hidden variable theory escapes the conclusion of the Bell-Kochen-Specker theorem.

The de Broglie–Bohm Theory

In his 1924 Ph.D. dissertation, Louis de Broglie proposed that for each type of physical particle (e.g., photon, electron), there is a corresponding type of wavefield that guides the particle's motion. In particular, de Broglie proposed that the equation $p = h/\lambda$ supplies the link between the momentum (p) of a particle and the wavelength (λ) of the corresponding wave. Although de Broglie's idea received some confirmation through the discovery of diffraction effects with electrons, it was soon squashed by the then-dominant orthodox interpretation of quantum mechanics. In 1952, David Bohm independently rediscovered de Broglie's idea and developed it into a full-blown (empirically equivalent) alternative to quantum mechanics. According to Bohm's theory, the quantum state ψ plays two roles: Its squared modulus $|\psi|^2$ gives a classical probability distribution over particle positions, and the gradient of its phase defines a field that guides the trajectories of individual particles. The resulting dynamics of individual processes is completely deterministic; the apparent randomness and indeterminism of quantum mechanics can be shown to follow from the impossibility of keeping track of the trajectories of individual particles.

The Bohmian probability distribution over particle positions reproduces the distribution that the quantum state assigns to the position observable Q. Thus, Bohm's theory can be thought of as the (unique) hidden variable theory, in which Q always has a definite value, regardless of the quantum state (see Bub and Clifton 1996). Moreover, Bohmians claim that operators that do not commute with Q do not represent genuine physical quantities, and purported measurements of these quantities can be interpreted as complicated position measurements (cf. Dürr, Goldstein, and Zanghì 1996).

Bohm's theory provides an explicit counterexample to any attempt to prove the impossibility of hidden variables. It also shows that quantum mechanics does not logically entail indeterminism. But critics are quick to point out that Bohm's theory has its own share of counterintuitive consequences. For one, it undermines the traditional understanding of space-time structure: there are no inertial trajectories of particles, and the nonlocal dependence of the values of hidden variables on measurement contexts violates the spirit, if not the letter, of special relativity. Nonetheless, some claim that quantum mechanics by itself is non-local, and so non-locality cannot be thought of as a peculiar drawback of Bohm's theory (Albert 1992; Maudlin 2002). But this position is by no means unanimous, and the debate over the locality issue continues (see Locality).

It has also been claimed that Bohm's theory is inconsistent with the principle of faithful measurement—that is, that the outcomes of measurements reflect the values that quantities possessed prior to measurement. (However, Bohmians will reply that this result is not surprising, since these unfaithful measurements are of pseudoquantities.) Finally, it has been claimed that Bohm's theory arbitrarily privileges position over momentum (which play symmetric roles in the canonical commutation relations). In fact, it has been claimed that one could formulate an alternative Bohmlike interpretation in which momentum is always definite (Stone 1994). However, Bohmians are likely to claim that there are good philosophical and physical reasons for privileging position over momentum (e.g., interaction terms in a system's Hamiltonian are typically functions of position and not of momentum).

Modal Interpretations: Dieks and van Fraassen

Let M be a measuring device, and let O be an object system. The state space of the composite system $M + O$ is represented by the tensor product $V_1 \otimes V_2$ of the state spaces of M and O (see Locality). A postmeasurement state v of $M + O$ will typically be entangled, in which case there is no property that O possesses with probability 1. Thus, the orthodox interpretation (which maintains the eigenstate-eigenvalue link) invokes the projection postulate to change the entangled state v into a product state.

Modal intepretations attempt to give a rule for picking out the definite properties of O without collapsing the postmeasurement state. The key move made by these interpretations is to make a distinction between the "dynamical state" of the composite $M + O$ (i.e., the state that changes deterministically, as if no special measurement process occurred), and the "value state" of the object system O (i.e., the state that determines the observed outcomes of a measurement of O). According to the *biorthogonal decomposition theorem*, for any state v of $M + O$, there is a set $\{y_1, \ldots, y_n\}$ of unit-length vectors in V_2, and coefficients λ_i (all between 0 and 1) such that

$$\mathrm{Prob}^{v}(E) = \sum_{i=1}^{n} \lambda_i \, \mathrm{Prob}^{yi}(E), \qquad (5)$$

for all experimental questions pertaining to the object system. So, ignoring the measuring apparatus M, the state v looks just like an ignorance mixture of $\{y_1, \ldots, y_n\}$. According to Dieks's modal interpretation, the value state of O is some y_i (which one obtains is not determined by the dynamical state), and in this case the real properties are those represented by projection operators that are compatible with E_{yi}. According to van Fraassen's modal interpretation, the value state of O can be any state x that is "possible relative to v" (i.e., not orthogonal to all of the y_i), and in this case the real properties of O are those compatible with E_x. In both cases, the set of real properties is constrained in some way by the quantum state and typically has different members at different times.

One argument put forward in favor of modal interpretations is that they solve the measurement problem by means of a "minimal" revision of standard quantum mechanics (without the projection postulate). However, it has also been argued that these interpretations do not provide a general solution to the measurement problem (Albert 1992, appendix). (For a summary of these debates, see Vermaas 1999, chap. 10.) In terms of being a minimal revision of standard quantum mechanics, it can be shown that the set of real properties in Dieks's modal interpretations is definable in terms of the quantum state alone (Clifton 1995; Vermaas 1999). Thus, modal interpretations do not need to invoke extraneous philosophical arguments for privileging certain properties over others—in contrast to Bohm's claim that position is privileged, and to Bohr's claim that the measured observable is privileged. However, modal interpretations have also been criticized for not providing a plausible dynamics for their hidden states. On the one hand, some modal interpreters (e.g., van Fraassen 1997) reply that a satisfactory interpretation of quantum mechanics does not need to supply dynamics for its hidden states. On the other hand, attempts to work out explicit dynamics for the hidden states have yielded some unpleasant results (Vermaas 1999; Dickson and Clifton 1998).

Bohr's Complementarity Interpretation

According to Niels Bohr's complementarity interpretation of quantum mechanics (see Complementarity), it is appropriate to speak about different quantities in different "measurement contexts." In particular, there are contexts in which it is appropriate to speak of an ensemble of particles with definite positions; and there are contexts in which it is appropriate to speak of an isolated system that is subject to dynamical conservation laws; but there is no single context in which it would be appropriate to employ both of these modes of description, which are complementary. Thus, to a first approximation, Bohr's interpretation can be seen as the (unique) nocollapse interpretation in which the measured observable R is always determinate (see Bub and Clifton 1996). Note, however, that in this interpretation, the set of real properties is not fixed by the quantum state (as in Dieks's modal interpretation), nor does it remain invariant (as in Bohm's interpretation). But what then determines which quantity it would be appropriate to speak about? Is the preferred determinate observable picked out by something physical (e.g., the Hamiltonian of the measurement interaction), or does it correspond merely to a perspective from which the object is viewed?

Other Interpretations

There are many other interpretations of quantum mechanics besides those previously mentioned.

Three of the most prominent are Everett's relative state interpretation (and its numerous philosophical spawn; see Barrett 2001), the consistent histories interpretation (see Griffiths 2001), and relational interpretations (see Mermin 1998; Rovelli 1996). The first two of these interpretations have been gaining ground recently among quantum cosmologists who argue that an observer-independent interpretation of quantum theory is needed in order to apply quantum mechanics to the universe as a whole.

Conclusion

Much of the landscape of contemporary philosophy of science has been shaped in some way or other by issues that arise in the foundations of quantum mechanics. Conversely, attitudes toward the interpretation of quantum mechanics reflect trends in the general philosophy of science. In particular, while many early attempts to interpret quantum mechanics (e.g., by members of Bohr's institute in Copenhagen) had operationalist overtones, more recent efforts have focused on developing "realistic" interpretations. Whether any of these realistic interpretations is adequate and plausible remains a matter of debate.

HANS HALVORSON

The author acknowledges the helpful input of John Norton, University of Pittsburgh.

References

Albert, David (1992), *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press.

Barrett, Jeff (2001), *The Quantum Mechanics of Minds and Worlds*. New York: Oxford University Press.

Bell, John S. (1966), "On the Problem of Hidden Variables in Quantum Mechanics," *Reviews of Modern Physics* 38: 447–452.

——— (1990), "Against 'Measurement'," *Physics World* 3: 33–40.

Bohr, Niels (1949), "Discussion with Einstein on Epistemological Problems in Atomic Physics," in P. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*. Evanston, IL: Open Court, 200–241.

Bub, Jeffrey, and Rob Clifton (1996), "A Uniqueness Theorem for 'No Collapse' Interpretations of Quantum Mechanics," *Studies in History and Philosophy of Modern Physics* 27: 181–219.

Clifton, Rob (1995), "Independently Motivating the Kochen-Dieks Modal Interpretation of Quantum Mechanics," *British Journal for the Philosophy of Science* 46: 33–57.

Dickson, Michael, and Rob Clifton (1998), "Lorentz Invariance in Modal Interpretations," in Dennis Dieks and Pieter Vermaas (eds.), *The Modal Interpretation of Quantum Mechanics* Boston: Kluwer, 9–47.

Dirac, P. A. M. (1958), *The Principles of Quantum Mechanics*, 4th ed. New York: Oxford University Press.

Dürr, Detlef, Sheldon Goldstein, and Nino Zanghì (1996), "Naive Realism about Operators," *Erkenntnis* 45: 379–397.

Ghirardi, G. C., A. Rimini, and T. Weber (1986), "Unified Dynamics for Microscopic and Macroscopic Systems," *Physical Review D* 34: 470.

Griffiths, R. (2001), *Consistent Quantum Theory*. New York: Cambridge University Press.

Kochen, Simon (1985), "A New Interpretation of Quantum Mechanics," in Pekka Lahti and Peter Mittelstaedt (eds.), *Symposium on the Foundations of Modern Physics*. Singapore: World Scientific, 151–170.

Kochen, Simon, and Ernest Specker (1967), "The Problem of Hidden Variables in Quantum Mechanics," *Journal of Mathematics and Mechanics* 17: 59–87.

Kuhn, Thomas (1978), *BlackBody Theory and the Quantum Discontinuity, 1894–1912*. New York: Oxford University Press.

Maudlin, Tim (2002), Quantum Nonlocality and Relativity, 2nd ed. Cambridge, MA: Blackwell.

Mermin, N. David. (1998), "What is Quantum Mechanics Trying to Tell Us?" *American Journal of Physics* 66: 753–767.

Rovelli, Carlo (1996), "Relational Quantum Mechanics," *International Journal of Theoretical Physics* 35: 1637–1678.

Shimony, Abner (1993), *Search for a Naturalistic World-view*, vols. I and II. New York: Cambridge University Press.

Sommerfeld, Arnold (1919), *Atombau und Spektrallinien*. Braunschweig: Vieweg.

Stein, Howard (1972), "On the Conceptual Structure of Quantum Mechanics," in R. G. Colodny (ed.), *Paradigms and Paradoxes: Philosophical Challenges of the Quantum Domain*. Pittsburgh, PA: University of Pittsburgh Press, 367–438.

Stone, Abraham (1994), "Does the Bohm Theory Solve the Measurement Problem?" *Philosophy of Science* 61: 250–266.

van Fraassen, Bas (1997), "Modal Interpretation of Repeated Measurement: Reply to Leeds and Healey," *Philosophy of Science* 64: 669–676.

Vermaas, Pieter (1999), *A Philosopher's Understanding of Quantum Mechanics*. New York: Cambridge University Press.

von Neumann, John (1932), *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer.

Wigner, Eugene (1962), "Remarks on the Mind–Body Question," in I. J. Good (ed.), *The Scientist Speculates*. New York: Basic Books, 284–302.

WILLARD VAN QUINE

(25 June 1908–25 December 2000)

Willard Van Quine was born in Akron, Ohio, and died in Boston, Massachusetts. He took an undergraduate degree in mathematics from Oberlin College in 1930. In 1932, he completed a Ph.D. at Harvard University with a dissertation in logic that generalized and simplified a portion of Whitehead and Russell's *Principia Mathematica.* From 1932–1933, traveling on a fellowship in Europe, Quine spent five months in Vienna, where he attended meetings of the Vienna Circle and met such notables as Schlick, Waismann, Gödel, Hahn, Reichenbach, and Ayer (see Vienna Circle). Six weeks in Prague brought the beginning of the famous personal and professional relationship between Quine and Rudolf Carnap (see Carnap, Rudolf). Quine then studied logic with Tarski, Leśniewski, and Łukasiewicz while in Warsaw for six weeks. In 1936, following three years as an inaugural Junior Fellow at Harvard, Quine took a faculty position at Harvard, teaching there (but for his service in the United States Navy during World War II) until his retirement in 1978. Quine published prolifically throughout his career until the year of his death.

Quine emerges from a tradition within analytic philosophy that has been called scientific philosophy. This tradition is characterized by a concern for the epistemology and ontology of science, logic, and mathematics; the exploitation of developments in logic and set theory; and an antipathy toward speculative metaphysics (Hylton 2001). In particular, Quine's work is best understood against the backdrop of Vienna Circle logical empiricism, especially the work of Carnap. Allowing for some necessary simplification, the logical empiricists were concerned to portray science as a unified system of knowledge, including not only logico-mathematical knowledge and the so-called hard sciences, but also psychology, sociology, and history (see Logical Empiricism). While the positivist conception of science was broader than typically portrayed, it is, of course, not the case that every claim of every discipline qualified as scientific. To so qualify, a claim or statement had to pass a test of cognitive significance by being either analytic (true solely in virtue of the meanings of constituent terms) or synthetic (empirically confirmable or disconfirmable) (see Cognitive Significance). Any claim that was neither analytic nor synthetic was considered cognitively meaningless, thus unscientific.

This conception of the analytic and the synthetic served a number of interrelated ends in the logical empiricist program. First, the claim that the truths of logic and mathematics are analytic provides an empirically respectable account of the supposed a priori status of logico-mathematical knowledge. The relevant claims are true in virtue of meaning alone, so no particular state of the world is relevant, and thus no appeal to observation is relevant. Yet neither is an appeal to special intuition or nonempirical realms required—understanding of the language is the key justifying component of such knowledge.

Second, the empiricist characterization of synthetic claims was central to providing an account of the unity of the a posteriori portion of science. In the early days it was thought that every synthetic claim would strictly reduce to (translate into) some claim in a basic observational language. This language would include vocabulary sufficient for logic, set theory, and some form of observational claim. The exact nature of the observational claims was much debated, even after strict reductionism had been abandoned. If feasible, this would show that all genuinely synthetic claims are ultimately about possible or actual observations. But strict verificationist reducibility of theoretical claims to observational claims is not to be had (see Reductionism; Verifiability). The required relation had to be loosened to some form of implication of observational claims by the theoretical (Carnap 1936–1937). In any case, the requirement that synthetic claims must be related to observable circumstances in ways to be made clear through logical analysis supports the notion of the unity of science (see Unity and Disunity of Science)—for every nonanalytic claim would bear the same (type of) relation to observational claims, and the process of confirmation would be fundamentally the same. Thus, with no fundamental epistemological or ontological distinctions made among synthetic claims, no fundamental distinctions of methodology or ontology were made among the sciences. Whether one is considering physics, psychology, sociology, or whatever else, to be scientific, the claims of the discipline had to be either analytic or synthetic.

Third, in this understanding of legitimate theorizing, much traditional philosophy was to be swept aside as unscientific. One might wield the requirement of cognitive significance like a scythe—cutting down any claim that failed to be either clearly analytic or clearly synthetic, thereby eliminating a host of metaphysical claims and problems. Or one might take a more considered, clinical approach, as Carnap did. Many traditional philosophical disputes (idealism versus realism, for example) were seen as pseudoproblems—situations in which what appear to be contradictory claims regarding matters of fact are, according to Carnap, more fruitfully viewed as disagreements over which language (linguistic framework) should be adopted. Since adoption of a language is logically prior to the process of meaningful inquiry, nothing decidable by inquiry (that is, no matter of fact) is at issue. Rather it is a question of which language to adopt for the purposes of inquiry, and competing proposals can be assessed only on pragmatic (and, so, for Carnap, nonfactual) grounds. Thus, what traditionally would be taken as a deep dispute requiring metaphysical inquiry is cast by Carnap as a question not of truth, but of methodological and linguistic efficacy. Provided a proponent is clear about the structure of the language, tolerance reigns when considering the very loosely constrained questions of how perspicuous, simple, and fruitful the proposed framework might eventually prove.

While this view substantially deflates the status of philosophy as queen of the sciences, it does not completely relegate her to the position of intellectual handmaiden. The logical empiricists maintained a role for philosophy in the use of logic, set theory, and mathematics to analyze, clarify, and simplify the ground, structure, and results of empirical theorizing. Since those disciplines were understood to be analytic, philosophy itself is understood to be analytic (or the pragmatic investigation of analytic frameworks) and is not expected to make synthetic claims, or produce knowledge—such is the job of (and only of) unified science. Rather, philosophy is an a priori discipline of linguistic and conceptual analysis, maintaining a status independent of and methodologically distinct from empirical science. Far from playing a passive or merely organizing part, however, the analytic work of philosophy was projected to play a significant role in the advance of knowledge by illuminating the epistemology of science and helping to diagnose, cure, and prevent outbreaks of pseudoproblems.

Quine emerges from this tradition and inherits its concerns, but in rejecting the analytic/synthetic distinction he radically transforms the manner in which they are addressed. The conception of analyticity was central to the logical empiricists' semantics, epistemology, and dismissal of metaphysics; moreover, it marked the frontier between science and what remained of philosophical inquiry. In place of the picture sketched above, Quine offers a holistic semantics and epistemology that allows for only a difference of degree (not type) between so-called analytic and synthetic sentences. All meaningful sentences (including those of logic and mathematics) have at least remote observational import, not when taken individually, but only insofar as they are part of a set of claims (up to the whole of science) having observational implications. Quine's rejection of analyticity and the holistic epistemology and semantics anchor his naturalism—a view of science and philosophy as fundamentally similar in subject and method, differing only in degree of contact with empirical considerations. Mathematics and logic are viewed, not as analytic a priori, but as central strands of ongoing theorizing, thus participating in the empirical content of the whole theory. Philosophy and science (and "common sense") are on a continuum. In a sense, philosophy becomes science—though, as will be shown, this is a misleading turn of phrase.

Although Quine's importance is little disputed, there is much disagreement over the success and exact import of his rejection of analyticity and transformation of logical positivism and philosophy. Critical views of Quine range from those who aim to reject semantic holism (Fodor and Lepore 1992) and/or defend some form of the analytic/synthetic distinction (Boghossian 1997; Grice and Strawson 1956; Katz 1966) to those who read Quine as the revolutionary who could not, himself, see the full implications of his break with tradition (Rorty 2001). These disputes will be important at various points in this essay.

Two Dogmas, Analyticity, and Philosophy

Two Dogmas of Empiricism (Quine [1951] 1980) is often looked to as the decisive moment in Quine's rejection of analyticity, and, indeed, as a (if not *the*) decisive moment in the development of twentieth-century analytic philosophy. Rorty (2001), with typical enthusiasm, hails it as the most important article of the century and writes that it "rocked the audience back on its heels." Such folkloric status is enhanced by its appearance right at the midpoint of the century—Quine presented the paper in December 1950 to the American Philosophical Association in Toronto, and it was published in *Philosophical Review* in January 1951—dividing the calendar perfectly. The date of Quine's death, moreover, nearly perfectly marks the silver anniversary of *Two Dogmas.* Yet, while it does contain the famous arguments against analyticity and the striking initial pronouncement of Quine's holism, and has been discussed and translated perhaps more than any other English-language article in philosophy, *Two Dogmas* must not be considered in isolation from its surrounding works. It constitutes only a part of Quine's attack on analyticity, and, in truth, contains only a sketchy statement of his metaphysical and epistemological views. Thus, a discussion of the article can be a starting point, but by no means an endpoint.

Quine opens *Two Dogmas* by proposing to examine the notion of analyticity (that certain truths are true in virtue of meaning and independently of fact) and the notion of reductionism (that each meaningful sentence is equivalent to some claim in an observational language). The ensuing criticisms strike at the center of the logical empiricist conceptions of science, the a priori, and philosophy.

Quine divides supposedly analytic truths into two classes: the logical truths and those that can be transformed into logical truths by appropriate substitution of synonyms (whether this classification exhausts the supposed analytic truths has been questioned; see, e.g., Boghossian 1997; Katz 1966). Quine proposes initially to take the first class for granted and focus on the second class of analytic sentences. Thus, *Two Dogmas* has very little explicit discussion of logical truths, even though his criticisms of the analyticity of logic are more fundamental than much of what goes on in *Two Dogmas*—this issue will resurface in greater detail further on. Since Quine focuses on statements that can supposedly be transformed into logical truths by appropriate substitution of synonyms, the initial problem is to gain a relevant understanding of synonymy, or sameness of meaning.

Definition is surveyed and rejected as helpful in explicating synonymy—for, Quine argues, definitions either depend on preexisting synonymies, thereby failing to explain them generally, or are explicit introductions of notational variants, again failing to explain the synonymy relation generally. Next, Quine considers the condition of interchangeability *salva veritate:* Two terms are synonymous if they can be interchanged in all contexts without change of truth value. The problem Quine finds is that in order to secure a relation stronger than mere coextension, one must either include a necessity operator in the language or modify the interchangeability requirement from preservation of truth value to preservation of analyticity. The latter is a nonstarter, as analyticity is what wants explanation. The former, though less obviously, is equally a nonstarter according to Quine. The only way he sees to make sense of a necessity operator is essentially to presuppose an understanding of analyticity; thus, again, one must presuppose what wants explaining. Quine concludes that explaining analyticity by way of synonymy fails.

Quine next considers an attempt to define analyticity directly, at least for artificial languages, via semantical rules. His complaint here is that while there are various ways of distinguishing a subset of the truths of some artificial language *L* and labeling them "analytic for *L*," this provides no understanding of what "analytic" means generally, for there is no indication of how this would generalize across languages ("*S* is analytic for *L*" with variable *S* and *L*), nor is there any indication of how the specific notion of analytic for *L* relates to the notion of analyticity for natural languages. Even if there were a specification of "analytic for *L*" that captured intuitions concerning natural language analyticities, no clarity would be gained, for the attempt to explain the natural language case was abandoned in hopes that an appeal to artificial languages would be more illuminating (though more on this below). As an alternate approach, the analytic truths of *L* might be specified by appeal to the semantical rules of *L* (for analyticity is supposed to have something to do with meaning relations). Then "*S* is analytic for *L*" (for variable *S* and *L*) becomes "*S* is true in virtue of the semantical rules for *L*." But, of course, "semantical rules for *L*" wants explaining in a general way, for any recursive specification of a set of truths of *L* could be labeled as semantical rules. Again, the proposal gives no way of identifying what the rules or analytic truths of one language supposedly have in common with those of other languages, no way of explaining a general notion of analyticity. Quine (1980, 37) concludes that belief in analyticity is an "unempirical dogma."

Quine then discusses reductionist verificationism. If, as the reductionist view claims, each meaningful statement could be translated into some statement in a logico-observational language, then there would be an eminently clear criterion of statement synonymy (translating into the same observational claim), from which a criterion of term synonymy could be derived. This understanding of synonymy would then yield an understanding of analyticity. This strict form of semantic reductionism had already been discredited by the date of *Two Dogmas,* but Quine claims that the notion that a sentence has a specifiable content independently of other sentences still survives in the doctrine of analyticity. That doctrine encourages the idea that each sentence has a clearly specifiable content, while the idea of specifiable sentential content (left over from strict reductionism) encourages the idea that some sentences lack empirical content, that what content they have is not at all empirical but purely a question of meaning relations. Thus, discrediting the notion of specifiable sentential content discredits the notion of analyticity, for Quine sees the two as inextricably linked. Moreover, by underlining the failure of strict reductionism and extending its moral to the then current doctrines concerning the empirical content of supposedly synthetic claims, Quine is criticizing both sides of the logical empiricists' conception of the analytic/synthetic distinction.

In place of the notion of specifiable sentential content, Quine offers an early version of his semantic and epistemic holism. There is a strong remnant of logical empiricist verificationism here—Quine (1980) continues to countenance the notion of empirical content, but not of sentences taken individually: "The unit of empirical significance is the whole of science" (42). Following Duhem and Neurath (see Neurath, Otto; Duhem Thesis), Quine emphasizes the holistic nature of theory. Since no hypothesis has observational implications independently of a host of auxiliary hypotheses, there is, for Quine, no sense in which any theoretical claim is meaningful independently of the theory in which it is embedded. Moreover, since only a conjunction of hypotheses has observational implications, a failed prediction falsifies not a specific hypothesis, but a conjunction of hypotheses. Where the theory should be modified in order to defuse the implication and maintain consistency is underdetermined by the evidence. The falsification determines only that one or more of the conjuncts must be rejected or changed, but nothing determines which. On the basis of this underdetermination Quine claims that in the face of failed prediction, any sentence may, in principle, be maintained by making the necessary adjustments elsewhere in the theory. Conversely, any sentence may be revised, again, so long as the concomitant adjustments are made elsewhere. Quine even countenances the possibility of rejecting logical or mathematical laws in order to defuse the inference. All that is necessary, initially, is to block the inference leading to the false predictions. If rejection of a law of logic or mathematical claim will defuse the inference, then such an avenue is open. Since logic alone cannot determine how a theory must be revised, Quine claimed that pragmatic considerations (including conservatism and simplicity) figure into the choices made. Again, this follows Neurath's emphasis on the role of pragmatic concerns in theorizing. Neurath's conception of those concerns, however, was much broader than Quine's—for Neurath included social, economic, and political issues among the relevant considerations (Neurath 1983).

This latter claim of radical revisability is often taken as a further argument against the analytic/synthetic distinction, especially given the linkage of analyticity to apriority. For, if analyticity and apriority coincide (as the logical empiricists would have it) and if a priori claims are unrevisable (as is, perhaps, intuitive), then radical revisability would imply that there are no a priori truths, and so no analytic truths. Such a reading is encouraged by the opening of a famous paragraph of *Two Dogmas.* Quine (1980) has been discussing holism:

> If this view is right, it is misleading to speak of the empirical content of an individual statement—especially if it is a statement at all remote from the periphery of the field. *Furthermore it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold come what may.* Any statement can be held true come what may, if we wish to make drastic enough adjustments elsewhere in the system. Even a statement close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. Conversely, by the same token, no statement is immune to revision. Revision even of the logical law of excluded middle has been proposed as a means of simplifying quantum mechanics; *and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?* (43) (emphasis added)

The first italicized portion suggests that Quine is appealing to mere revisability in his rejection of the analytic/synthetic distinction. But this common interpretation fails to account for the final lines of the paragraph and to take into account the views of Carnap, the main target of Quine's criticisms. Unrevisability was no part of Carnap's view of analyticity. Indeed, a central pillar of Carnap's view was that competing analytic frameworks could be chosen or revised based on pragmatic considerations and that such changes are in principle different from changes made in the synthetic portions of theory. This is the heart of Carnap's deflation of metaphysics and the notion of the constitutive a priori. If Quine is appealing only to the revisability of supposed analytic claims, then this criticism flies wide of Carnap's conception. Thus, there must be more going on in the preceding paragraph, and it occurs in the final lines. Quine is appealing not just to revisability, but to there being no principled difference between the revision of supposedly analytic claims (e.g., logical laws) and supposedly synthetic claims (e.g., that planets move only in perfect circles).

The point for Quine is that any revision made to the overall theory is supposed to improve its fit with sense experience while maintaining as much simplicity and usability as possible. This is just what a theory is for Quine—a linguistic construct facilitating interaction with—and understanding of—the world, constrained by predictive test and pragmatic considerations of simplicity and efficacy. There is no difference of type in the considerations that might lead to the revision of a law of logic and those that might lead to the revision of a so-called synthetic claim. Rather, there are only differences of degree—a difference in how directly linked to observations a claim is, and a difference in the amount of readjustment a revision would require in the rest of the theory. The natural (and pragmatic) tendency toward conservatism and simplicity inclines theoreticians away from revising logic and mathematics and toward revising claims more closely linked to observation. Given this conservatism, revision of the more fundamental portions of theory (mathematics, logic, ontology), though always an option, will be considered only when either prediction meets with gross and extended failure in some domain, or some important and widespread gain of theoretical simplicity is in the offing (or both). Despite this difference in degree, and despite the difference in intellectual focus it occasions when revising theory, every decision to revise or accept a theory is a question not of this or that specific hypothesis, but of the whole theory. And any such decision is constrained by a combination of pragmatic and empirical considerations. Hence, for Quine, there is no difference of the kind Carnap conceived—no distinction between kinds of revision, between the analytic and the synthetic, between the purely pragmatic and the fully factual. Thus, setting aside some qualifications to be addressed below, Quine's holism and radical revisability do generate a further argument against analyticity. But it is the lack of a principled difference in kind of revision, not mere revisability itself, which is operative.

Finally, having rejected reductionism in favor of holism regarding empirical content, there is no need for analyticity as a special explanation for the supposedly nonempirical, a priori claims of logic and mathematics. Despite being linked to observation only remotely, logic and mathematics participate in the empirical content of the whole system, for they are ubiquitous in and essential to the inference of observational consequences from sets of hypotheses. The apparent unrevisability, apriority, and necessity of logic and mathematics are explained via the unwillingness to revise such central strands of theory and the usual availability of simpler revisions.

Note that despite how Quine opens *Two Dogmas,* holistic considerations do address the analyticity of logic (the first class of analytic claims). Indeed, the arguments against the semantic rule conception also address the analyticity of logic, though this may not be entirely transparent while reading *Two Dogmas.* But the full-fledged attack on the analyticity of logic and math occurs in *Truth by Convention* (Quine [1936] 1976a) and *Carnap and Logical Truth* (Quine [1960] 1976b).

It is in these pieces straddling *Two Dogmas* that some heavy work is done in attempting to dismantle Carnap's conception of analyticity. Indeed, it has been argued (most recently by O'Grady [1999] and George [2000], but see also Gregory [2003]) that *Two Dogmas* actually does very little to damage Carnap's conception of analyticity and the use to which he puts it. The main problem is that Carnap was not, ultimately, interested in explaining and grounding an analyticity distinction applicable to natural language, nor was he interested in a general definition of analyticity applicable across artificial languages. As Carnap ([1952] 1990) notes, all he himself is after is a formally articulated distinction within an artificial language that, to some extent, though by no means perfectly, captures intuitions regarding analyticity and that, more importantly, clearly delineates framework commitments from theoretical commitments. This is consonant with Carnap's overall metaphysical deflationism and his view of philosophy as a discipline of linguistic analysis aimed at clarifying and examining analytic frameworks. In order to contribute to the advance of knowledge by helping to diagnose, cure, and prevent outbreaks of pseudo-problems, a concept of analyticity may not need to be grounded in natural language or generalizable across artificial languages. All that "analytic for L_0" need do, it seems, is adequately formalize and identify those sentences taken to be most fundamental to whatever theoretical conception is under examination. If "analytic for L_1" does the same for an alternate theoretical conception, then sufficient clarity has been gained for pragmatic issues of framework choice to be considered. If this view of Carnap's program is correct, then Quine's criticisms of analyticity for artificial languages appear simply to miss their mark. For the upshot of Quine's criticisms is that analyticity for artificial languages fails either to be nonarbitrary and generalize across languages or to capture and explain the concept of analyticity for natural languages. But on the above account, neither full generality nor explanation of natural language analyticity is required.

The situation is not so simple, however—for, along with the explication and comparison of competing theoretical frameworks, Carnap wanted to hold a deflationary stance toward the choice of analytic frameworks. As noted, such choice was understood as governed by purely pragmatic considerations such that framework decisions carry no genuine metaphysical import. Thus, the analyticity distinction is crucial to Carnap's antimetaphysical program and to the conception of philosophy as unique in its method of analysis. But it is not entirely clear that these aspects of the view evade the conjunction of the *Two Dogmas* criticisms with those extracted from *Truth by Convention* and *Carnap and Logical Truth.* In those works Quine argues (among other things) that the method of legislative postulation (as it is called in the later article), while promising to establish certain sets of sentences as true by convention or analytic, can easily be extended beyond logic and mathematics, beyond what is taken to be fundamental to a theory proposal, to include even empirical truths—indeed, every supposed truth of the theory in question. There is no principled stopping point to the legislative postulation of truths. Thus, there is no principled stopping point (on this account) to the circumscription of analytic truths. That is, if one wished, one could define the whole of a theory as analytic (this point is implicit in the *Two Dogmas* criticism of semantical rules). This is no problem for Carnap's explicative aims. One can still delineate competing theories, restricting the postulation of truths to those sets of sentences the proponents of a theory take to be most fundamental. This, surely, will facilitate understanding and pragmatic comparison of theories, thereby aiding the advance of science.

The lack of a principled analytic/synthetic distinction is, however, a problem for Carnap's metaphysical deflationism. The notion of analyticity was supposed to support the deflation of metaphysics by distinguishing sets of sentences whose acceptance is a matter of pure pragmatic decision from sets of sentences whose acceptance constitutes a judgment of truth. The former are the analytic sentences of the language, and, as their acceptance is supposed to be logically prior to all meaningful inquiry in that language, their acceptance cannot constitute a judgment of truth—rather, it is supposed to be a pragmatic decision regarding which tool to use. But if the exact delineation of a set of analytic sentences is constrained, not by some logical principle or any deep understanding of natural language synonymy, but by only what seems fundamental to the supporters of a particular proposal, then the metaphysical deflationism loses its force. This is because the distinction between pragmatic framework decision and synthetic judgment carrying metaphysical import is essentially arbitrary and can be varied at will. The whole of a theory might be defined as analytic, or none of it, or some proper portion.

If the whole is taken to be analytic, then any change in theory is supposed to count as a purely pragmatic framework decision, where such decisions are constrained by simplicity, coherence, facility of use, and the overall empirical fit of the theory. But, since the whole theory is analytic, no other kind of change can be made. So there is no distinction here between pure pragmatic decision and genuine judgment. If none of the theory is taken as analytic, then any change is supposed to count as a genuine judgment of truth. But, again, such changes will be constrained by overall simplicity, coherence, facility, and empirical fit. No real distinction is being made here either. If one of the many middle roads is taken, then nearly any proper portion of the theory may be taken as analytic. However, without some principled ground (in logic or natural language) for analyticity, all this amounts to is the assigning of provisional protected status to certain sets of sentences, such that revising *those* sentences is taken to be a more fundamental sort of revision than revising others. But, depending on current pragmatic and empirical concerns, including the intuitions of the theoreticians involved, exactly which portion of the theory is so protected can be varied at will. Such a view, Quine (1976a, b) argues, supports a distinction not of metaphysical status, but only of the theoreticians' current willingness to revise certain portions of the theory as opposed to others—a willingness that can evolve as theory, evidence, and pragmatic concerns evolve. If this is correct, then, for want of a principled analytic/synthetic distinction, Carnap's deflationary metaphysics collapses into a view nearly identical to that of section 6 of *Two Dogmas*—a view that accords equal metaphysical import to all truths of the theory, distinguishing them mainly by the theoreticians' willingness to revise (Gregory 2003).

In addition to attempting to undermine Carnap's deflationism, Quine's rejection of analyticity is supposed to result in an erasure of the logical empiricists' distinction between philosophy and science. Quine still has a conception of the unity of science, but now this includes mathematics, logic, and philosophy—these being understood not as analytic disciplines, but as empirically meaningful in virtue of their contribution to the whole theory. Metaphysical deflationism is rejected, but Carnap's explicative aims persist. Logico-mathematical analysis of theoretical proposals is still central to the practice of philosophy, but the judgments and decisions based on such analysis are not considered devoid of metaphysical import. Thus, Quine reinflates metaphysical inquiry, but only so long as the claims of such inquiry participate in the empirical content of the whole theory. Moreover, it is not only the philosopher who recognizably engages in such philosophical activity, it is open to any self-conscious theorizer. The main differences among the layperson, the scientist, and the philosopher are simply in the frequency and degree of sophistication with which that individual engages in abstract reflective analysis. Thus, while philosophy becomes science, science is recognized as having always been philosophical.

It is not easy to be clear on what this means. There are two natural, yet polarized, ways of misinterpreting the impact of Quine's rejection of analyticity and reinflation of metaphysics. On one side, there is the view that Quine (re)instates a certain liberalism regarding metaphysics and philosophy. On the other is the view that Quine has rejected philosophy altogether in favor of a rigid scientism. In the liberalist interpretation, Quine's rejection of logical empiricist constraints on inquiry reopens the door to traditional metaphysics; or, in conjunction with his stress on pragmatism, it opens a new door to a hypertrophic pragmatism in which inquiry is constrained only by social practice. In the scientistic interpretation, Quine's rejection of the boundary between science and philosophy and his insistence on empirical significance leave no room for philosophical inquiry—all is science, and science is all.

Both interpretations mistake what Quine took to be the nature of his own views. Against the liberalist interpretation, Quine consistently maintains the importance of observational constraints on inquiry and that even despite observational underdetermination of theory, these constraints can distinguish better from worse theories (1969, 1975, and 1998). Moreover, Quine is best understood as a form of metaphysical realist—at least in the internalist sense. For Quine claims that there is no standard transcending the best scientific methodology from which to make meaningful claims of anti-realism regarding ongoing theory (Quine 1981b, d; [1990] 1992). The scientistic interpretation is closer to being accurate but ignores both Quine's attitude toward current science and the way in which he is trying to reconceive philosophy. On the first count, the charge of scientism implies a blind faith in the methodology and deliverances of science or scientists, but Quine accounts for both the possibility of large-scale theoretical change (taking even foundational commitments as tentative), as well as small- and large-scale methodological change (see below). Quine (1981c, 22–23; [1990] 1992, 20–21) even countenances the possibility of rejecting physicalism and empiricism (though not intersubjective testability). On the second count, to view Quine as plumping for science and dismissing philosophy is to maintain a simplistic distinction between the two, such that one must, by embracing science, be rejecting philosophy. But, as the discussion of analyticity begins to reveal, Quine sees no fundamental distinction between the two, and not simply because what is best or acceptable in philosophy is what is scientific. Rather, it is because scientists and philosophers alike speculate and theorize about the world in an attempt to understand it; and any such theorizing is constrained by holistic empirico-pragmatic concerns. The more closely tied to observation inquiry is, the more scientific it is; conversely, the more remote from observation, the more philosophical—it matters not what academic department one reports to. Philosophy does not disappear; indeed it is understood as ubiquitous. The grain of truth in the scientistic interpretation is that Quine was thoroughly (though in principle tentatively) committed to the findings and methodology of science. Moreover, Quine was deeply committed to the notion that as one's distance from intersubjective checkpoints increases, so does one's risk of moving beyond science or philosophy, into fantasy or gibberish. The scientistic interpretation fails, however, to recognize that Quine, via philosophical analysis and argument, maintained a unique caution and skepticism regarding science, based in large part on his (very philosophical) recognition of how tenuous is the connection between intersubjective checkpoints and the vast theoretical structures erected upon them.

Observation, Theory, and Naturalized Epistemology

At the most general level of description, Quine has a hypothetico-deductive model of science. Hypotheses are generated and sets of them tested by the observational predictions deducible from those sets. When prediction is successful, then so far so good—confidence should always be tentative. When prediction is not successful, then new or revised hypotheses are called for. Given Quine's naturalism, understanding the details of hypothesis generation and testing is a task for science itself. Many subdisciplines will be relevant—physics, neurology, psychology, evolutionary biology, linguistics, history of science, etc.—especially on the generative side of the tale. But Quine ([1990] 1992, 2) believed he had "by means of little more than logical analysis" shed significant light on the structure of prediction and testing. The observation sentence is central to this analysis.

Observation sentences are supposed, in some sense, to be those sentences most closely associated with concurrent sensory stimulation and on which members of a language community will largely agree when presented with the same stimulus situation. To make this more precise, three criteria pick out observation sentences relative to a community of speakers. First, observation sentences are occasion sentences. That is, they are true at some times and places, and false at others (e.g., "There's a dog"). This is in contrast to standing sentences, which are true always or false always (e.g., "Electrons carry negative unit charge"). Second, a sentence is observational for an individual speaker if that speaker responds affirmatively (at the time of stimulation) for some range of stimulations of the speaker's sensory receptors, and negatively for some other range (there may also be a range in which the speaker is noncommittal). Stimulation of a subject on a given occasion is understood as "the temporally ordered set of all those of his exteroceptors that are triggered on that occasion" (Quine [1990] 1992, §2). Given this definition of stimulation, two subjects cannot share the same stimulation unless they share nerve endings. Hence, a careful way of stating the component of communitywide agreement is needed. So, third, a sentence is an observation sentence for the community if it is observational for each member individually and if community members would agree in their verdicts upon witnessing the same (or a similar) occasion of utterance (Quine [1990] 1992, §§2, 15–16). Thus, an abbreviated definition might run: An observation sentence is an occasion sentence that commands assent or dissent outright upon query in a given stimulus situation, and this pattern of assent and dissent is consistent across a community.

Observation sentences have a dual semantical and epistemological importance for Quine. Semantically, they are both the locus of empirical content and the first rung on the ladder of language acquisition. Epistemologically, they are the intersubjective checkpoints of science. The very same intersubjectivity of utterance and prompting occasion that normalizes usage and affords a way into language for the neophyte also allows for the testing of sets of hypotheses. Since hypotheses consist mainly of standing generalizations, they do not imply individual observation sentences (particular occasion sentences). Rather, sets of hypotheses imply observation categoricals. These are generalized conditionals of observation sentences (e.g., "Whenever there's an apple, then it's red"). Indeed, observation categoricals are a sort of minimal hypothesis, expressing generalized or habituated expectation. Unlike observation sentences, they are testable—one instantiates the antecedent and checks to see if the consequent obtains. If it does, then so far, so good. If not, then the conjunction of implying hypotheses is falsified, and revision is called for. Outside of his definition of observation sentences and categoricals, Quine takes a rather straightforwardly Popperian and Humean line on the logic of testing (see Popper, Karl Raimund). The testing of sets of hypotheses via the testing of observation categoricals they imply can, strictly speaking, only refute the conjunction of hypotheses. The continued success of predictions embodied in implied observation categoricals reinforces the habit of reliance on and confidence in the categoricals and their implying hypotheses (Quine 1981a, 28; [1990] 1992, §§5–6).

It is important to avoid some misunderstandings regarding observation sentences. Though Quine saw them as playing the role classical empiricists had wanted of sensory evidence, they are not Humean impressions, nor Russellian sense data. Nor, in contrast to certain logical empiricist conceptions of protocol sentences, are they reports of sensory phenomena. They are occasion sentences so strongly associated with ranges of stimulation that utterance or assent/dissent is practically immediate. Such immediacy is supposed to minimize, though by no means eliminate, reliance on learned theory. Viewed as undifferentiated wholes (holophrastically), observation sentences are nontheoretical responses to stimuli. This is part of what allows the novice to acquire a language. But observation sentences are not simply undifferentiated wholes. They contain terms that appear in more theoretical sentences, and it is in virtue of these shared terms that observation categoricals are implied by hypotheses. This dual nature has a number of implications. First, though when taken holophrastically they appear theory neutral, observation sentences are theory laden in virtue of the inferential connections to and terms shared with theoretical sentences. Second, in addition to direct conditioning, observation sentences may be learned via description and inference. Third, observationality of a sentence is relative to the community specified. What counts as observational for one community ("That's a red giant," "That's a middle C"), in virtue of the members' spontaneity of judgment in a stimulus situation, may not count so for a broader community. At any given time, however, the more specialized speakers could instruct those less specialized, in part by reverting to observation sentences common to both. It should also be stressed that observation sentences are not incorrigible. Assent to the utterance of an observation sentence may be rescinded either in the face of further observation or as the result of theoretical considerations. Thus, while observation sentences play a fundamental role in the testing of hypotheses, they are not a form of sensory given or simples forming an incorrigible foundation for knowledge.

Thus, logical analysis and some armchair psychology yield the prediction and testing side of the story, at least in outline. The generative side of the story, however, is highly unconstrained by logic and observation. Understanding the generation of expectations, projections, and hypotheses thus requires more than just logical analysis.

On the one hand, theoretical claims cannot be deduced from observations because there is no logic of ampliative inference, nor do logical constraints determine how to revise in the face of failed prediction. On the other hand, the rejection of analyticity, apriority, and reductionism involve repudiating both the Cartesian goal of externally justifying scientific methods and the Carnapian goal of rationally reconstructing the logico-empirical structure of science (Quine 1969). The natural sciences themselves are to be used to "address the question how we, physical denizens of the physical world, can have projected our scientific theory of that whole world from our meager contacts with it" (Quine 1995, 16). Thus, it is not that ampliative inference is devoid of all system or structure. Rather, what system or structure can be imputed to it will be largely extralogical—a matter of the evolution of innate similarity standards in the species, in the cognitive development of individuals, and in the community's ongoing theorizing. In the earliest pronouncements, such as "Epistemology Naturalized," Quine (1969) focuses especially on psychology as the science wherein this project is to be pursued. In later writings, such as *Pursuit of Truth* (Quine [1990] 1992), neuroscience, evolutionary genetics, and the history of science are included. The epistemologist is to investigate the complex and various ways humans actually do arrive at theories, including neurological, psychological, sociological, and historical factors.

Interestingly, despite arguing in favor of the naturalization of epistemology, Quine engaged in no hands-on investigation of the sort he urged. Perhaps this was a consequence of his professed dislike of laboratory work and enjoyment of popular science literature (see Hahn and Schilpp 1998, 5, 43; Quine 1985, 37). He did, however frequently theorize on how to gather behavioral evidence of a subject's similarity standards and their evolution, including, of course, the development of language (see e.g., Quine 1974 and 1981c). Foley (1994) takes this as a sign that Quine was not doing epistemology in any new way, but unless one accepts a naive distinction between philosophy and science, it should not be expected that all naturalistic epistemologists be lab rats as opposed to abstract theorizers.

Three interrelated objections have typically been raised regarding Quine's naturalized epistemology—the circularity objection, the normativity objection, and the change-of-subject objection. As the responses to these objections are also interrelated, it is worth considering them en masse. If epistemologists are to engage in a scientific study of science, then it seems that the results must be circular, thereby vitiating the normative/justificatory project of epistemology (the circularity objection). Given the circularity issue, and the fact that science is a purely descriptive endeavor, no scientific epistemology could ever be a normative epistemology, but the normative aspect is a crucial part of any philosophical epistemology (the normativity objection). While the epistemology Quine advocates may be of interest to psychologists, it is not a properly philosophical epistemology, since, being circular and purely descriptive, it fails to address the fundamental normative questions of traditional epistemology—Quine is simply changing the subject. These will be addressed in reverse order.

It is correct that in some sense Quine is attempting to change the subject, to motivate a departure from or reconception of traditional epistemology, but it is naive to conclude that naturalized epistemology is no longer philosophical, losing all contact with traditional issues. This maintains the simplistic dichotomy between science and philosophy that Quine was repudiating, and it seems to treat traditional formulations of questions as somehow sacrosanct. Quine's importance, and that of his naturalism, rests in the attempt to reconfigure the field of inquiry in a philosophically and scientifically fruitful manner. Hence the change-of-subject objection has the air of mere dismissal, as opposed to critical engagement.

The normativity objection presupposes that science is purely descriptive, that all a scientific investigation of belief and theory formation can do is list the various and sundry things that go on. But scientific theory and practice have significant normative dimensions, as evidenced by idealizations used in theory development and testing and the normative role of theory in engineering. Moreover, these norms are applied to scientific practice itself, allowing differentiation of practices according to their measure along different parameters. As these measures of instrumental efficacy are theoretical claims, they are fallible and open to revision. Quine cited predictive success as the ultimate parameter. So, naturalized epistemology is the assessment of the instrumental value of cognitive and social practices toward the goal of predictive success. Quine supports this through his analysis of the structure of theory and evidence, which reveals that prediction of intersubjectively available checkpoints is the fundamental norm of science. Since, as science reveals, information comes through the five senses, success in sensory prediction is the "final arbiter." As mentioned during the discussion of his alleged scientism, Quine ([1990] 1992, 20–21) recognized the possibility of admitting sources of information and testing other than the senses (extrasensory perception, revelation), were this ever warranted. He stopped short of speculating about giving up on intersubjective predictive tests altogether.

One further point bears mentioning. Clearly, to say that predictive success is the end against which methods are assessed is not to say that predictive success is the goal of science or cognition. It is likely one of the goals, but truth, understanding, and aesthetic enjoyment are surely others. Finding value in various practices, even if only distantly or loosely connected to predictive success, is entirely consistent with taking predictive success as the test parameter.

The obvious circularity of this approach is not a problem from Quine's point of view. The rejection of analyticity and apriority and the view of common sense, science, and philosophy as continuous imply that inquiry (epistemological or otherwise) cannot begin from a position independent of all theory. Hence any such inquiry is ultimately circular. But this is not to be understood as defeatist resignation. It is not that the demand for an independent justification of science and its methods is well grounded, but, alas, it cannot be met. Rather, the point is that such a demand is itself based in a misconception of the nature of theory and language—a misconception of epistemology. Quine tried to offer a better conception. There is a further worry that because natural epistemology begins from within ongoing scientific theory, it is doomed to reinforce the norms already at work in science. But this, while not impossible, is as unlikely as the possibility that no new theories will be developed, because all inquiry begins from within ongoing theory. Theory changes in the light of new evidence and new understandings of old evidence and theory. Methodological norms are fallible and may change with developing theory. It is no more likely that natural epistemology will become stuck in a loop of blind stagnation than that science in general will. For Quine, the epistemology of science is on a par with science itself (Gregory 1999; Quine 1969 and [1990] 1992).

Conclusion

It is worth illuminating points of contact with a few other philosophies of science. There are, of course, the connections to and departures from logical empiricism. The stress on falsification links Quine with Popper, as noted above, though Quine never maintained a sharp criterion of demarcation between science and pseudoscience. Quine's recognition of a lack of purely logical constraints on theory change, his assertion of the theoretical nature of normative constraints, his holism, and his stress on conservatism while recognizing the possibility of fundamental revisions—all these suggest that theory development will usually be a rather mundane affair but that given extended and serious failure in some domain, dramatic and very loosely constrained change will occur. This is all perfectly consonant with Kuhn's account of normal science and paradigm change (see Kuhn, Thomas). Of course, Quine's view idealizes theories as formal linguistic structures and conceives of theory change as modification of such structures. Kuhn's view of theory change is much richer, paying detailed attention to sociological, technological, and practical issues. Moreover, in Quine's view, revolutionary change is more highly constrained (mainly by predictive test) than in the typical reading of Kuhn. Finally, views of philosophers such as Lakatos (1977), who distinguish portions of theory that are less apt to be revised and are revised only under special circumstances, are less in tension with Quine's views than it might appear. Quine's views can countenance distinctions among sentences of more or less protected status, and naturalized epistemology is supposed to illuminate the nature of those distinctions. Tensions arise only insofar as such distinctions are taken to ground metaphysical and epistemological conclusions antithetical to Quine's naturalism. This is not to deny the presence of tensions between Quine's views and others' or to deny the possible value in rejecting Quine's naturalism. But it is important to note that while Quine is notorious for rejecting or blurring conceptual boundaries, his views can countenance certain kinds of distinctions.

Quine's criticisms of analyticity heralded the waning of logical empiricism. His holistic naturalism offers a unique view of philosophy as progressive, metaphysically committed, and continuous with science. His articulation of a version of naturalized epistemology was one of the major impetuses for the development, in philosophy, of naturalistic studies of science and cognition. Whether one accepts his naturalism in detail, in outline, or not at all, Quine's importance cannot be overestimated. Willard Van Quine shaped philosophy and philosophy of science in the second half of the twentieth century.

Paul A. Gregory

References

Boghossian, P. (1997), "Analyticity," in B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*. Oxford: Blackwell.

Carnap, R. (1936–7), "Testability and Meaning," *Philosophy of Science* 3 and 4: 419–471, 1–40.

——— ([1952] 1990), "Quine on Analyticity," in R. Creath (ed.), *Dear Carnap, Dear Van*. Berkeley and Los Angeles, CA: University of California Press, 427–432.

Fodor, J. A., and E. Lepore (1992), *Holism: A Shopper's Guide*. Oxford: Blackwell.

Foley, R. (1994), "Quine and Naturalized Epistemology," *Midwest Studies in Philosophy* XIX: 261–282.

George, A. (2000), "On Washing the Fur Without Wetting It: Quine, Carnap, and Analyticity," *Mind* 109: 1–24.

Gregory, P. A. (1999), *Language, Theory, and the Human Subject: Understanding Quine's Natural Epistemology*. Ph.D. thesis in Philosophy. University of Illinois at Chicago.

——— (2003), "'Two Dogmas'—All Bark and No Bite? Carnap and Quine on Analyticity," *Philosophy and Phenomenological Research* 61: 633–648.

Grice, H. P., and P. F. Strawson (1956), "In Defense of a Dogma," *The Philosophical Review* 65: 141–158.

Hahn, L. E., and P. A. Schilpp (eds.) (1998), *The Philosophy of W. V. Quine*. Expanded edition. LaSalle, IL: Open Court.

Hylton, P. (2001), "W. V. Quine," in A. P. Martinich and D. Sosa (eds.), *A Companion to Analytic Philosophy*. Oxford: Blackwell, 181–204.

Katz, J. J. (1966), *The Philosophy of Language*. New York: Harper & Row.

Lakatos, I. (1977), *Philosophical Papers* (Vol. 1). Cambridge: Cambridge University Press.

Neurath, O. (1983), *Philosophical Papers, 1913–1946*. Edited by R. S. Cohen and M. Neurath. Dordrecht, Holland: Reidel.

O'Grady, P. (1999), "Carnap and Two Dogmas of Empiricism," *Philosophy and Phenomenological Research* 59: 1015–1027.

Quine, W. V. (1969), "Epistemology Naturalized," in *Ontological Relativity and Other Essays*. New York: Columbia University Press, 69–90.

——— (1974), *The Roots of Reference*. LaSalle, IL: Open Court.

——— (1975), "The Nature of Natural Knowledge," in S. Guttenplan (ed.), *Mind and Language*. Oxford: Oxford University Press, 67–81.

——— ([1936] 1976a), "Truth by Convention," in *The Ways of Paradox and Other Essays*. Cambridge, MA: Harvard University Press, 77–106. Originally published in *Philosophical Essays for A. N. Whitehead*. New York: Longmans.

——— ([1960] 1976b), "Carnap and Logical Truth," in *The Ways of Paradox and Other Essays*. Cambridge, MA: Harvard University Press, 107–132. Originally published in *Synthese* 12.

——— ([1951] 1980), "Two Dogmas of Empiricism," in *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 20–46. Originally published in *Philosophical Review* 60.

——— (1981a), "Empirical Content," in *Theories and Things*. Cambridge, MA: Harvard University Press.

——— (1981b), "Reply to Stroud," *Midwest Studies in Philosophy* 6: 473–475.

——— (1981c), *Theories and Things*. Cambridge, MA: Harvard University Press.

——— (1981d), "Things and Their Place in Theories," in *Theories and Things*. Cambridge, MA: Harvard University Press.

——— (1985), *The Time of My Life*. Cambridge, MA: MIT Press.

——— ([1990] 1992), *Pursuit of Truth*. Revised edition. Cambridge, MA: Harvard University Press.

——— (1995), *From Stimulus to Science*. Cambridge, MA: Harvard University Press.

——— (1998), "Reply to Morton White," in E. Hahn and P. A. Schilpp (eds.), *The Philosophy of W. V. Quine*. Expanded edition. LaSalle, IL: Open Court.

Rorty, R. (2001, February 2), "An Imaginative Philosopher: The Legacy of W. V. Quine," *Chronicle of Higher Education* 47: 21. http://chronicle.com/free/v47/i21/21b00701.htm. (Aug 13, 2005)

See also **Analyticity; Carnap, Rudolf; Duhem Thesis; Logical Empiricism; Neurath, Otto; Popper, Karl Raimund; Reductionism; Verifiability; Vienna Circle**

R

FRANK PLUMPTON RAMSEY

(22 February 1903–19 January 1930)

Frank Plumpton Ramsey made important contributions to philosophy, economics, logic, and mathematics. The son of a Cambridge mathematician, from an early age he was well known to members of the Cambridge intellectual community. He received a degree in mathematics from Trinity College in 1923, became a Fellow of King's College in 1924, and in 1926 he was made a University Lecturer in Mathematics, the post he held until his untimely death in 1930. Ramsey was an original thinker and no one's disciple, but his work clearly shows the influence of Russell, Keynes, Wittgenstein, Moore, W. E. Johnson, and Peirce. (Biographical information on Ramsey can be found in Mellor 1995, Sahlin 1990, and the introductory sections of Ramsey 1931 and 1990.)

Ramsey's earliest publications include a criticism of Keynes' theory of probability (Ramsey [1922] 1989) and a critical notice on Wittgenstein's ([1922] 1961) *Tractatus.* From 1925 to 1927, he published the philosophical papers "The Foundations of Mathematics," "Universals," "Mathematical Logic," and "Facts and Propositions" (Ramsey 1927). In 1927–1928, he published two influential papers in economics, "A Contribution to the Theory of Taxation" and "A Mathematical Theory of Saving," and his single mathematical publication, in which was introduced what is now known as Ramsey's theorem, "On a Problem in Formal Logic." Shortly after his death, the collected works of Ramsey (1931) appeared, containing all of the aforementioned works except the economics papers and the discussion of Keynes. It also contains writings unpublished until then, including the paper "Truth and Probability," written in 1926, and papers dating from 1928–1929 such as "Theories" and "General Propositions and Causality." Similar collections of Ramsey's work have also been published (Ramsey 1978 [which has the economics papers] and 1990). Two more recent books (Ramsey 1991a, b), contain material from the manuscripts in the Ramsey Collection at the University of Pittsburgh's Archives of Scientific Philosophy. The most comprehensive treatment to date of Ramsey's philosophical work is by Sahlin (1990).

The significance of some of Ramsey's work—for example, his contributions to the foundations of mathematics—was quickly appreciated. But it took time for many of Ramsey's ideas to become widely known and appreciated (on this, see Ramsey 1990, xi–xxiii, and Mellor 1995). As one encounters those ideas, it is worth remembering that they were all produced before his twenty-seventh birthday, which he did not live to see. Some are expressed in papers or notes not meant for publication and not developed to a point that fully satisfied him. It is remarkable how much illumination has been found, and how much can still be found, by studying the work that Ramsey had time to produce.

Logic and *The Foundations of Mathematics*

Ramsey was well acquainted with Whitehead and Russell's (1910) *Principia Mathematica* (*PM*), and with Wittgenstein ([1922] 1961). He commented on and corrected the proofs of the second edition of *PM*, and he was the major contributor to the first English translation of the *Tractatus.* In his lengthy "The Foundations of Mathematics (FM)" and the subsequent paper "Mathematical Logic (ML)," Ramsey strongly argued that the system of *PM* needed serious revision in order to remain true to the project of capturing mathematics within logic (see Russell, Bertrand). He identified three crucial shortcomings of the system that, in his view, undermined both the legitimacy of its fundamental principles and the system's adequacy as a basis for mathematics. As Sullivan (1995) observes, there is a correspondence between the three fundamental problems that Ramsey found and the three problematic axioms of *PM* with which Russell struggled—the axioms of choice, reducibility, and infinity. The upshot of the solutions Ramsey offered was that the formal system of *PM* was drastically reinterpreted but largely preserved; changes were confined to a simplification of its theory of types and the elimination of one of its axioms (reducibility). In the course of doing this, Ramsey drew a distinction that has become a standard way of classifying the various paradoxes that Whitehead and Russell cataloged in *PM.* Only brief indications of Ramsey's criticisms and proposals can be given here. (For background and more thorough discussion, see Sullivan 1995 or Grattan-Guinness 2000; the latter contains an extensive bibliography of further sources.)

At the heart of Ramsey's revisions to the system of *PM* is a move toward extensionality, based on an understanding of logic strongly influenced by Wittgenstein. Ramsey followed the *Tractatus* in taking propositions to be truth functions of atomic propositions. He divorced propositions from the symbolic formulas that express them; a single proposition (truth function) may be expressed by many different formulas, and some propositions may not be expressible at all. The same is true for propositional functions. (In *PM*, propositional functions are, roughly, what yield propositions when all their occurrences of free variables are bound or replaced by names. In Ramsey's conception, they turn out to be functions from individuals to propositions.) Some truth functions with infinitely many arguments can be expressed by use of the quantifiers, which Ramsey understood as convenient ways of writing infinite conjunctions and disjunctions. Given infinitely many atomic propositions, however, many possible truth functions of the set of atomic propositions will not be directly expressible, though they are relevant to the truth or falsehood of universally and existentially quantified propositions.

In the system of *PM*, every class (set whose members all have the same type) is defined by a symbolically expressible propositional function, or defining property. Ramsey's *first criticism* was that this is too restrictive for mathematics, which at least leaves open the possibility of infinite classes not definable by the propositional functions of *PM.* So the system of *PM* misinterprets important mathematical assertions about some or all classes, including the axiom of choice (*PM*'s multiplicative axiom) (*FM,* §II). Given the restricted availability of classes in *PM*, this would be an empirical truth rather than a logical truth, if it is true at all. In Ramsey's reconstruction of *PM*, many classes were available beyond those definable by *PM*'s propositional functions, and he regarded the axiom of choice as an obvious tautology, a necessary truth.

Paradoxes and the Theory of Types

Ramsey's *second criticism* of *PM* is now the one best remembered. Whitehead and Russell had cataloged seven logical contradictions that the system of *PM* must avoid; these included the existence of Russell's class of all classes not members of themselves, the liar sentence, and Richard's construction of a decimal that both is and is not finitely definable. They attributed a common source to the contradictions—"a certain kind of vicious circle"—and they invoked a *vicious circle principle*: "Whatever involves *all* of a collection must not be one of the collection" (*PM,* introduction, Ch. 2). The system of *PM* adhered to this principle through its theory of ramified types.

Ramified type theory arranged propositional functions into a twofold hierarchy of orders and of types within individual orders. The *type* of a propositional function was determined by the members of its domain. Individuals, monadic functions of individuals, monadic functions of monadic functions of individuals, and so on, had distinct and increasing types, which might be labeled 0, 1, 2, and so on. (Types of relations, and of classes defined with them, were more complex.) A function was only meaningful when applied to entities of type one less than itself, and it was not meaningful to say that classes defined by such functions either contain or fail to contain themselves as members. The hierarchy of *orders* of propositional functions was generated by how the functions were defined—specifically, by what quantifications over functions (think of them as quantifications over properties) were used in their definitions. In keeping with the vicious circle principle, the idea was that the definition of a propositional function could not quantify over all functions, or even over all functions of its own type. This restriction generated functions of increasing order: Those whose definition involved no quantification over functions were order zero (Ramsey called these *elementary*), those whose definition involved quantification over elementary functions were first-order, those whose definition involved quantification over first-order functions were second-order, and so on. It was not meaningful to quantify over propositional functions of all orders, and Whitehead and Russell showed that the dual hierarchy of ramified type theory defused the threat posed by the contradictions to the system of *PM*.

But the complications that the hierarchy of orders brought to type theory presented a serious problem for *PM*'s adequacy as a foundation for mathematics. The example Ramsey emphasized most fully in *ML* is crucial to real analysis. The least upper bound of a set of real numbers would be defined in *PM* by a function having an order greater than the order of a function defining the class of real numbers, so that it would fail to be a member of that class, that is, fail to be a real number itself. Whitehead and Russell introduced their *axiom of reducibility* to deal with such problems. It asserted that to each propositional function within a given type, there corresponded an equivalent function of lowest order for that type, so that the same class was defined by both functions. By invoking the axiom, then, real numbers and least upper bounds could be defined by functions of the same (lowest) order. Ramsey forcefully asserted, however, that the axiom of reducibility is far from obvious, and certainly not a principle of logic. Ramsey followed Peano in pointing out that the logical contradictions on Whitehead and Russell's list are dissimilar. (He added to the list Grelling's "heterological" contradiction, which he attributed to Weyl.) Ramsey placed them in two groups—(1) contradictions that could arise within a logical system and (2) contradictions that "cannot be stated in logical terms alone; for they all contain some reference to thought, language, or symbolism, which are not formal but empirical terms" (*FM*, Ramsey 1990, 183). The latter "all involve some psychological term, such as meaning, defining, naming, or asserting. They occur not in mathematics, but in thinking about mathematics" (*ML*, Ramsey 1990, 239). This has come to be known as a distinction between logical paradoxes and semantic paradoxes. Ramsey argued that those in the first group, including the contradictions of Russell's set and Burali's greatest ordinal, can and must be avoided by features of the logical system; in the case of *PM*, the theory of types, without ramification, would do the job. The second group of contradictions included the paradoxes of the liar, the least indefinable ordinal, the least integer not definable in fewer than nineteen syllables, Richard's finitely undefinable decimal, and Grelling's property "heterological." These are what motivated *PM*'s very general vicious circle principle and the ramified theory of types and orders.

How did Ramsey deal with the semantic paradoxes (which he described as "epistemological") and dispense with reducibility? Consider reducibility first. Ramsey agreed that a propositional function's type, determined by the members of its domain, was a real feature of it. But he thought of *PM*'s orders as features of the particular symbolic expressions that point to functions, rather than as features of the functions themselves. Since the functions themselves served to define classes, there was no need to regard their orders as relevant to the definitions of classes, and so no need for the unwanted axiom of reducibility. In taking this direction, Ramsey departed from *PM*'s strict observance of the vicious circle principle, which he regarded as much too broad. He treated definitions as ways of specifying particular functions and classes, rather than as ways of constructing them. It can be acceptable, in the course of successfully specifying what an entity is, to refer to the whole of a class of which that entity is a member. Whatever circularity is involved need not be vicious. This approach relies, of course, on the idea that the entities are already somewhere out there to be selected by the specifications. The Platonistic flavor of his account led Carnap (1931) to label it "theological mathematics," in contrast to the "anthropological mathematics" of the intuitionists. Ramsey later moved away from the account and this particular aspect of it.

What, then, about the semantic paradoxes? They all exploit the relation of *meaning* between symbols and propositional functions they express. Orders are no longer features of functions themselves and no longer play a role in defining classes, but it does make sense to think of orders as features of symbolic expressions that reflect levels of quantificational complexity—over individuals, or over functions (properties or relations) of individuals, or over functions of functions of individuals, and so on. In a move resembling later treatments that developed hierarchies of languages, Ramsey argued that relation(s) of meaning (and so of definability) are ambiguous for symbolic expressions of different orders. The meaning relation that holds for an elementary function is distinct from the meaning relation that holds for a first-order function, and so on up the hierarchy of orders. He further argued that when one keeps track of the distinct meaning relations, each of the semantic paradoxes can be shown not to yield a real contradiction (*FM*, §III).

Impredicative Functions, Infinity, Abandonment of Logicism

Ramsey pushed the extensionality of his system quite far. As mentioned earlier, he arrived at the point of regarding propositional functions as functions from individuals to propositions, rather than as open symbolic expressions. Ramsey called all propositional functions that are truth functions of atomic propositions *predicative* functions (this is not the same meaning that Whitehead and Russell gave the term). Predicative functions include more than just functions of individuals; Ramsey shows that all of the propositional functions of *PM* are predicative in his sense. But in Ramsey's extensionalization of *PM*, there are more propositional functions than this: *Impredicative* functions are those among the mappings from individuals to propositions that cannot be built up as truth functions of atomic propositions. In Ramsey's system, quantification over propositional functions is understood to range over all *functions in extension*, predicative and impredicative.

Ramsey's *third criticism* of *PM* was directed at its treatment of identity. In *PM*, identity was defined by appeal to a principle of indiscernibility of elementary properties. But it is no truth of logic, Ramsey said, that two things cannot share all elementary properties. He instead took *sameness of individual* as a primitive nonlogical idea, and made use of the rich collection of functions in extension to give an account of the propositional function $x = y$. The proposition that *PM* interpreted as saying that indiscernible individuals are related by '=' was instead interpreted as saying that '=' holds between individuals that share all functions in extension, predicative or impredicative. The latter include all mappings from individuals to propositions, however arbitrary, and this amounts to saying that '=' holds among individual(s) that are the same. In this interpretation, the assertion $(\phi_e)(\phi_e\, x \equiv \phi_e\, y)$, where ϕ_e ranges over functions in extension, does turn out to be a tautology exactly when $x = y$, and Ramsey took it, so understood, to be the defining condition for $x = y$. Ramsey applied this account to *PM*'s axiom of infinity, which, in Whitehead and Russell's interpretation, asserted that there are infinitely many individuals distinguishable by predicative functions (in Ramsey's sense of 'predicative'). Ramsey regarded that as at best an empirical claim. Since his reconstruction included functions in extension, his interpretation of the axiom just amounted to the assertion of the existence of an infinity of individuals, whether distinguishable by predicative functions or not. Ramsey acknowledged that this may still appear to be an empirical claim, but he used his account of identity to argue that, though the axiom is unprovable, it is a tautology (necessary) if it is true. At the end of *FM* he advocated adopting it, both on these grounds and because it is indispensable to mathematics.

This is still not an entirely satisfactory result for a logicist theory of mathematics, and Ramsey increasingly departed from that view after the publication of *FM*. The paper *ML* appeared one year later, and in it he still defended a generally logicist outlook against the alternative approaches of Hilbert, Weyl, and Brouwer. In *ML* Ramsey remained dissatisfied with the status of the axiom of infinity, yet he was still convinced that it is needed. There is evidence, however, that his views soon began to change. Braithwaite reports that in 1929 Ramsey "was converted to a finitist view which rejects the existence of any actual infinite aggregate" (Ramsey 1931, xii), and among the notes written in the last years of his life there is clear evidence of his strong interest in finitism (Ramsey 1991a, notes 53 and 54). There is also this statement, written in 1929 at the end of an unpolished set of notes on theories:

> It is obvious that mathematics does not require the existence of an infinite number of things. We say at once that imaginary things will do. . . . But there are no imaginary things, they are just words, and mathematicians and physicists who use the infinite are just manipulating symbols with some analogy to propositions. (Ramsey 1991a, note 58)

There is no further indication how he thought to dispense with an axiom of infinity.

Ramsey's Theorem

All of Ramsey's work in logic and the foundations of mathematics predated the dramatic developments of the 1930s. Who can say how his views would have continued to evolve if he had lived? Before moving on to other topics, it is worth mentioning Ramsey's most lasting contribution to mathematics itself, Ramsey's theorem. In the paper "On a Problem in Formal Logic," Ramsey took up Hilbert and Ackermann's Decision Problem and proved a special case, the decidability of validity for ∃∀-form sentences with identity. In 1936 Alonzo Church showed that the problem for full predicate logic is unsolvable. As a preliminary to the main topic of his paper, however, Ramsey established a remarkable result in combinatorics that has proved seminal to a great deal of subsequent mathematical research. Details can be easily found in mathematical sources on *Ramsey theory* or *Ramsey numbers.*

Probability and Partial Belief

Ramsey's greatest influence on the philosophy of science today is through his work on probability and degrees of belief. His writings on laws and theories introduced other important ideas that are now well remembered and widely used (see below), but Ramsey lacked the opportunity to develop them as fully as the ideas in the remarkable paper "Truth and Probability" (TP). The paper was not widely appreciated prior to Savage's (1954) work, but Ramsey and Bruno de Finetti are now recognized as the two key figures in the origin and early development of contemporary accounts of subjective, or Bayesian, probability. *TP* is a rich paper and cannot be covered fully here (see also Zabell 1991, Jeffrey [1965] 1983, Skyrms 1990, Sahlin 1990, and Galavotti 1991). The paper has five sections—Braithwaite (Ramsey 1931, introduction) says that at one time Ramsey planned to add a sixth, on probability in science, and to publish the paper separately. He later developed plans to include it in a book, or books, on logic, truth, and probability. Drafts of other portions of that project, chapters devoted mainly to truth and judgment, have been published (Ramsey 1991b), and other works contain notes on probability, degrees of belief, and chance (Ramsey 1931 and 1991a).

Ramsey opens *TP* by allowing that there may be two distinct interpretations of probability, one appropriate to logic, the other to statistics and physical science, and he made clear that his subject was the former. The framework he had in mind began by conceiving of logic as the science of rational thought, subdivided into what he called the 'logic of consistency' and the 'logic of truth.' The former contains formal deductive logic and mathematics. Most of *TP* is devoted to developing the idea that the logic of consistency also contains a theory of probability, and to explaining that the sort of probability it must contain is *subjective.* In the final section, Ramsey turned to the logic of truth with the observation that "we want our beliefs to be consistent not merely with each other but also with the facts." Conformity to the logic of consistency gives no guarantee of that, so there is room for a broader human logic, "which tells men how they should think" or what it would be reasonable to believe (Ramsey 1990, 87). Ramsey's remarks on the logic of truth will be discussed in a later section.

Ramsey sets the stage for his own account with a review of the two rival interpretations of probability that were familiar to his readers. (Zabell 1991 is particularly good on the background and context of Ramsey's theory.) One is the frequency account, and the other is J. M. Keynes' logical interpretation of probability. Ramsey made no attempt to refute a frequency interpretation of probability—it is clearly mathematically viable and appears to be useful to science. It turns out not to be a suitable basis for a logic (of consistency) for partial belief, however, and he set aside frequencies until he later took up the logic of truth. Keynes' interpretation, on the other hand, was clearly meant to be a part of logic, and Ramsey gave serious attention to its shortcomings. Keynes advocated the view that a probability is an objective logical relation holding between one proposition and another. He held that such logical relations are unanalyzable, yet they are at least sometimes perceivable, and they serve as guides to rational belief. The degree of belief it is rational to have in an unknown proposition *p* is given by the probability relation that holds between the proposition describing what one knows and *p.* Keynes' degrees of probability were not generally quantitative, but in some cases, by appeal to the principle of indifference, they could be compared and calculated. Especially since the failure of Carnap's sophisticated later attempts at a similar approach, accounts like this are today no longer widely held. Suggestions for their resurrection still emerge from time to time in areas of philosophy where this history is not well known, and Ramsey's criticisms of Keynes remain relevant. They begin with the straightforward observation that

> there really do not seem to be any such things as the probability relations [Keynes] describes. He supposes that, at any rate in certain cases, they can be perceived; but speaking for myself I feel confident that this is not true. . . . [M]oreover I shrewdly suspect that others do not perceive them either, because they are able to come to so very little agreement as to which of them relates any two given propositions. (Ramsey 1990, 57)

Ramsey developed this criticism thoroughly in the second section of *TP*. Later, after the presentation of his own view, three further criticisms of Keynes' theory were given in the fourth section of *TP*. Keynes' account (a) failed to make clear why the logical relations should obey the axioms of probability, (b) attempted to lay down a priori logical constraints such as the principle of indifference to generate what are surely empirical probability values in science, and (c) did not recognize or explain "a probable belief founded not on argument but on [uncertain] direct inspection," since the logical relation indicates the probability of a proposition based only on what is *known* (Ramsey 1990, 86). Ramsey thought it clear that his own theory had none of these flaws.

Degrees of Belief

The third section of *TP* is its longest and most significant. It introduced, as well as anything written since, a conception of *degrees of belief* that is now widely used. Ramsey offered two lines of justification for taking *rational* degrees of belief to be probabilities. One was based on a betting model of action, and led to the Dutch Book Argument. The second provided a groundbreaking generalization of the betting model. Ramsey stated an axiomatic theory of rational preference and derived from it an expected utility theory that put the agent's degrees of belief in the role of probabilities (it is considered in the next section).

What *are* degrees of belief, and why think that they can be quantified? Ramsey developed an account that applies to dispositional beliefs as much as to occurrent beliefs. He considered and rejected the suggestion that the degree of a belief corresponds to the strength of an introspective feeling one might have about it, proposing instead that it must be a causal property of the belief, "which we can express vaguely as the extent to which we are prepared to act on it. This is a generalization of the well-known view, that the differentia of belief lies in its causal efficacy" (Ramsey 1990, 65). Ramsey did not deny, of course, that beliefs are sometimes accompanied by feelings of various intensities—say, of various degrees of conviction. The apparent advantage of such feelings is that they can be known through introspection, which supports the view that one knows how strongly one believes things. But Ramsey doubted that one always knows how strong belief is, and suggested another way of judging a belief's strength that need not rely on internal observation of some belief-feeling: One imagines how one would act in various hypothetical circumstances. Even if there were some quantitative scale on which feelings could be measured, unlikely as that is, he argued, they contribute little to the role of belief as a basis for action.

In its focus on their action-guiding role, Ramsey's account of partial beliefs fits well with his wider perspective on belief, which was strongly influenced by his study of Peirce. That will be neglected here, except to note that in other writings (Ramsey 1927; 1991b, esp. Ch. 3), he offers a broadly functionalist account of a belief's content, or, in his terms, its *propositional reference.* Several discussions of Ramsey's treatment of belief appear in Mellor (1980). Ramsey clearly thought it a good working hypothesis that internal patterns of relevant causal properties or dispositions are present and susceptible of measurement, though he was under no illusion that such measurement is easy. He repeatedly drew analogies between the difficulties of measuring these causal properties and the complications that arise in physical measurements of, for example, length or electric current.

A person's degree of belief in *p* is a causal property contributing to the belief's influence over that person's choices and actions. A familiar "old-established" technique for measuring it, which Ramsey regarded as "fundamentally sound," is to offer that person bets on *p* and see what bets, at what odds, that person is willing to accept. The person's willingness to give high odds on *p* indicates a high degree of belief in *p*; an insistence on receiving high odds indicates that the person's degree of belief is low. More precisely, if the least favorable bet that the person is willing to make on the truth of *p* is one where the person pays \$$a$ if *p* is false and wins \$$b$ if *p* is true, the person's degree of belief in *p* is the betting quotient $a/(a + b)$. A conditional degree of belief in *p given q* is similarly measured by the odds of the least favorable conditional bet—a bet on *p* that is in effect only if *q* is true. (A degree of belief in *p given q*, Ramsey said, is not always the same as the degree to which one would believe *p* if one believed *q* for certain.) Ramsey noted that there are many complications that undermine the generality and precision of the method: the diminishing marginal utility of money or of whatever goods are the payoffs, the possibility that one has particular eagerness or reluctance to make bets, the possible disturbance of one's opinion by the act of making the offer, and so on. Ramsey compared these to similar difficulties in carrying out physical measurements—it can be difficult to isolate one out of a number of forces at work, and measurement may require a physical intervention that alters the system being measured. The use of the betting model to characterize strength of belief predates Ramsey—it was used by Borel (1924) and de Finetti (1937), for example.

The betting model has since been the subject of much discussion and criticism, yet it remains the best way of getting across the idea on which contemporary Bayesian theory is founded. Explicit wagering is a specialized form of activity, and most persons engage in it only occasionally. But the model has the two advantages of familiarity and flexibility, since one can readily imagine betting on a wide variety of propositions. So in a theoretical effort to understand the action-guiding role of degrees of belief, betting makes a good, though hardly perfect, stand-in for actions of all sorts. Many actions can be regarded as expressions of implicit wagers, "Whenever we go to the station we are betting that a train will really run, and if we had not a sufficient degree of belief in this we should decline the bet and stay at home" (Ramsey 1990, 79). Ramsey's own remarks about the usefulness and limits of the idealized model are more sensible than a great deal of what has come after. According to the model, a given person at a given time has *a single* (definite) degree of belief in the proposition *p*. This involves assumptions of both *precision*—degrees of belief have precise numerical values—and *stake-insensitivity.* The latter holds that the action-guiding strength of a belief is unaffected by the size of the stakes involved in the actions so guided (or, at least, unaffected over some significant range of stakes). The degree of belief one has, for example, in "It will storm this afternoon" is unaffected by whether the stakes involve the inconvenience of carrying an umbrella, or more seriously, the risk of death in a small sailboat. There is no doubt that the magnitudes of the stakes affect how one makes choices, but Ramsey's theory and its descendents locate the effects not in changes to the strengths of one's beliefs, but in the interplay between beliefs and desires that yields choice (this point is further discussed in Armendt 1993).

The Logic of Consistency for Degrees of Belief

The fundamental claim is that *rational* degrees of belief satisfy the principles of probability. The argument Ramsey actually presented for it is based on his axiomatic theory of rational preference. But he also stated, for the first time, what has since become known as the Dutch Book Argument. That argument concludes that when degrees of belief violate the axioms of probability, they are flawed, because under their guidance the believer who holds them would be willing to accept a combination, or book, of bets that together yield a sure loss to him, whatever turns out to be true and however the bets pay off—a Dutch Book. Ramsey (1990) did not fill in the argument, but he clearly could have: "If anyone's mental condition violated these laws . . . he could have a book made against him by a cunning bettor and would then stand to lose in any event" (78). He also stated the conclusion of the Converse Dutch Book Argument:

> Having any definite degree of belief implies a certain measure of consistency, namely willingness to bet on a given proposition at the same odds for any stake. . . . Having degrees of belief obeying the laws of probability implies a further measure of consistency, namely such a consistency between the odds acceptable on different propositions as shall prevent a book being made against you. (Ramsey 1998, 78–79)

Both the argument and its converse also appear in de Finetti (1937). In the voluminous later literature on the Dutch Book Argument, Skyrms best addresses the heart of the topic (see e.g., his essay "Higher Order Degrees of Belief" in Mellor 1980; Skyrms 1990, Ch. 5). The key element of Skyrms' interpretation, in keeping with what seems to be Ramsey's own, is that the Dutch Book is a dramatic device. The believer's susceptibility to a Dutch Book illustrates the presence of a flaw in his beliefs; it does not in itself constitute the flaw. Armendt (1993) follows Skyrms' interpretation and further discusses the relationship between the betting model and axiomatic preference theories.

Ramsey used the betting model to present the intuitive idea of degree of belief, but his general account of subjective probability as a norm for degrees of belief really comes from his preference theory, only an overview of which can be given here (for further details, see Jeffrey [1965] 1983 or Sahlin 1990). Ramsey replaced betting with the more general device of a *gamble.* A gamble yields one outcome or another, say α or β, depending on whether a proposition *p* is true. The outcomes α, β were taken to be states of the world very fully specified with respect to things that the believer or agent cares about. In a nutshell, the account was this: The agent has a systematic set of preferences among various outcomes, as well as among a large set of imaginable gambles on them. His preferences are subject to certain principles, which Ramsey stated as axioms of the system and its preference relation. Some axioms are richness assumptions—the system includes preferences among fine-grained arrays of gambles on every proposition in which the agent has a degree of belief. To the extents that this richness is an acceptable idealization and the other axioms (e.g., transitivity and connectedness) are plausible components of a model of rationality, systems that satisfy the axioms can be regarded as systems of rational preference. Ramsey showed that for any such system, two appropriately unique, and so nonarbitrary, measures can be derived. One is a real-valued measurement of the values of the outcomes and gambles (a utility function, though he did not call it that), and the other is a numerical measurement of the propositions that is provably a probability function. The two measures together obey a principle of expected utility, and Ramsey interpreted the probability function as the measure of the agent's degrees of belief. Ramsey's account provides a foundation for both the theory of subjective utility and the theory of subjective probability, though the latter is what he emphasized in *TP*. The theory is a forerunner of many later axiomatic treatments of belief, utility, and decision developed by philosophers, economists, statisticians, and others. The later literature is immense and impossible to survey here. Savage's (1954) theory was extremely influential, and it contributed to the later recognition of Ramsey's work; particularly well known among philosophers is Jeffrey's ([1965] 1983) theory.

The Logic of Truth

Ramsey took the result just described to establish the legitimacy of a logic of consistency for partial beliefs. When the account is regarded as a decision or utility theory, its direct inclusion of degrees of belief clearly makes it an *epistemic* account. When viewed either as a doxastic theory or a utility theory, it is also *subjective*, in that the constraints applying to degrees of belief and preferences are internal to the system, not imposed by what is true or objectively desirable in the world. In later subjective probability theory, and especially in the *radical probabilism* of de Finetti, Jeffrey, and others, subjectivism is fundamental to the account—the apparent objectivity of some probabilities is held to be explicable by the dynamic behavior of interacting systems of subjective (conditional) probabilities. Ramsey himself expresses such a view in the later note "Chance" (Ramsey 1990). The remaining part of *TP*, however, does not contain an explicit commitment to radical probabilism. Here Ramsey sought a fuller account of good epistemic practice, one that goes beyond the logic of consistency by ascertaining the standards and sources of successful belief. On the other hand, in this most exploratory part of the paper, Ramsey asserted little or nothing with which a radical probabilist need disagree.

What is the connection between probability in the sense of (rational) degrees of belief, and probability in the sense of frequencies, or "class-ratios"? While still considering the logic of consistency, Ramsey (1990) allowed that "experienced frequencies often lead to corresponding partial beliefs, and partial beliefs lead to the expectation of corresponding frequencies" (83), but he denied that a general connection along such lines can be made out. What can be said to connect the interpretations is that, "supposing goods to be additive, belief of degree *m/n* is the sort of belief which leads to the action which would be best if repeated *n* times in *m* of which the proposition is true" (84). This is the sense in which the calculus of frequencies was linked to a calculus of consistent partial beliefs. (In the later note, "Reasonable Degree of Belief" [Ramsey 1990], he explores difficulties and refinements associated with this idea.) C. S. Peirce's pragmatism greatly influenced Ramsey's approach to the question of what, beyond consistency, makes degrees of belief reasonable. The *habits* by which one arrives at and maintain beliefs should be at the focus of attention. This led Ramsey (1990) to regard inductive inference as a central topic in the logic of truth, which is not to say that the logic of consistency is silent concerning induction:

> [I]f *p* is the fact observed, my degree of belief in *q* after the observation should be equal to my degree of belief in *q* given *p* before. . . . When my degrees of belief change in this way we can say that they have been changed consistently by my observation. (88)

Beyond conditionalization, though, what inferential habits produce reasonable beliefs, and what standards guide judgments about them? A richer account of induction than Ramsey's, based on the idea of exchangeable sequences of events, was soon given by de Finetti (1937). As to goals and standards, *always fully believe the truth* comes to mind, but it is not very helpful for human belief or for the habits followed in the many predicaments where certainty would be misplaced. A better standard, Ramsey proposed, is that a habit should yield partial beliefs whose strengths correspond to the frequencies with which relevantly similar beliefs are true. He illustrated the point with an example concerning the wholesomeness of toadstools; one might also think of weather forecasts. A habit that yields a degree of belief in p equal to x is reasonable when the frequency with which such beliefs are true is x. (For a discussion of how this proposal may be linked to deeper decision-theoretic standards of success, see Adams 1988).

Ramsey's further remarks in *TP* are a prolegomenon to the logic of truth, rather than a development of it. He asserts that scientific inductive reasoning will be indispensable to the project of identifying and assessing mental habits—judging when and how well they work. Induction is itself such a habit, and a useful one. To understand it better, and to determine how useful it is, one cannot avoid employing it. There is a circle here, but as Ramsey conceives the point of the project, nothing vicious about it.

The Value of Knowledge

TP is by far Ramsey's most substantial and polished effort in the area of partial belief, decision, and probability, but a number of other notes on related topics are among his papers. Two were mentioned above; another particularly interesting note is "Weight or the Value of Knowledge" (Ramsey 1991a, 285–287), in which Ramsey demonstrated a result that was independently rediscovered by Savage and by I. J. Good. Can a decision maker generally expect to be better off by acquiring more information before making a choice? Ramsey showed, in the context of his decision theory, that when free information is available, acquiring it will not lower, and may increase, the expected utility of the decision. Skyrms (1990, ch. 4) points out that the setting Ramsey uses in his treatment suggests that he may also have partly anticipated the generalized form of belief updating developed by Jeffrey ([1965] 1983) known as *probability kinematics.*

Scientific Theories, Laws, and Causality

During 1928–1929, Ramsey wrote several notes and papers on these topics that contain ideas of lasting influence. Since the papers are unfinished, and his views were in some respects clearly evolving and unsettled, the focus here will be on their general direction and on several ideas that were later taken up and developed more fully than Ramsey himself had opportunity to do. One theme that runs through the papers is an instrumentalist view of laws and theories.

The 1929 paper "Theories" investigated the formal structure of scientific theories. Ramsey's approach reflected his study of recent work by Nicod, Carnap, and Russell; he was interested in the content of theoretical assertions and in how such content is related to the observational assertions that the theory explains. Assume for the moment that the two sorts of assertions can be clearly distinguished. An idea attractive to logical positivists was that, in principle, anything expressed by theoretical assertions could also be expressed in a more roundabout way by observational assertions alone. One way to do this is to show that theoretical terms are explicitly definable from observational ones and that the definitions can be inverted. Ramsey presented a simple, toy example and explored what would be required to show this. He concluded that it might be done, but only in a way so complex and cumbersome that it would never be worth doing. That was not surprising, but a more significant problem is that the theory obtained by the method of explicit definition is too rigid. Further observations might suggest additions to the theory, but additions cannot be made without altering the definitions and thereby changing the meanings of its terms (this point was further developed by Braithwaite 1953, Ch. 3).

Ramsey offered another proposal. Assume a theory T that is axiomatized in first-order logic, with a distinct theoretical vocabulary whose terms appear in T and in a set C of correspondence rules (or as Ramsey says, a dictionary) relating theoretical and observational assertions. Conjoin all the axioms of T and all the rules of C, replace the occurrences of each distinct theoretical term with a second-order variable, and introduce for each distinct variable a second-order existential quantifier that binds its occurrences. The resulting sentence contains only observational terms, it is entailed by $(T \wedge C)$, and the particular observation sentences it entails are the same as those entailed by $(T \wedge C)$. This device is now known as the *Ramsey sentence* of the theory, and it has since been widely used in diverse treatments of the content, meaning, and truth of theories (see e.g., Hempel 1958, Carnap 1966, and Lewis 1970; there are many others). One notable area in which Ramsey sentences have been widely used is in the philosophy of mind. Functionalists in particular, following Lewis (1972), have often advocated employing Ramsey sentences to characterize mental states.

How much is accomplished or shown by any particular use of the Ramsey sentence technique clearly depends upon the application. Is the matter at hand one in which a relationship between two distinct linguistic, syntactic systems is really the main target of interest? Proponents of semantic or model-based approaches to understanding theories generally think that this is not the significant issue, and believe that more light can be shed on theories by thinking of set-theoretic or model-theoretic structures.

Ramsey explored two ways of thinking of causal laws. The brief 1928 note "Universals of Law and of Fact" (Ramsey 1990) explains the difference between universals of law and universals of fact (between lawlike and accidental generalizations) by appeal to an ideal future system of complete knowledge about the world. If one knew everything, Ramsey said, one would want to organize that knowledge in a deductive system in a way that strives for simplicity. The general axioms of the system would be the fundamental laws of nature, and the generalizations derivable from them without reference to facts of existence are derivative laws of nature. The choice of axioms "is bound to some extent to be arbitrary, but what is less likely to be arbitrary if any simplicity is to be preserved is a body [of such generalizations]" (Ramsey 1990, 143). This is as close as Ramsey came to a uniqueness claim, and he did not invoke a set of independently natural properties, or ways of carving up the world, that the ideal theory need capture. "As it is," he said, "we do not know everything; but what we do know we tend to organize as a deductive system and call its axioms laws, and we consider how that system would go if we knew a little more and call the further axioms or deductions there would then be, laws" (ibid). In this unpublished note, Ramsey did not cite him, but the account may well have developed from Ramsey's familiarity with John Stuart Mill's work. In any case, within a year Ramsey discarded the view. Its influence has endured, however, in the work of more recent philosophers, notably Lewis (1973) and Earman (1986) (see Laws of Nature).

The paper "General Propositions and Causality" contains Ramsey's second, revised treatment of causal laws. It was written in 1929, at a time when Ramsey had frequent conversations with Wittgenstein, who had just moved to Cambridge. The new view was that a law is not a summary of propositions about particular events; its causal force lies in our trust of it as a guide to inferences about particular events. Causal generalizations "are not judgments but rules for judging 'If I meet a ϕ, I shall regard it as a ψ.' This cannot be *negated* but it can be *disagreed* with by one who does not adopt it" (Ramsey 1990, 149). To assert a causal law is to assert a formula from which one can derive propositions about particular events. Its *causal* character lies in the temporal ordering of the events (ψ does not precede ϕ). The special importance attached to rules for judgments so ordered is traceable to how one thinks about one's actions; in deliberation one gives special importance to forward-looking rules, those that look to the future. Ramsey's views here have affinities to a number of subsequent treatments that strive for pragmatic, reductive accounts of causal necessity; a noteworthy example is Skyrms (1980).

This article concludes by mentioning a suggestion Ramsey (1990) made in a footnote to a discussion of conditionals in "General Propositions and Causality," at a point where he was distinguishing "hypotheticals" from material implications: "If two people are arguing 'If *p,* will *q*?' and are both in doubt as to *p*, they are adding *p* hypothetically to their stock of knowledge and arguing on that basis about *q*. ... We can say that they are fixing their degrees of belief in *q* given *p*" (155). The idea is that the acceptability of an indicative conditional corresponds to the acceptability of its consequent after the antecedent is hypothetically added to one's beliefs. Ramsey took the latter to be measured by the conditional probability of the consequent on the antecedent. The idea has been extensively developed by Adams (1975). In later philosophical literature on conditionals, Ramsey's suggestion, known as the *Ramsey test*, is widely embraced, at least up to a point. Precise characterizations of the idea vary, however, as do opinions about the scope of its adequacy. The large body of literature cannot be covered here (a recent survey of much of it is in Bennett 2003).

Brad Armendt

References

Adams, Ernest (1975), *The Logic of Conditionals.* Dordrecht, Holland: Reidel.

——— (1988), "Consistency and Decision: Variations on Ramseyan Themes," in Harper and Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics.* Dordrecht, Holland: Kluwer, 49–69.

Armendt, Brad (1993), "Dutch Books, Additivity, and Utility Theory," *Philosophical Topics* 21: 1–20.

Bennett, Jonathan (2003), *A Philosophical Guide to Conditionals.* Oxford: Oxford University Press.

Borel, Emile (1924), "A propos d'un traité de probabilités," *Revue philosophique* 98: 321–326.

Braithwaite, Richard (1953), *Scientific Explanation.* Cambridge: Cambridge University Press.

Carnap, Rudolf ([1931] 1964), "Die logizistische Grundlegung der Mathematik," translated in Benacerraf and Putnam (eds.), *Philosophy of Mathematics.* Englewood Cliffs, NJ: Prentice-Hall, 31–49. Originally published in *Erkenntnis* 2: 91–105.

——— (1966), *Philosophical Foundations of Physics.* Edited by M. Gardner. New York: Basic Books.

de Finetti, Bruno (1937), "La Prévision: ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré* 7: 1–68.

Earman, John (1986), *A Primer on Determinism.* Dordrecht, Holland: Reidel.

Galavotti, Maria Carla (1991), "The Notion of Subjective Probability in the Work of Ramsey and de Finetti," *Theoria* 57: 239–259.

Grattan-Guinness, I. (2000), *The Search for Mathematical Roots, 1870–1940.* Princeton, NJ: Princeton University Press.

Hempel, Carl (1958), "The Theoretician's Dilemma," in H. Feigl et al. (eds.), *Minnesota Studies in the Philosophy of Science* (Vol. II). Minneapolis: University of Minnesota Press.

Jeffrey, Richard ([1965] 1983), *The Logic of Decision.* Chicago: University of Chicago Press.

Lewis, David K. (1970), "How to Define Theoretical Terms," *Journal of Philosophy* 67: 427–446.

——— (1972), "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy* 50: 249–258.

——— (1973), *Counterfactuals.* Cambridge: Harvard University Press.

Mellor, D. H. (1995), "F. P. Ramsey," *Philosophy* 70: 243–262.

——— (1980), *Prospects for Pragmatism: Essays in Memory of F. P. Ramsey.* Cambridge: Cambridge University Press.

Ramsey, Frank P. ([1922] 1989), "Mr. Keynes on Probability," *British Journal for the Philosophy of Science* 40: 219–222. Originally published in *The Cambridge Magazine* 11: 3–5.

——— (1927), "Facts and Propositions," *Aristotelian Society Supplementary Volume VII*, 153–170.

——— (1931), *The Foundations of Mathematics and Other Logical Essays.* Edited by R. B. Braithwaite. London: Routledge & Kegan Paul.

——— (1978), *Foundations.* Edited by D. H. Mellor. Atlantic Highlands, NJ: Humanities Press.

——— (1990), *Philosophical Papers.* Edited by D. H. Mellor. Cambridge: Cambridge University Press.

——— (1991a), *Notes on Philosophy, Probability, and Mathematics.* Edited by Maria Carla Galavotti. Naples: Bibliopolis.

——— (1991b), *On Truth.* Edited by Nicholas Rescher and Ulrich Majer. Dordrecht, Holland: Kluwer.

Sahlin, Nils-Eric (1990), *The Philosophy of F. P. Ramsey.* Cambridge: Cambridge University Press.

Savage, Leonard J. (1954), *The Foundations of Statistics.* New York: Wiley & Sons.

Skyrms, Brian (1980), *Causal Necessity.* New Haven, CT: Yale University Press.

——— (1990), *The Dynamics of Rational Deliberation.* Cambridge: Harvard University Press.

Sullivan, Peter (1995), "Wittgenstein on 'The Foundations of Mathematics,' " *Theoria* 61: 105–142.

Whitehead, Alfred North, and Bertrand Russell (1910), *Principia Mathematica* (3 vols.). Cambridge: Cambridge University Press.

Wittgenstein, Ludwig ([1922] 1961), *Tractatus Logico-Philosophicus.* Translated by F. P. Ramsey and C. K. Ogden. London: Routledge & Kegan Paul.

Zabell, S. L. (1991), "Ramsey, Truth, and Probability," *Theoria* 57: 211–238.

See also **Decision Theory; Dutch Book Argument; Epistemology; Laws of Nature; Probability; Russell, Bertrand**

RATIONAL RECONSTRUCTION

Philosophers of science do many things. Nevertheless, the demand arises on occasion for them to give a general account of the relation of the work of philosophy of science to work in the sciences. To this demand, there are several responses (e.g., logical analysis of scientific and metascientific concepts, an explicit account of scientific method). One answer that has been employed in various places and times since the twentieth century is that philosophy of science engages in a rational reconstruction of science. This seems to raise as many questions as it answers. Does not the need for a "rational reconstruction" of science at the hands of philosophers seem to indicate that science as practiced is in some important sense not (wholly) rational? What is it about science that needs to be reconstructed in order to better exhibit its rational structure? What are the proper tools for reconstructing science?

This article seeks to provide a brief but balanced account of the point of rational reconstruction in the two projects that most importantly advanced that understanding of the proper business of philosophy of science: logical empiricism and Imre Lakatos's methodology of scientific research programs (see Logical Empiricism; Research Programs). Rational reconstruction is connected with many central issues of philosophical method within analytic philosophy of science. Arguments over rational reconstruction connect also to large debates about proper method in philosophy generally and, in particular, to the debates in the late nineteenth and early twentieth centuries over the relations of philosophy to psychology and sociology (for excellent introductions to these debates, see Kusch 1995 and 1999).

Logical Empiricism

Within analytic philosophy of science, Rudolf Carnap's *Der logische Aufbau der Welt* [The Logical Construction of the World] of 1928 is the work often cited as first articulating an understanding of the business of philosophy of science as rational reconstruction (see Carnap, Rudolf). Carnap's project in the *Aufbau* is a rational reconstruction of the objects of science through exploitation of the technical resources of the new mathematical logic. He sketches a system of logical definitions that shows how all proper scientific claims can be translated back into a language that makes reference only to experience. Logical definition and inference are, of course, exemplary rational procedures, and thus the constructional system rationalizes psychologically attained knowledge, exhibiting its objective conceptual meaning.

It is sometimes suggested that rational reconstruction was introduced by the young Carnap (Reichenbach 1938) and is an expression of a radical break between early logical empiricism and the German epistemology that was pursued in the generations before it. Carnap's use of 'rational reconstruction' [*rationale Nachkonstruktion*] in the *Aufbau* was, in fact, a move within a well articulated debate among neo-Kantians, phenomenologists, positivists, Marxists, and others over the proper methods of epistemology (Richardson 1999). For example, the positivist Theodor Ziehen (1914), in his meta-epistemological work *Zum gegenwärtigen Stand der Erkenntnistheorie* [On the Present State of Epistemology], distinguished between two methods used in the epistemology of his day. He called the first the 'genetic method,' since it traced representations used in judgment back to their psychological origins; and he called the second the 'reconstructive method' because it "to a certain extent reconstructs the world of impressions from representations and judgements" (29). That is, the proponents of the reconstructive method say that the world of impressions is not the content of knowledge but is itself only an object of knowledge when placed within a system of representations and judgments. Ziehen associated the genetic method with positivism and naturalism, and the reconstructive method with neo-Kantianism.

Within this general framework, 'reconstruction' was closely associated with 'rational.' Indeed, some neo-Kantians used the language of reconstruction and tied it directly to the rationalizing of the experiential world. For example, Jonas Cohn (1908), in his *Voraussetzungen und Ziele des Erkennens* [Presuppositions and Goals of Knowing], calls pure mathematics an a priori "constructive" science because it constructs its own objects. Pure experience is not, for him, the source of knowledge so much as is the rationally opaque starting point of knowledge that is brought under rational control through mathematized natural sciences. These sciences use the resources of mathematics to rationalize and objectify experience:

> All reconstruction is partial rationalization. The concepts of the particular reconstructive sciences are also, considered in and of themselves, constructions and, as such, fully transparent. But the new epistemic task that they serve gives them a close relation to the inexhaustible and opaque experiential reality . . . . One can say, therefore, that the particular reconstructive sciences concern themselves with individuals only insofar as these can be captured under general determinations. (Cohn 1908, 342)

Carnap's articulation in 1928 of the point of epistemology as rational reconstruction of the results of cognition is a move, therefore, in an ongoing meta-epistemological debate. The formal logic of Russell and Whitehead's *Principia mathematica* was to be used to show how the objects of knowledge that are picked out by ordinary cognitive processes and the more articulated processes of science can be arrived at logically and discursively. Thus, the process of rational reconstruction has a curious feature: It yields the results independently already attained in science but replaces rationally opaque processes with transparently rational definitions and inferences:

> The fact that we take into consideration the epistemic relations does not mean that the syntheses or formations of cognition [*Erkenntnis*], as they occur in the actual process of cognition, are to be represented in the constructional system with all their concrete characteristics. In the constructional system, we shall merely reconstruct these manifestations in a rationalizing or schematizing fashion; intuitive understanding is replaced by discursive reasoning. (Carnap 1967, 89)

Rational reconstruction replaces the rationally opaque psychological processes by which knowledge is typically attained with explicit logical definitions and inferences that show how the results of those processes are genuinely objects of rational knowledge.

The rational reconstructionism of the *Aufbau* fits into two more disputes central to German epistemology at the time. The first was over the rational status of metaphysical claims. Carnap's principal philosophical use of rational reconstruction is not so much positive regarding the sciences as negative regarding metaphysics. Carnap argues that the claims of the metaphysicians *cannot* be rationally reconstructed: Such talk has no touchstone in experience and no rational control from logic; it is without content and form. The second dispute was the by-then long-standing question of the relation of epistemology to psychology. Carnap's rational reconstructionist epistemology distinguishes the questions of the objectivity of knowledge that are properly epistemological from the workings of cognition in the causal order that are a matter for empirical study of psychology. These two debates come together, since presumably there is some psychosocial causal story about why some people talk about metaphysical matters such as the relation of *Dasein* and Nothingness. Carnap's view was that whatever causal story there is that explains the existence of such talk in a culture, the talk itself cannot be rationalized and has no meaning. Thus, the task of rational reconstruction is not simply to reproduce willy-nilly whatever the results of any psychological process may be. By 1932, Carnap (1959) explicitly claimed that the processes giving rise to metaphysical talk were affective or conative, not cognitive, psychological processes.

The first logical empiricist work in English to make much of the notion of rational reconstruction was Hans Reichenbach's (1938) *Experience and Prediction*. In this work, the need to properly distinguish epistemological from psychological concerns is very much in the foreground, and rational reconstruction is marshaled exactly here:

> Epistemology does not regard the processes of thinking in their actual occurrence; this task is entirely left to psychology. What epistemology intends is to construct thinking processes in a way in which they ought to occur if they are to be ranged in a consistent system; or to construct justifiable sets of operations which can be intercalated between the starting-point and the issue of thought-processes, replacing the real intermediate links. Epistemology thus considers a logical substitute rather than real processes. For this logical substitute the term *rational reconstruction* has been introduced; it seems an appropriate phrase to indicate the task of epistemology in its specific difference from the task of psychology. (5f)

Indeed, Reichenbach (1938) introduces his famous distinction between the contexts of discovery and justification as a "more convenient determination" (6) of the notion of rational reconstruction. He stresses the way in which the justification of scientific claims is a matter of public communication, whereas the psychology of scientific discovery can be a subjective and intuitive matter. Only the former is the proper concern of epistemology, and even in this area, epistemology substitutes a fully logically articulated structure for the more inchoate and suggestive arguments actually found in the writings of scientists engaged in reasoned persuasion. Following Reichenbach, the term 'rational reconstruction' was routinely employed by logical empiricists to explain the point of their philosophical enterprise.

Logical empiricist rational reconstruction was not without its critics. The most famous is, of course, Willard Van Orman Quine (1969), who argued that once logical empiricists themselves rejected radical reductionism, "the translation of all significant discourse into the language of experience," there was no further point for rational reconstruction:

> To relax the demand for definition, and settle for a kind of reduction that does not eliminate [the defined terms], is to renounce the last remaining advantage that we supposed rational reconstruction to have over straight psychology; namely, the advantage of translational reduction. If all we hope for is a reconstruction that links science to experience in explicit ways short of translation, then it would seem more sensible to settle for psychology. Better to discover how science is in fact developed and learned than to fabricate a fictitious structure to a similar effect. (78)

Quine's move toward psychology and naturalism here has been widely followed (see Quine, Willard Van). However, one can read Carnap's movement away from the term 'rational reconstruction' to 'explication' as a sort of response to Quine (see Explication). The point of philosophical work is not to find out how science has developed and been learned, but to provide precise resources for its future development through the conceptual tools of logic. That is, for Carnap, rational reconstruction or explication consists not so much of forensic norms for the purpose of understanding and evaluating how things have gone in science, but of deliberative resources for aiding in a clearer language of science for the future. This is, in fact, how he connected rational reconstruction and explication in the preface to the second edition of the *Aufbau* in 1961:

> By rational reconstruction is here meant the searching out of new definitions for old concepts. The old concepts did not ordinarily originate by way of deliberate formulation, but in more or less unreflected and spontaneous development. The new definitions should be superior to the old in clarity and exactness, and, above all, should fit into a systematic structure of concepts. Such a clarification of concepts, nowadays frequently called "explication", still seems to me one of the most important tasks of philosophy. (Carnap 1967, vi)

Rational reconstruction, especially when deployed as a strict distinction between scientific discovery and justification, has been subject to other important objections. Thomas Kuhn and others in the historical philosophy of science have stressed a continuity of discovery and justification; Kuhn (1977) remarks, for example, "considerations relevant to the context of discovery are then relevant to justification as well; scientists who share the concerns and sensibilities of the individual who discovers a new theory are ipso facto likely to appear disproportionately frequently among that theory's first supporters" (328). Indeed, Kuhn's work on multiple discovery suggests an even deeper lesson: In the absence of successful justification, scientific discovery has not even happened (see Kuhn, Thomas). Discovery is generally assigned to whomever can best marshal the resources of persuasion, and a claim that is not substantiated in the scientific community is not a scientific discovery at all.

Lakatos

The second major attempt to explain philosophy of science as rational reconstruction was at the very center of the historical philosophy of science that had rejected logical empiricist rational reconstructionism. Imre Lakatos importantly revived the notion of rational reconstruction in a new setting, with his view that philosophy of science engaged in the rational reconstruction of the history of science (see Lakatos, Imre). Lakatos argued that any history of science begins with a prior normative sense as to what counts as a scientific achievement in the first place, and this normative sense is an implicit or explicit philosophy of science. An explicit philosophy of science is, thus, an account of the constitutive norms of scientific achievement, and a history of science based on such a philosophy presents an internal history of scientific rationality according to this norm. Lakatos (1978b) summarizes his view of the relation of philosophy of science and the history of science:

> [P]hilosophy of science provides normative methodologies in terms of which the historian reconstructs "internal history" and thereby provides a rational explanation of the growth of objective knowledge; (b) two competing methodologies can be evaluated with the help of (normatively interpreted) history; (c) any rational reconstruction of history needs to be supplemented by an empirical (socio-historical) "external history." (102)

A philosophy of science thus determines what internal history of science is, which in turn determines which historical questions about science are relegated to an external, social, and psychological history. Philosophical presuppositions, therefore, set the problems that a historian must solve in order to write a history of the rational development of science while simultaneously assigning some problems to a nonrational or even irrational external history. For example, a falsificationist methodology of science can assign a rational meaning to an experiment only if it can find a theory that the experiment attempts to falsify. Thus, if an experiment is presented by the scientists who designed it as independent of theory, the falsificationist historian must find hidden in the historical record a theory that the experiment in fact tested. In the absence of such a theory, the falsificationist cannot assign a rational point to the experiment. If no such theory can be found, then the falsificationist must assign both the scientist's own account of science and the experimental activity to a nonrational external history—perhaps a history of false consciousness due to faulty scientific pedagogy.

It does seem plausible that a historian of science would have to go to the historical record with certain presumptions about which activities are properly scientific. What Robert Boyle did with his air pump is part of the history of science; what Robert Boyle had for his breakfasts and what his butler did with the air pump are neither of them events properly in the history of science. Lakatos's pronouncements on how to write history, however, struck many as so theory laden as to require a deliberate falsification even of what was obviously internal to science. For example, in a famous passage, Lakatos (1978b) writes:

> Internal history is not just a *selection* of methodologically interpreted facts; it may be, on occasions, their *radically improved version.* One may illustrate this using the Bohrian programme. Bohr, in 1913, may not have even thought of the possibility of electron spin. He had more

> than enough on his hands without the spin. Nevertheless, the historians, describing with hindsight the Bohrian programme, should include electron spin in it, since electron spin fits naturally in the original outline of the programme. Bohr might have referred to it in 1913. Why Bohr did not do so is an interesting problem which deserves to be indicated in a footnote. (119)

When called upon to diagnose Lakatos's curious attitude toward history, some theorists, notably Hacking (1981) and Koertge (1976), have, quite rightly, pointed to Lakatos's lingering Hegelianism. Unlike most empirically minded Anglo-analytic philosophers, who think that the set of historical events is simply the set of all things that have happened, Hegelians have a teleological sense of history, and any event that does not fit into the thread of the story is not part of history. To have a diagnosis of Lakatos's attitude is not yet to convince anyone else to adopt his view. Many have found Lakatosian rational reconstructions of history of science to be grotesque parodies that cover up more than they reveal about the dynamics of science. Kuhn indicated, indeed, how Lakatosian rational reconstructions (which must have looked like highly theoretical versions of the textbook histories of science he inveighed against) violate the norms that historians work within:

> The problem is not that philosophers are likely to make errors—Lakatos knows the facts better than many historians who have written on these subjects, and historians do make egregious errors. But a historian would not *include in his narrative* a factual report which he *knew* to be false. If he had done so, he would be so sensitive to the offense that he could not conceivably compose a footnote calling attention to it. (Kuhn 2000, 151)

Conclusion

Philosophy of science has been, and continues to be, motivated in large measure by a desire to explain the rationality of science. Rational reconstruction was, to an important degree, a response to the demise of inductivist accounts according to which science was rationally constructed via an inductive method. Rational reconstruction worked together with conventionalism and hypothetico-deductivism to grant a freedom and creativity to the work of scientists while rescuing the rationality of science, which was now located in *post facto* rational evaluation. Debates over rational reconstruction are thus ultimately debates about whether philosophers should be defending the rationality of science and what resources are available to do so.

Alan Richardson

References

Carnap, Rudolf (1967), *The Logical Structure of the World and Pseudoproblems in Philosophy*. Translated by Rolf George. Berkeley and Los Angeles: University of California Press.

——— (1959), "The Elimination of Metaphysics through Logical Analysis of Language," in A. J. Ayer (ed.), *Logical Positivism*. New York: Free Press, 60–81.

Cohn, Jonas (1908), *Voraussetzungen und Ziele des Erkennens*. Leipzig: Wilhelm Englmann.

Friedman, Michael (1999), "Epistemology in the *Aufbau*," in *Reconsidering Logical Positivism*. Cambridge: Cambridge University Press, 114–162.

Hacking, Ian (1981), "Lakatos's Philosophy of Science," in Hacking (ed.), *Scientific Revolutions*. Oxford: Oxford University Press, 128–143.

Koertge, Noretta (1976), "Rational Reconstructions," in Robert S. Cohen, Paul K. Feyerabend, and Marx W. Wartofsky (eds.), *Essays in Memory of Imre Lakatos*. Dordrecht, Holland: Reidel, 359–369.

Kuhn, Thomas S. (1977), "Objectivity, Value Judgment, and Theory Choice," in *The Essential Tension*. Chicago: University of Chicago Press, 320–339.

——— (2000), "Reflections on My Critics," in James Conant and John Haugeland (eds.), *The Road Since Structure*. Chicago: University of Chicago Press, 123–175.

Kusch, Martin (1995), *Psychologism*. London: Routledge.

——— (1999), "Philosophy and the Sociology of Knowledge," *Studies in History and Philosophy of Science* 30: 651–685.

Lakatos, Imre (1978b), "History of Science and Its Rational Reconstructions," in John Worrall and Gregorie Curry (eds.), *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press, 102–138.

Quine, Willard Van Orman (1969), "Epistemology Naturalized," in *Ontological Relativity and Other Essays*. New York: Columbia University Press, 69–90.

Reichenbach, Hans (1938), *Experience and Prediction*. Chicago: University of Chicago Press.

Richardson, Alan (1998), *Carnap's Construction of the World*. Cambridge: Cambridge University Press.

——— (1999), "How to Be a Good Non-Naturalist: Epistemology as Rational Reconstruction in Carnap and His Predecessors," in Julian Nida-Rümelin (ed.), *Rationality, Realism, Revision*. Berlin: De Gruyter, 856–861.

Scholz, Gunter (1998), "Rekonstruktion," in Joachim Ritter and Karlfried Gründer (eds.), *Historisches Wörterbuch der Philosophie*. Basel: Schwabe, 570–578.

Ziehen, Theodor (1914), *Zum gegenwärtigen Stand der Erkenntnistheorie*. Wiesbaden: Bergmann.

See also **Carnap, Rudolf; Explication; Lakatos, Imre; Logical Empiricism; Research Programs**

RATIONALITY

See **Incommensurability; Kuhn, Thomas; Scientific Change; Scientific Progress**

REALISM

Realism in the philosophy of science is basically the thesis that unobservable entities posited by empirically successful theories exist (see Theories). Theoretical posits, like electrons or genes, are not just useful ideas but real entities. Realism takes the explanatory and predictive success of theories to warrant an ontological commitment to the existence of the entities they posit. But it is certainly possible for theoretical posits to be useful even if the entities posited do not exist. Scientists themselves use many theories that they disbelieve. Antirealism claims that the explanatory and predictive success of theories testifies only to the utility of their posits and cannot warrant the belief that the posited entities are real. At issue is the reach of empirical evidence (see Instrumentalism).

Virtually all the positions and arguments advanced in the debate over scientific realism are implicated in this statement of the issue. Interpretive problems will be addressed first so as to construct a defensible formulation of the realist thesis. Subsequent sections will evaluate the major arguments.

Formulating the Realist Thesis

Components of Realism

Scientific realism, as described, has semantic, metaphysical, and epistemic components. Semantically, it claims referential success for theoretical concepts and truth for existence claims of theories. This much semantics is redundant. It reduces to the metaphysical claim that posited entities exist. To assert that *p* is true is to assert that *p*, and to assert that *a* refers is to assert that there are *a*s. These disquotational properties eliminate 'true' and 'refers' from contexts where *p* and *a* are explicit.

But realism makes a further commitment to the descriptive success of theories. Whether or not reference to an unobservable entity requires describing it truly, realism must attribute some truth to a theory's description of the entity if evidence for the theory is to warrant realism's metaphysical claim. For, as the entity is unobservable, its mere existence, independently of its nature and relations to other entities, carries no observable consequences. So the realist believes not only that the entity exists, but also that it has certain properties. Because theories develop and change in response to new evidence, and competing or successive theories sometimes differ in the properties they attribute to the same entity, this further realist commitment is vague and inconstant. It is therefore doubtful that the semantic component of realism can be made entirely redundant.

It is instructive, on this point, to contrast the thesis of "entity realism" advocated by Ian Hacking (1983). Hacking eschews all talk of truth and all theoretical description; his realism is exclusively metaphysical, committed to the entities alone in virtue of their technological applications in the study of yet more speculative aspects of nature. He believes that electrons exist because scientists say they use them to study other things, but he professes no theoretical beliefs as to the nature of electrons. As Hacking disallows himself the resources to explain how a theoretical entity manages to be technologically useful, deference to how scientists describe their own practice is his only means of identifying the entities to which his realism is committed. This leaves him wide open to an antirealist reading of such description as a useful manner of speech without ontological import.

A viable realism must endorse some theoretical properties as well as a theoretical ontology. The difficulty, to which Hacking's minimalism is responsive, is to make this further commitment specific. Characteristically, realism speaks of there being "some truth" to theory, but declines to be fully explicit as to what this truth is. Realism claims that there is truth enough to account for the predictive and explanatory effectiveness of unobservable posits. But not knowing how much truth or what truth is enough, realism cannot cash its truth attributions in for specific theoretical assertions.

Epistemically, realism claims that empirical evidence justifies theoretical beliefs. This component is, again, only partially redundant. Specification of the propositional content of the beliefs justified can constitute the realist thesis without adverting to justification. But the denial that any specific theoretical proposition is justified is not yet antirealism. Antirealism claims not that particular propositions are unjustified, for the realist could agree and endorse others, nor even that *no* theory is justified, for the realist could attribute this to a contingent lack of evidence, perhaps to lack of resources or diligence. Antirealism claims that it is not in the nature of empirical evidence to justify theoretical belief. Realism and antirealism divide over the *capacity* of the kind of evidence in principle available to warrant beliefs as to the existence and nature of entities that cannot, in principle, be observed. To claim that theoretical beliefs are epistemically justifiable is already realism, and this claim does not reduce to any endorsement of a particular theory.

Of course, realists characteristically claim much more. They claim that many theories in fact enjoy an evidential warrant sufficient to show that the unobservable entities they posit exist. But then they meet the additional argumentative burden that this additional commitment incurs with vacillation or vagueness; the favored theories are unspecified or their specification is disputed. Typically, they are identified only as the "best supported" theories of current science, without specifying a standard of support or determining the degree of support of any particular theory. This has not mattered much to the debate. The principal arguments for and against realism are largely insensitive to differences among its semantic, metaphysical, and epistemic components.

The Observational-Theoretical Distinction

The distinction between observation and theory is crucial to realism, and it might seem that the whole debate is obviated by the disrepute into which this distinction has fallen (Maxwell 1962; Churchland 1985) (see Observation). If the distinction is denied, then the alternative to realism is an expansion of antirealism into a thoroughgoing skepticism. The antirealist needs the distinction to delimit such incredulity. But surely some, at least rough and ready, distinction is allowable; its difficulties go not so much to showing that the distinction is untenable as that there are too many ways to draw it, none of them neat. An epistemically pristine concept of observability independent of science is elusive in principle. And if one follows science in countenancing theory-mediated methods of observation, then the boundary of the observable is ever-shifting (Leplin 1984b; Shapere 1982). Atoms and microbes used to be examples of unobservable entities; now the status of atoms is unclear. Elementary particles seem safe, unless one begins to worry about whether they are rightly conceived of as "entities" at all and whether it makes sense even to entertain their observability (see Particle Physics).

One can only assume that the issue is rightly joined at some point, for in claiming that empirical evidence can warrant theoretical belief the realist is not predicting that all objects of such belief will become observable. Whatever the point of engagement, beliefs on the observational side are not in dispute. The realist need not account for their justification but is entitled to assume it in defending the justifiability of beliefs on the theoretical side. Conversely, an argument strong enough to impugn observational beliefs is out-of-bounds to the antirealist. It is not the burden of realism to refute skepticism.

The Scope of Realism

Realism's obvious attraction, and motivation, is to side with science. Realism brooks no philosophical impediment to what science discovers. This advantage depends on delimiting the range of theoretical entities, and propositions about them, that realism endorses. For science itself declines to treat many of its theoretical posits realistically. If it is not to disagree with science, realism must disqualify some entities.

Empirical success *tout court*—explanatory and predictive utility—does not seem an adequate basis for doing so. Stipulating that success be uniform—uncompromised by failure—disqualifies replaced theories of continuing utility, but this stipulation is both too strong and too weak. The records of entrenched, credible theories are not unblemished; there are reasons for tolerating empirical problems rather than blaming the theories that face them, nor is there any expectation that the best theories scientists could ever have would be trouble free. And theories at the frontiers of research, theories too speculative and underdeveloped to believe, may, for a time at least, face no empirical difficulty whatever. They may be supported by all available evidence simply because too little pertinent evidence is available. Worse, rival theories may all be successful; the realist must not be ontologically indiscriminate where science is skeptical.

The standard solution is to restrict realism to mature, entrenched, well-tested, current theories. This is inadequate on several counts. Not all the entities such theories posit are meant to have ontological standing. Some are conceptual devices that the theory is not committed to. And their status can change. Scientists come to interpret realistically entities originally introduced as artifacts of mathematics or aids to computation—positrons and quarks, for example. And entities originally intended realistically get reinterpreted as conceptual devices whose existence is prohibited by theory—electron orbits, for example (see Chemistry, Philosophy of). Some theoretical entities are part of a conceptual background assumed in interpreting a theory but uninvolved in the application of the theory to observable situations—the electromagnetic ether, for example. The conceptual framework may be dispensable or replaceable without empirical cost. Some entities are conceptually anomalous to the point that commitment to them is unreasonably precipitous however well the evidence supports them and however indispensable to the progress of science they appear—gravitational mass, for example. A defensible realism must somehow unburden itself of a profusion of theoretical entities that scientists—whose epistemic ambitions are entangled with pragmatic, aesthetic, and heuristic interests philosophically infected by sociology—admit into their best theories.

To address these problems, the following statement of scientific realism is proposed:

> *SR*: Theoretical entities that are needed to explain or predict empirical results, and that are posited by well-supported theories free of empirical or conceptual difficulties, exist and have those of the properties these theories attribute to them that enable them to fulfill their explanatory and predictive roles.

To motivate the unusual complexity of *SR*, it will be contrasted with two popular versions of realism.

Contrasting Formulations

According to Bas van Fraassen (1980), realism is the thesis that science aims at truth and that acceptance of a theory includes believing that it is true (see Empiricism). Notice that this version acquiesces totally in realism's propensity for vagueness as to what, exactly, one is to be realist about. Van Fraassen's realist is not explicitly committed to anything epistemic or metaphysical. He is, however, semantically profligate; truth is the defining concept. He is also, like Hacking, committed to a particular reading of scientific practice. Van Fraassen's realist need not believe that any theory is true but that the scientist believes this in accepting a theory. *SR*, by contrast, is a thesis about what science achieves, regardless of its aims or epistemic attitudes. Science's predominant interest and direction could be nonepistemic, so far as *SR* is concerned. They could be indecipherable or as diversified as the proclivities of individual scientists. Unwilling to defer to the supposed values of science as an institution, *SR* must itself specify the conditions for metaphysical and epistemic commitment.

An older realist tradition, associated with Hilary Putnam (1978) and Richard Boyd (1984), holds that successful theories are approximately true (see Putnam, Hilary). The qualification leaves room for theory change and refinement. This version is epistemically profligate. *SR* achieves the same flexibility with much less generosity. A theory does not have to be "approximately" true or "close" to the truth to be empirically successful. It can contain lots of falsity that fails to make a difference at the observational level. So *SR* endorses only those entities and only those of their property attributions needed to make sense of what transpires at that level. *SR* is irreducibly semantic and epistemic because it is difficult to specify what these entities and properties are. They must be determined case by case through analysis of the evidence, of how the evidence is predicted or explained, and of whether alternative modes of prediction and explanation, invoking different entities or properties, are open.

Arguments for Realism

The Burden of Argument

Debates over realism are complicated by disagreement, usually implicit, about the argumentative burden. Does it fall upon the antirealist because realism's endorsement of science gives it an initial plausibility? This appears to be the presumption of Philip Kitcher (1993), who thinks that realism wins if antirealist arguments are countered. Or does it fall upon the realist because realism makes the stronger epistemic commitment? This is generally the view of antirealists, notably van Fraassen, who recommends antirealism for its epistemic minimalism. Since many realists, in the tradition of Putnam and Boyd, accept the burden of argument, it seems appropriate to begin with pro-realist argumentation.

This priority does not, however, reduce the antirealist's burden to criticism. Because of an important asymmetry in the opposing positions, antirealism requires an independent line of argument. It was noted that theoretical entities are given varying and changeable roles within science. Their ontological standing is subject to dispute, uncertainty, and investigation. A good example is electron orbitals. The original quantum theory required them, and experimentalists have claimed to observe them (Scerri 2001). Yet quantum mechanics denies their existence; only stationary states of the atom as a whole are real (see Chemistry, Philosophy of). Realism reflects this diversity in circumscribing the range of its intended applicability. Antirealism, by contrast, counsels a sweeping ontological skepticism with regard to theoretical entities as such, apart from any scientific issues affecting their interpretation. Only autonomous philosophical reasoning, presumed to trump debates within science, could ground so indiscriminate a position.

The Explanationist Strategy

The main line of argument for realism is ''explanationist'' or abductive. This means, essentially, that the explanatory achievements of theories count favorably in their epistemic evaluation (see Abduction). The need to advance theoretical hypotheses to explain empirical results is justification for believing the hypotheses, that is, for taking them to be true. In Putnam's version, if there were no truth to theory, one would be at a loss to explain, not just what is observed, but also the success of theory in explaining and predicting what is observed. In Boyd's version, truth attributions to theory derive from the success of the scientific method in generating empirically successful theories. To explain the success of method, one must suppose it grounded in a fundamentally correct picture of nature.

As it stands, this line of argument is seriously deficient. Just as realism is selective in the entities and properties it endorses, so must it be selective in the explanatory and predictive successes in which it invests epistemic import. For many, if not all, such successes admit of nonrealist explanation. Scientific method recognizes this. Scientists do not claim epistemic warrant from all the empirical results that a theory predicts. Experiments capable of warranting a theory must be carefully designed to make errors in the theory detectable. If this cannot be done, the theory is not warranted, however rich its explanatory and predictive success.

Thus, it was clear to Mendel that his laws of dominance and segregation, and the unobservable hereditary particles they introduced, got no support from the ratios of dominant and recessive traits in third-generation plants (see Genetics). That is why he designed the ''backcross test,'' which predicted new ratios uninvolved in the construction of those laws. Watson and Crick did not recommend their helical model of DNA for the amount of water that it accommodated, although the amount of water accompanying DNA was important information that an acceptable model would have to provide. They needed x-ray photographs. Some empirical information is built into theories. Some is predictable by rival theories. Some is explainable by theories—in geology and evolutionary biology, for example—that could also be used to explain contrary results. An explanatory argument for realism must identify those scientific successes that realism is needed to explain.

Novel Prediction

The most promising strategy is to focus on predictive novelty (Leplin 1997a). Scientific practice accords special probative weight to a theory's successful prediction of results that were unknown, unexplained, unappreciated, unanticipated, uninvolved in the theory's construction, unrelated to previous tests, unlikely or counterindicated on the basis of rival theories, unpredictable apart from the theory—a host of loosely related ideas are encapsulated in the scientific conception of novelty. The prediction of gravitational deflection of starlight by general relativity is a paradigm case of a novel result highly influential in the evaluation of a theory by scientists. The prediction from Fresnel's wave theory of a bright spot in the center of the shadow cast by a circular disk in spherical diffraction is a favorite example of philosophers (Giere 1983).

These examples suggest that novelty matters because it represents discovery, which is the business of science. A successful novel prediction is generally the discovery of a new phenomenon. But as the variations in the cognates of the notion suggest, this simple desideratum is incomplete. Sometimes what is discovered is not the phenomenon itself but how it is to be explained. A known phenomenon may continue to be regarded as novel so long as it defies theoretical explanation. For philosophical purposes, an understanding of novelty is needed that will advance the distinctively epistemic interest of licensing theoretical belief, rather than the loose and variable understanding of the scientist concerned to learn about the natural world.

The twentieth-century philosopher who made the most of novelty was Imre Lakatos (1970) (see Lakatos, Imre). Novelty was his criterion of the progressiveness of science; new theories must make new predictions. This is instructively ambiguous: Is the newness in what is predicted or in the prediction of it? Lakatos's conception of progress was primarily methodological, not epistemic. As it takes no account of a theory's conceptual coherence or negative evidence, Lakatosian progress cannot assure an advance in knowledge or truth. What the realist wants from novelty is a property that requires realist explanation. A novel empirical result must be one whose successful prediction by a theory requires the existence of the theory's unobservable posits to explain.

For this purpose it is unnecessary that the result be previously unknown, nor even uninvolved in constructing the theory, for its involvement could have been innocent. Perhaps, although the theorist used it, it need not have been the case; its use was incidental in that its omission would not have affected the theory's development. Perhaps the result was like a superfluous premise in a valid argument. It may have had heuristic importance, but without it the same conclusion would eventually have been reached. Conversely, that a result be totally unknown, new in every respect, does not guarantee novelty. For the result might instantiate a general principle assumed in developing the theory, or be interchangeable in the theory's development with known results that were crucial in determining the theory's predictive capacities. In these cases, the theory's provenance is such that it would naturally predict the result whether its posits warrant a realist interpretation or not.

Novelty requires that a result have a certain kind of *independence* from the provenance of the theory predicting it. Neither the result nor any general law or principle that subsumes it can have a role in determining what the theory will predict. Then the realist can argue that the theory's successful prediction of the result was not foreordained, but requires special explanation. This form of independence captures the idea of unexpectedness or distinctiveness ingredient in the common conception of novelty.

A remaining barrier to supposing that realist explanation is needed is that the result might also be predicted by rival theories with respect to which it is also independent. The realist must not endorse incompatible theories, and might lack a principled basis for favoring one over others. A number of additional measures pertaining to the weight or quantity of evidence might be introduced to handle this problem. But the simplest expediency is to make it a condition of a result's novelty for a theory that no viable rival theory predicts the same result. Then the realist can argue that the only available explanation of the theory's predictive success is to be found in the theory's own epistemic merits. This is a *uniqueness* condition. It captures the idea that, not only is a novel result in fact unexpected, but it is one that scientists had no, even unrecognized, reason to expect.

This requirement is clearly historical. Either novel status is historically variable or it must be indexed to a particular state of science. Given the realist's epistemic purposes, it is best to relativize novelty to the theories available at a particular time. A result R is *novel* with respect to a theory T if it satisfies the independence condition with respect to T and, at the time that T first predicts it successfully, no viable rival to T also predicts it. Then, the later development of rivals that predict R affects not R's novelty for T but its epistemic weight in the evaluation of T. Empirical results that satisfy the independence and uniqueness conditions for novelty are the ones whose successful prediction the realist claims provide warrant for theoretical beliefs.

Defending SR

With this delimitation of the range of scientific successes to which to apply his explanationist strategy, the realist's argument for *SR* is straightforward. The theoretical entities and properties that a theory uses to achieve predictive success are *needed* to achieve this success if this success is novel, for novel success is explainable neither on the basis of a theory's provenance nor on the basis of rival theories. A novel predictive success provides some reason to think that the theoretical entities used to achieve it are real, for if they are fictitious, if the theory is wholly wrong in positing and describing them, then one is at a loss to understand how the theory manages to achieve this success.

Notice that realism's explanandum here is not a novel result as such, or its prediction, or the truth of this prediction. The result is a fact about the world; the theory explains that. The theory's prediction of the result is explained by the theory's semantic content and the logical relations of this content to a statement of the result. The explanation of the prediction's truth is that this is the way the world is observed to be. What the truth of the theory is invoked to explain is the complex relational fact that the theory predicts the result *successfully*; that is, it yields the prediction and the prediction proves correct. How the theory manages to do this is unexplained by its semantic content or the observable way of the world.

A number of philosophers have argued that the appeal to truth gains the realist no explanatory advantage. They point out that the reality of theoretical entities adds nothing to their explanatory resources. Reality, according to Hacking (1983, 54), is not part of a theory's explanation of anything. A theory never explains anything just "by being true" (Levin 1984, 126). Such objections are beside the point, because realism's explanandum is not that of the theory, but is a second-order fact about the theory.

That a theory is sufficiently rich to yield novel predictions and that some such predictions prove true might be unremarkable. At what point appeal to chance becomes intellectually inadequate is disputable. But surely a sustained record of novel success unblemished by failure invites theoretical belief. To persist in agnosticism in the face of such ever-mounting achievement would seem to be a position that refuses to acknowledge the very possibility of evidence against it, thereby preempting its own eligibility for support. It is also a position at odds with scientific method, which seeks explanation not only of properties of the physical world but also of the success of theories in predicting and explaining these properties. Understanding why theories work as well as they do and no better is necessary for improving them. Unless the antirealist thinks there are a priori reasons to prefer methods of inquiry different from those found to be successful in experience, the project of understanding the success of theories cannot very well be rejected. But to the extent that such success is novel, realism is the outcome of this project.

The Antirealist Response

The antirealist's response to the sort of argument sketched is to attack abductive reasoning. Van Fraassen claims that explanation is a pragmatic rather than an empirical virtue; that is, the reasons for which it is valued have nothing to do with the realist's epistemic aims. Explanatory success cannot betoken truth, because it is impossible for empirical evidence to decide whether a theory's claims about its unobservable posits are true. No evidence available in principle can distinguish a theory's truth from its utility and reliability in prediction. The weaker, nonrealist hypothesis that a theory is empirically adequate—that it gets the observable facts right—is the strongest hypothesis testable. As a stronger hypothesis, realism is untestable and cannot be warranted.

One might wonder whether anyone but the skeptic can accept van Fraassen's constraints on the reach of evidence. Van Fraassen himself is certainly prepared to advance theses, like the empirical adequacy of a theory, that are stronger than alternatives among which observations cannot discriminate. Why is the thesis that a theory gets the *observed*, rather than the observable, phenomena right not the strongest thesis warrantable?

But a more fundamental question to ask is whether van Fraassen's reasoning argues against abduction or just presupposes that abduction is illegitimate. The realist contends that a theory's predictive and explanatory success is evidence for it. Van Fraassen contends that only the empirical facts themselves—not the fact that they were predicted successfully or explained, nor their novelty—are evidence for anything, and these facts can have no differential bearing on a theory's truth as against its mere empirical adequacy. Van Fraassen has given no reason for this limitation of the scope of relevant evidence. But he might ask whether the realist is any better off. Why should how the empirical facts are predicted or explained be an evidential matter? The criticism would then be that it is incumbent upon realists to defend their expansion of the scope of evidence.

There is a naturalistic answer to this challenge. Explanatory reasoning is indispensable in ordinary life. People leave the room by the door rather than the window, not because over extended trials people have had better results the one way than the other, but because they believe general laws that explain experience. Neither enumerative nor eliminative induction without abduction can make sense of the warrant for commonplace, practical decisions. Explanatory reasoning is equally indispensable within science. The progress of modern science depends squarely on the "method of hypothesis," the willingness to advance theoretical hypotheses and to judge theories by their explanatory resources. Recognizing that the methods and standards for investing credence have no a priori warrant, the naturalist relies on those that have been found successful in experience. Abduction works.

Arthur Fine (1984) thinks that this defense of abduction is circular. Because abduction is used in science to decide what conclusions to draw about the world, it must not be used in the philosophy of science to judge the epistemic status of these conclusions. This would be to *pre*judge their legitimacy. If philosophy is to judge the methods of science, it must not presume the reliability of those very methods. According to Fine, whether or not realism is correct, abduction is bound to support it, because realism is an explanatory hypothesis and abduction simply assumes that explanation betokens truth. If it does not, then the consistent abductive reasoner is systematically deceived. Only methods of reasoning that, unlike abduction, are potentially sensitive to a disconnection between truth and explanation could ground realism, because only they are capable of delivering a nonrealist conclusion.

There are a number of replies to this criticism, their theme being that Fine's requirements for a permissible realist argument are unreasonably stringent. Since abduction is used successfully in science to reach conclusions about observable phenomena, there is an immediate danger that Fine's requirements amount to skepticism. If theoretical beliefs are preempted by their abductive basis, so are many beliefs about observables. Could beliefs about observables be recovered on some other basis? Van Fraassen (1989) thinks so. But Leplin (1997a, Ch. 5) has argued that abduction is as fundamental as other forms of ampliative inference and that its inclusion in one's inferential repertoire is necessary to block paradox. In particular, enumerative induction is legitimate only for properties connected by an explanatory relation. More generally, the reply to Fine is that one cannot investigate scientific practice with methods more fundamental than those of science, because there are no more fundamental methods; indeed, there are no known methods at all of any plausibility and reliability that science does not already use. Hence, skepticism is the only alternative to the innocence of such circularity as there may be in the realist strategy.

Arguments Against Realism

Underdetermination and Empirical Equivalence

Because the relation of empirical evidence to theory is ampliative, any body of evidence is strictly consistent with the falsity of any theory. As it is not open to the antirealist to reject ampliation altogether—that produces skepticism—the antirealist cannot fault the realist for believing a theory on the basis of evidence that leaves open the possibility that some rival theory is true instead. Nevertheless, the possibility of unrefuted rivals has been used to mount an attack on realism. The key is to claim not only that any given body of evidence leaves alternatives to the realist's preferred theory open, but the stronger thesis of empirical equivalence:

> *EE*: Every theory T has a rival theory T' whose observable consequences are identically the same as those of T.

According to *EE*, the existence of T' is not merely a logical possibility but a fact. A multiplicity of theoretical options are invariably available, whatever the observable evidence. Moreover, observations used to support T cannot in principle discriminate T from T'; it is not just that they fail to do so presently, so that a warranted choice requires further evidence. Transitory evidential indecisiveness is a perennial situation; in it scientists attempt to design new experiments capable of discriminating among the contending theories. According to *EE* they may thereby eliminate some contenders, but can never reduce the options to a single theory. Frequently it is claimed that all theories have indefinitely many or infinitely many empirically equivalent rivals, so that crucial experimentation necessarily leaves a multiplicity of options (Kukla 1998, chap. 5). But if realism can contend with *EE* as formulated, these bolder versions need not be addressed. *EE* is a basis for claiming that theories are underdetermined by all possible evidence:

> *UD*: No amount of evidence suffices to warrant any theory.

It is supposed that, given *EE*, no evidence for any theory can support it over some rival (see Underdetermination of Theories). And it is further supposed that if rival theories are equally supported by the evidence, then a choice between them can be justified only pragmatically and cannot be epistemically warranted. If *UD* is correct, then realism—*SR* in particular—is indefensible.

As *EE* is stronger than the uncontested logical inconclusiveness of evidence, it requires argument. One approach is to cite cases. A Newtonian theory in which gravitation is fundamental may predict the same particle trajectories as an Einsteinian theory in which space-time is curved (Earman 1993). Standard von Neumann quantum mechanics may be empirically equivalent to a Bohmian formulation of quantum mechanics with hidden variables

(Cushing 1994). But the realist has no interest in denying the possibility of empirical equivalence, and it is unclear how a few real cases can generalize into anything like *EE*. What impresses one about real cases are the tentativeness and complexity of the judgment that empirical equivalence is irremediable, and the difficulty of formulating a rival theory that achieves it.

A more general argument exploits the leeway some important theories leave open for fixing certain of their parameters. Van Fraassen (1980, chap. 3) points out that any theory differing from Newtonian theory only in ascribing a nonzero constant absolute velocity to the center of mass of the universe is empirically equivalent to Newtonian theory. John Earman argues that the global topologies of some cosmological models allowed by general relativity are empirically indeterminable in principle. Rival hypotheses about the compactness of space, for example, will be empirically equivalent. As these examples depend on accepting a substantial body of theory, the realist should welcome them (Laudan and Leplin 1991; Stanford 2001). As shown, the realist has no interest in claiming that successful theories are to be believed in every detail or that every theoretical issue must be epistemically resolvable.

The most influential defense of *EE* is algorithmic. For example, define $T' = \neg T \wedge O(T)$, where $O(T)$ says that T's observable consequences are true. This approach substitutes artifactual cases for real ones to achieve generality. But it is unclear whether *EE* can be sustained artifactually. The antirealist cannot claim that the mere logical indecisiveness of evidence establishes *UD*, for then the result is skepticism. For *EE* to advance beyond the logical indecisiveness of evidence, the rival whose existence it guarantees must be a theory. Is $\neg T \wedge O(T)$ a theory? Impatient with the controversy this question stirs up (Leplin and Laudan 1993; Kukla 1998, Ch. 5), one might become inclined to dismiss it as "semantic." This would be a mistake, for what is really at issue is whether *EE* is capable of underwriting *UD* (Leplin 1997b). It is *UD* that threatens realism, not *EE* as such. If *EE* is read simply as asserting the consistency of $\neg T$ with $O(T)$ (or with all possible evidence for $O(T)$), if, in other words, $\neg T \wedge O(T)$ is to count as the rival theory that *EE* promises, then any underdetermination consequent upon *EE* is too sweeping for the antirealist to embrace. Either no underdetermination of real theoretical belief follows from *EE*, or any belief logically stronger than the observational evidence in hand is underdetermined. The first option sustains realism; the second is skepticism.

There is, in fact, no algorithm for generating theories, and it is by no means clear that theoretical alternatives are always available. And where they are available, there is no general impediment in principle to observational discrimination. Apart from the profligacy of its promise to deliver theories without end, there are two grounds for suspicion about *EE*.

The observational consequences of a theory depend on auxiliary information with which it is conjoined to generate predictions. One worry is that this information is inconstant, augmentable, and defeasible in unforeseeable ways. Equivalence with respect to one state of auxiliary knowledge may give way to decidability with respect to another. Unless the potential scope of auxiliary information is somehow delimited, equivalence has no guarantee. Even an equivalence of the idealized "total" theories of a final science is in principle amenable to changes in auxiliary knowledge (Leplin 1997b).

The second worry is that if *UD* were true, the theoretical auxiliaries needed to generate observational consequences from a theory would have no epistemic warrant. As different observational consequences are obtainable with different auxiliaries, the inability to privilege any specific auxiliaries leaves the theory's class of observational consequences indeterminate. But then *EE* itself is indeterminate, and cannot be used to infer *UD*. Thus either *UD* is false and there is no warrant for theoretical beliefs as realism contends, or *EE* is undecidable, in which case *UD* is unsupported. Either way, the antirealist's argument from underdetermination fails.

The Skeptical Historical Induction

A more serious challenge to realism is to be found in the historical record of once successful theories that science has come to reject. If successful theories have regularly proven to be false, then the success of a theory cannot be evidence that it is true. Larry Laudan (1981) has compiled a list of once prominent theories whose unobservable posits are unacceptable to contemporary science. Included are electromagnetic and optical ether theories, phlogistic chemistry, and the caloric and vibratory theories of heat. He invites the induction that the unobservable posits of contemporary science will in turn be rejected by future science, challenging the realist to explain why theoretical entities that happen to be current should be held exempt from the lessons of history.

Laudan's list, though embarrassing to realist intuitions, is not immediately generalizable. Even if many successful theories are false, success may yet indicate truth if its frequency among false theories is low. It may be that during the historical periods from which Laudan's examples come, virtually all theories were false. It is the small minority of successful ones that require notice. Perhaps success is far more frequent among true theories, which predominate in contemporary science, than among false ones. After all, the realist thinks not only that the best confirmed theories are at least partially true, but also that methods of theorizing and protecting against error have improved with scientific experience. It is a natural corollary to the realist position that truth is more frequent among current theories than past ones. But this is speculation. At best it prevents Laudan's argument from being definitive; it does not prevent it from shifting the burden to the realist.

A number of realist responses suggest themselves. The original response was Putnam's restriction of the abductive argument for realism to "mature" science. But of course the criterion of maturity cannot be currency. If realism is to claim any naturalistic grounding, if it is to be responsive to the historical record rather than imposed a priori, then its delineation of what is relevant in this record must not disqualify negative evidence. There must be an independent standard that Laudan's examples fail, a standard of experimental objectivity or severity of testing, some methodological stricture whose violation disqualifies the examples as the sort of successes that warrant realist explanation. Perhaps it can be argued that the quantum-mechanical revolution extends even to the criteria for theoretical success, so that it is simply a mistake to credit Laudan's examples with success whatever their coeval estimation. Short of that, it will be difficult to disqualify major theories of nineteenth-century physics. But with that, it will be difficult not to forecast further changes of method that disqualify contemporary physics. If standards of success have shifted to the point that scientific judgments of it are unreliable, then current science cannot be privileged. Either way, the induction goes through.

A different response is to point to the presupposition of current science in the determination that past theories have proved wrong, for, in denying the existence of previous theoretical entities, Laudan assumes a contemporary perspective. But to assume this perspective is to declare current science true and referential. The realist can ask for no more.

Whether Laudan's examples can be reinstated on a basis less congenial to realism is a difficult issue largely unexplored. If past theories were simply refuted by further evidence, they may be pronounced false without endorsing successor theories. But then, they may no longer be pronounced, on balance, successful, and it is unclear that realism need apply to them. Alternatively, one might fashion a new induction from the fact that once-successful theories eventually proved unsuccessful to the eventual decline of current theories. Another possibility is to appeal to inconsistencies among successful theories to argue that some of them must be false regardless of which, if any, are true. Then success is no guarantee of truth, and the realist's abduction fails. But judgments of logical relations among theories have been historically variable; a definitive judgment usually requires a contemporary perspective. And if somehow sustained, such judgments will deplete Laudan's list, weakening the induction.

A variant on the induction due to Kyle Stanford (2001) finesses these problems. Stanford connects the induction to the thesis of evidential underdetermination by propounding a transitory form of empirical equivalence. Instead of claiming that an equivalent rival exists for any theory, he claims that any body of evidence supporting a theory supports some rival theory equally well. The rival need not be available; indeed, Stanford's induction to the existence of such rivals in general is based on cases in which it was *unavailable* over a substantial period of the theory's success. That is, he induces from cases of theory succession rather than theory competition. The development of science exhibits a pattern in which successful *unrivaled* theories are eventually replaced by new theories that predict and explain all the evidence that, before they existed, supported their predecessors. One is to infer that the evidence for the best current theories is equally supportive of some alternative theory as yet unidentified, so that the best current theories are underdetermined.

Some critics of realism respond to the difficulties of *EE* by founding *UD* on something weaker: At any particular time the observable consequences of T are the same as those of some rival or other, possibly of different rivals at different times (Kukla 1998, Ch. 5). The trouble has been that no general argument, such as the algorithmic strategy attempts, has been advanced for this thesis; it appeared as an ad hoc reaction to challenges to *EE*. Stanford's induction appears to supply such an argument, if only availability of the rival is not required.

There is a serious difficulty with the argument, however. Many philosophers and historians contend that successor theories do *not* predict or explain the evidence for their predecessors. There are explanatory losses as well as gains in theory change; according to the more radical voices (Kuhn 1962), successive theories address altogether different questions and cannot recognize any common body of evidence at all.

But even with Stanford's more traditional view of progress (see Scientific Progress), it is problematic to maintain that the evidence explained by a new theory supports this theory as well as it supported its predecessor. In paradigm cases the new theory posits different, more powerful predictive machinery whose acceptance depends on new empirical successes. The original evidence that warranted the old theory would not be sufficient to warrant the new. How, then, can it be decided that the *extent* of the support it offers the new equals that which it offered the old? The only ready answer is to assimilate the notion of warrant or evidential support to that of subsumption; both theories subsume the old evidence. But status as a consequence of a theory is not the sole determinant of an observation's evidentiary weight with respect to the theory. It is an error, which the inference from *EE* to *UD* commits and *SR* is formulated to avoid, to assume that all of a theory's predictive successes support it at all, let alone equally. Novel ones count more, for example. In effect, Stanford supposes that since evidence supporting a theory is recoverable by inference from any stronger theory, it must support a stronger theory as well. This thesis generates well-known paradoxes of confirmation (Hempel 1965, 3–51) (see Confirmation Theory; Induction, Problem of).

The skeptical induction is designed to be naturalistic: The success of science is supposed to warrant realism, whereas the actual record of successful theories manifestly does not. As an argument from history, the induction takes at face value historical judgments of successfulness. But of course these judgments respond to multiple desiderata, pragmatic as well as epistemic. The realist will wonder whether counterexamples from history exhibit the kind of success for which realism is the appropriate explanation, and whether, if they do, the realist commitments they require are actually defeated. Even if the skeptical induction refutes a realism that infers theoretical truth from success straightaway, it may fail against a more discriminating thesis like *SR*.

To tell against *SR*, the inductive basis must include the particular theoretical entities and properties that were needed to achieve predictive success. Only if these have regularly been rejected by subsequent science, so that they were not simply misdescribed or incompletely understood but nonexistent, is *SR* in jeopardy. Moreover, the realist may require that the success achieved be novel, on the grounds that otherwise realism is not needed to explain it. Whether a significant record of novel success may be credited to many of the theories and entities from which antirealism is induced is questionable. More to the point, where novel success was achieved it is questionable whether the entities invoked to achieve it are rejected by current science.

A number of realists, including Kitcher (1993), Leplin (1997a), and Psillos (1999), respond to the skeptical induction by claiming referential success for some of the posits of abandoned theories, or at least for some specific predictive and explanatory applications of these posits. In certain experimental situations, the use of the term 'dephlogisticated air' by practicing chemists referred to the oxygen present before them, which they misdescribed. In saying that dephlogisiticated air sustained combustion, it was oxygen that they were referring to. Psillos goes so far as to identify the luminiferous ether, Laudan's prime example, with the electromagnetic field. This is a principled but risky strategy for the realist. Admitting the legitimacy of the induction, it makes its stand on what the historical evidence proves to be.

The resolution of this question depends on how reference to unobservable entities is fixed. The realist's strategy assumes that referential success does not require descriptive success, or at least that it tolerates some descriptive failure; a theoretical entity can be real although the theory that posits it is wrong. The best current theory of reference, the causal theory originated by Saul Kripke (1980), is the natural foundation for this strategy. But apart from internal problems, this theory requires an ostensive access to the referent, which is unavailable in the case of theoretical posits. Evidently, some degree of descriptive accuracy is required for theoretical reference. Pending a better understanding of the theory of reference, it is unclear how much discarded science the realist must resurrect to answer the skeptical induction.

Jarrett Leplin

References

Boyd, Richard (1984), "The Current Status of Scientific Realism," in J. Leplin (ed.), *Scientific Realism*. Berkeley and Los Angeles: University of California Press, 41–83.

Cushing, James (1994), *Historical Contingency and the Copenhagen Hegemony*. Chicago: University of Chicago Press.

Churchland, Paul (1985), “The Ontological Status of Observables: In Praise of Superempirical Virtues,” in Paul Churchland and Clifford Hooker (eds.), *Images of Science*. Chicago: University of Chicago Press.

Earman, John (1993), “Underdetermination, Realism and Reason,” *Midwest Studies in Philosophy* 18: 19–38.

Fine, Arthur (1984), “The Natural Ontological Attitude,” in J. Leplin (ed.), *Scientific Realism*. Berkeley and Los Angeles, CA: University of California Press, 83–108.

Giere, Ron (1983), “Testing Theoretical Hypotheses,” in John Earman (ed.), *Testing Scientific Theories: Minnesota Studies in the Philosophy of Science*, vol. 10. Minneapolis: University of Minnesota Press.

Hacking, Ian (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Science*. Cambridge: Cambridge University Press.

Hempel, Carl (1965), *Aspects of Scientific Explanation*. New York: Free Press.

Kitcher, Philip (1993), *The Advancement of Science*. New York: Oxford University Press.

Kripke, Saul (1980), *Naming and Necessity*. Oxford: Blackwell.

Kuhn, Thomas (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kukla, André (1998), *Studies in Scientific Realism*. New York: Oxford University Press.

Lakatos, Imre (1970), “Falsification and the Methodology of Scientific Research Programs,” in Imre Lakatos and Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 91–196.

Laudan, Larry (1981), “A Confutation of Convergent Realism,” *Philosophy of Science* 48: 19–49.

Laudan, Larry, and Jarrett Leplin (1991), “Underdetermination and the Empirical Equivalence of Theories,” *Journal of Philosophy* 88: 449–472.

Leplin, Jarrett (ed.) (1984a), *Scientific Realism*. Berkeley and LA: University of California Press.

——— (ed.) (1984b). “Truth and Scientific Progress,” in J. Leplin (ed.), *Scientific Realism*. Berkeley and Los Angeles: University of California Press, 193–218.

——— (ed.) (1997a), *A Novel Defense of Scientific Realism*. New York: Oxford University Press.

——— (ed.) (1997b), “The Underdetermination of Total Theories,” *Erkenntnis* 47: 203–215.

Leplin, Jarrett, and Larry Laudan (1993), “Determination Underdeterred,” *Analysis* 53: 8–15.

Levin, Michael (1984), “What Kind of Explanation Is Truth?” in J. Leplin (ed.), *Scientific Realism*. Berkeley and LA: University of California Press, 124–140.

Maxwell, Grover (1962), “The Ontological Status of Theoretical Entities,” in Herbert Feigl and Grover Maxwell (eds.), *Minnesota Studies in the Philosophy of Science, Volume III: Scientific Explanation, Space and Time*. Minneapolis: University of Minnesota Press, 3–27.

Psillos, Stathis (1999), *Scientific Realism: How Science Tracks Truth*. New York: Routledge.

Putnam, Hilary (1978), *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul, lecture II.

Shapere, Dudley (1982), “The Concept of Observation in Science,” *Philosophy of Science* 49: 485–525.

Scerri, Eric (2001), “The Recently Claimed Observation of Atomic Orbitals and Some Related Philosophical Issues,” in Jeffrey Barrett and J. McKenzie Alexander (eds.), *PSA00, Part I, Contributed Papers*. East Lansing, MI: Philosophy of Science Association, 76–89.

Stanford, Kyle (2001), “Refusing the Devil’s Bargain: What Kind of Underdetermination Should We Take Seriously?” in Jeffrey Barrett and J. McKenzie Alexander (eds.), *PSA00, Part I, Contributed Papers*. East Lansing, MI: Philosophy of Science Association, 76–89.

van Fraassen, Bas C (1980), *The Scientific Image*. Oxford: Clarendon Press.

——— (1989), *Laws and Symmetry*. Oxford: Clarendon Press.

See also **Abduction; Empiricism; Instrumentalism; Observation; Scientific Models; Scientific Progress; Theories; Underdetermination of Theories**

REDUCTIONISM

Reductionism is the thesis that the results of inquiry in one domain—be they concepts, heuristics, laws, or theories—can be understood or are explained by the conceptual resources of another, more fundamental domain (Nagel 1961; Sarkar 1998). Thus chemistry is supposed to be reducible to physics (see Chemistry, Philosophy of); within physics, thermodynamics is supposed to be reducible to the kinetic theory (see Kinetic Theory); Mendelian genetics is supposed to be reducible to molecular genetics (see Genetics; Molecular Biology); and so on. Reductionism can be viewed both as describing a research strategy or heuristic (how research should be pursued) and as a claim that the results of such research justify the assertion that one domain is being reduced to another. This article will concern both these aspects.

From the mid-1930s to the mid-1970s, philosophers commonly regarded reduction as a relation among theories, theoretical vocabularies, and laws.

Nagel's (1961) influential analysis reflected the then contemporary linguistic orientations: One theory reduced to another if the theoretical vocabulary referring to its entities and properties were definable, and its laws (logically) derivable, from that of the other—connected by empirical identifications, correlations, or reconstructive definitions (see Nagel, Ernest). Essentially this model of reduction extends the deductive-nomological model of explanation to the situation in which the explanandum is itself a law (of the reduced theory) (see Explanation; Hempel, Carl Gustav). Nagel also added "nonformal," or pragmatic, conditions that had to be satisfied for a reduction to be scientifically valuable. These conditions are perhaps the most lasting contributions of Nagel's account, but in the 1960s and 1970s, they were seldom noted—the formal properties were of central concern.

Schaffner (1967) extended Nagel's account to allow approximations and "strong analogies" in connecting theories—situations in which theories did not match exactly after a putative reduction, agreeing in some predictions but diverging in others: when one theory succeeded another, or a higher-level theory was explained and corrected by a more exact lower-level account. Theories at different levels of phenomena could supposedly thereby be successively reduced to those of the lowest compositional level or most fundamental theory, indicating the derivative character and the *in principle* dispensability of the things reduced.

This familiar philosophical gambit should invite suspicion: Whatever else it accomplishes, an in principle claim reliably indicates that it has not yet been achieved in practice. Should science be satisfied with such in principle claims? If so, how are these ever to be established? If not, what alternatives should be explored (Wimsatt 1976)? Typically, "methodological" reductionists urge the superiority of reductionist aims but seldom discuss methods—no practical reductionist problem-solving heuristic, or anything from the supposedly irrelevant "context of discovery." These are bothersome lacunae: If scientific work is reductionist, it should be discernible in the practice of science.

This unitary account of reduction has long dissolved, leaving a polyphonic disunity. Wimsatt (1979), Hooker (1981), and Sarkar (1998) review the relevant literature. On the one hand, studies from different sciences have become needlessly decoupled, seldom citing one another; on the other, they have become more responsive to actual scientific practice—especially in biology. This article concentrates on biology, but the analysis is intended to extend to mechanistic explanations throughout the cognitive, social, and physical sciences (see Explanation; Mechanism).

The perceived unity of the Nagel-Schaffner account was an artifact of the focus on structural or logical rather than functional features, when interest in reduction served foundationalist aims of increasing philosophical rigor, epistemological certainty, and ontological economy. These philosophical goals seldom matched the goals of scientific research even within contexts in which that research explicitly involved the pursuit of reductions (Schaffner 1974; Wimsatt 1976). Sarkar (1998) distinguished three important types of scientific reductions (see also Nickles 1973):

- Intralevel reduction, including successional reduction (Wimsatt 1976), the type of reduction involved in theory succession;
- Abstract interlevel reduction, in which levels of organization are distinguished, and upper-level features explained, using lower-level ones; and
- Spatial interlevel or strong reduction, in which the levels of organization are defined compositionally in physical space.

Prior formal accounts commonly conflated these three different types of reduction. Scientific reductions of any of these sorts are not the global and complete systematizations traditionally envisioned by philosophers. They are usually partial, local, conditional, and context dependent—that is, dependent on the specific mechanisms involved, and their associated *ceteris paribus* conditions (for interlevel reductions), or on the character and conditions of approximation used (for intralevel reductions). Thus, reductions do not necessarily lead to the unification of the sciences into a whole that satisfies traditional philosophical criteria for coherence and unity (see Unity and Disunity of Science).

Intralevel reductions are common in mathematically expressed theories and models. The most important type of intralevel reductions are successional reductions (see below under "Intralevel Reductions"). These typically localize formal similarities and differences between earlier and later, or more approximate and exact, theories of the same phenomena through mathematical transformations, thereby aiding succession and elaboration of the later theory and delimiting conditions for safe and effective heuristic use of the former.

In abstract interlevel reductions, levels of organization are distinguished in an abstract space. For instance, in classical genetics before the molecular era (see Genetics), the genome was represented hierarchically from a single allele to multiple alleles at a single locus, to linkage groups, and finally to the entire multilocus genotype (Sarkar 1998, Ch. 5). Properties of the entire multilocus genotype (that is, the whole organism) were supposed to be reducible to those at lower levels of this hierarchy. In condensed matter physics (Batterman 2002), hierarchical models of physical systems in phase space provide another potential example of such reductions. In biology, abstract interlevel reduction is typically a prelude to spatial interlevel reduction (for instance, from classical genetics to molecular biology). However, in physics, these abstract reductions often occur in phase-space representations and, thus, do not show the same pattern. This type of reduction does not raise any unique significant philosophical concerns not shared by the spatial type of interlevel reduction and will not be discussed further here (for more discussion, see Sarkar 1998, Chs. 3 and 5); from here on, interlevel reduction will be taken to mean spatial interlevel reduction.

Spatial interlevel reductions are compositional—localizing, identifying, and articulating mechanisms that explain upper-level phenomena and entities, all represented as entities in physical space. Reductionist accounts in complex sciences are commonly interlevel—for instance, explaining Mach bands in terms of lateral inhibition in neural networks (von Bekesy 1967), behavior of genes in terms of DNA (Sarkar 1998), or gases in terms of clouds of molecules (see Kinetic Theory). Reductions of this sort have also traditionally been called *mechanistic explanations* in the literature (Nagel 1961) (see Mechanism). *Aggregativity*, the claim that the whole is nothing more than the sum of its parts, is also commonly associated with such interlevel reductions. It is the proper opposite to *emergence* (see Emergence). Like interlevel reductions, aggregative relations are compositional. But aggregativity requires more. System properties that are aggregates of parts' properties represent degenerate cases where the organization of parts does not matter: They are invariant over organizational rearrangements. It is—roughly—a reduction without a mediating mechanism. Mechanistic models often start with many aggregative simplifying assumptions but add organizational features as they develop. As with intralevel replacements, some things (the aggregates) seem dispensable, though for different reasons: They are not required *in addition* because they are "nothing more than" the reducing things.

Intralevel Reduction

In intralevel reductions, the representation of the system being investigated is not assumed to be hierarchical or otherwise organized into levels. For instance, in biology, during heritability analysis (Sarkar 1998, Ch. 4), phenotypic variation in populations is presumed to be at least partly explainable by, or reducible to, genotypic variation: No assumption is made about the structure and organization of the genotype. The reduction of geometrical optics to physical optics similarly involves no claim about organization (see Schaffner 1967); the reduction of Newtonian gravitation to general relativity concerns all of space-time and makes no claim about hierarchical organization (see Space-Time).

The most interesting type of intralevel reduction is successional reduction, where one theory succeeds another and the earlier theory gets reduced to the later one. These reductions relate theories or models of entities at the same level of organization or theories that are not level specific. They are relationships between theoretical structures where one theory or model is transformed into another (often via limiting approximations) to localize similarities and differences between them. Since such derivations involve approximations, they are not truth-preserving deductions (see Approximation). 'Derivations' in this context is not the logical notion of deduction: In one sense, it is weaker insofar as it requires weaker formal assumptions; in another sense, it is stronger because the approximations and idealizations involved typically make implicit empirical assumptions (Leggett 1987 provides a discussion, though limited to the context of physics).

A terminological issue about the use of 'reduction' in contexts of theory succession should be noted: Sometimes, it is said that the later, more exact, or more complete theory *reduces in the limit* to the other (Nickles 1973). Thus, special relativity 'reduces' to classical mechanics in the limit as $v/c \rightarrow 0$ (where v is the velocity of a body and c is the velocity of light); either by letting velocity of the moving entity, $v \rightarrow 0$ (a "real" engineering approximation for velocities much smaller than the speed of light, c) or by letting $c \rightarrow \infty$ (a counterfactual transformation yielding Newton's conception of instantaneous action at a distance). This usage of 'reduction' is nonstandard in scientific contexts and parasitic on that of reduction in mathematics to indicate the specification of a particular value to a variable that can range over a set of values or some other operation that shows that one problem is a special case of another problem. There is obviously no question of the predecessor theory explaining its successor.

Localizing similarities by successional reduction and differences between theories (in aspects not captured in the reduction and in the transformations used) serves multiple functions in the succession of the newer theory (Wimsatt 1976; Nickles 1973). It co-opts the evidence for, and legitimates the use of, the older theory where they agree (as $v \rightarrow 0$ in the case of Newtonian mechanics and special relativity); may establish conceptual connections between them (as $c \rightarrow \infty$ in the same case); and locates contexts to pursue confirmation, testing, and elaboration of the newer theory where they disagree. Limiting conditions show where the older theory is a valid approximation and how rapidly it breaks down.

Finally, successional reduction is a kind of similarity relation. A series of pairwise successional reductions will usually be intransitive: Differences accumulate in theoretical successions because of approximations and idealizations, ultimately becoming too great to manage. Also, reductionist transformations involve scientific work, and are not achieved gratuitously. Since their scientific functions usually require relating the new theory only to its immediate predecessor, one rarely goes any further. (There is no scientific relevance in tracing special relativity back to Aristotelian physics!) Thus, successional reduction is often intransitive by default, even when possible otherwise—which it often is not, because of cumulative differences. If mappings are too complex to construct transformations relating immediate successors, then reduction fails, and the older theory and its ontology may be discarded. Instead of reduction, there is theory replacement. So this kind of reduction can be eliminative—but characteristically only when it fails (Wimsatt 1976). Ramsey (1994) and Batterman (2002) elaborate such reductions, and Batterman's discussion of the role of singularities shows another way in which reductions can fail without being eliminative. Sarkar (1998) discusses approximations, and Wimsatt (1987) connects these issues to related uses of false models.

Interlevel Reduction

By contrast, interlevel reductions generally do not relate theories (Sarkar 1992). They are driven by referential identities (Schaffner 1967; Wimsatt 1976) or localizations (Bechtel and Richardson 1993; Sarkar 1998) between entities at the reduced and reducing levels, and not theoretical similarities. These identities or localizations comprise what were called bridge laws or "reduction functions" in traditional accounts of reduction. Darden and Maull (1977) envision the construction of a single theory tying together two domains or levels but do not require prior theories. Interlevel reductions explain phenomena (entities, relations, lawlike regularities) at one level through the operations of often *qualitatively* different mechanisms at a lower level—such qualitative differences sometimes lead to claims of emergence (see Emergence).

Such mechanistic (or "articulation-of-parts") explanations (Kauffman 1972) are paradigmatically reductionist in biology (Wimsatt 1976; Glennan 1996; Sarkar 1998). They are compositional—upper- and lower-level accounts are supposed to refer to the same thing. Unlike the similarity relations of successional reductions, these references are transitive across levels, though the explanations may not be. Levels are defined by spatial inclusion of parts in a whole. For instance, Mendel's factors were successively localized through mechanistic accounts

(a) in chromosomes by the Boveri-Sutton hypothesis (Wimsatt 1976; Darden 1991),
(b) relative to other factors (now genes) in the chromosomes by linkage mapping (Wimsatt 1992)
(c) to bands in the physical chromosomes by deletion mapping, and finally
(d) to specific sites in chromosomal DNA.

Identities and localizations are powerful hypothesis generators in the reductionist's heuristic toolkit, suggesting new predictions at one level from properties or relationships at the other, with ample cues for how to construct explanatory accounts (Bechtel and Richardson 1993; Wimsatt 1976). In contrast, more traditional "correspondence theories" (Kim 1966) lack these resources and look "empirically equivalent" to identificatory theories only in static ad hoc comparisons made after the fact. (Work in scientific discovery motivates more realistic dynamical rather than merely static accounts of these processes. They are often revealingly different.) When genes were tentatively localized to positions on chromosomes, it spawned a research program dedicated to the elucidation of what different regions of chromosomes did and consisted of. Subsequent identification of genetic specificity with DNA sequences led to projects of sequencing parts of and eventually the entire genomes of organisms (see Molecular Biology).

Localizations are not logically as strong as identities: If two entities are identical, then *any* property of one is a property of the other. But localizations preserve all relevant spatiotemporal properties of identities, and thus all of their local mechanisms. Consequently, contrary to Schaffner (1967), identities are not required for successful reductions. For identities or localizations, interlevel reductions are transitive when compositional claims preserve boundaries between entities, and functional localization fallacies (attributing a system property to what is only a very important part of that system) often result when they do not. Sarkar (1998) explicates this point about identities more formally, showing that claims incorporating localizations may have the form of conditional statements and yet allow reductions to be achieved; they need not have the form of biconditionals, which identities must.

Boundaries between entities sometimes change for good reasons, and explanations may be intransitive for other reasons (e.g., different interests at different levels). But transitivity of explanation (and its central connection to compositional relations) is still reflected in the *modus tollens* form: Failures to explain upper-level phenomena in lower-level terms are inevitably blamed (by reductionists) on incomplete or incorrect descriptions of the relevant system at one level or another, generating expectable mismatches.

As noted earlier, failed successional reductions may eliminate objects of an older theory, but failures of interlevel reduction make upper-level objects and theory indispensable—there is typically no other way to organize the phenomena. Interlevel reductionist explanation—successful or not—is never eliminative. A mythical philosophical invention, eliminative reduction, reflects older aims of ontological economies since abandoned (Wimsatt 1979). There is no evidence for such elimination in the history of science, and there is no reason—in terms of scientific functions served—to expect it in the future. Claims to the contrary (see, e.g., Churchland 1986) may arise through conflations of successional and interlevel reduction. Designers of optical instruments continue to use geometrical optics rather than physical optics, though they may need to make corrections due to diffraction or exploit the phenomenon of polarization in some contexts, for instance, to eliminate the effects of glare (Batterman 2002). Engineers designing combustion engines use thermodynamics, not the kinetic theory of matter.

Analyses of reduction presuppose (and should provide) correlative analyses of levels whose objects, properties, and relationships are supposed to be related. These typically show that robust (multiply detectable) higher-level entities, relations, and regularities (Wimsatt 1981) do not disappear wholesale in lower-level scientific revolutions. Conceptions of them transmute, add (or occasionally subtract) dimensions, or turn up in different ways but do not disappear (Wimsatt 1994). In the case of interlevel reductions, claims of eliminativism rest upon exaggerations of unrepresentative cases.

Importantly, reductionist explanatory mechanisms are not themselves—nor directly legitimated by—exceptionless general laws. Exceptionless generalizations would be unmanageably complex. Useful, simple, broadly applicable generalizations about composed systems are richly qualified with *ceteris paribus* exceptions explicable in terms of mechanisms operating under an open-textured variety of applicable conditions (Glennan 1996). Mechanisms are not pragmatically translatable into laws (Wimsatt 1976; Cartwright 1983) (see also Laws of Nature; Mechanism).

How successful is interlevel reductionism? Historically, as characterized here, it goes back to the mechanical philosophy of the seventeenth century, according to which properties of extended bodies were to be explained by the contact interactions of their constituent parts (Sarkar 1992). The mechanical philosophy was recognized to have failed as a universal epistemology for physics by the nineteenth century, but the reductionist program continued to be pursued, though only in a piecemeal fashion. Its major nineteenth-century achievement was the kinetic theory of matter (see Kinetic Theory), in particular, Boltzmann's mechanical interpretation of the second law of thermodynamics. Einstein's kinetic account of Brownian motion in 1905 was another major success (see Sarkar 2000). While details of these—and other—reductions in physics continue to be debated, the general philosophical point, that reductions have been successful in many areas of physics, remains correct.

The major success of interlevel reductionism since the twentieth-century, as often alluded to earlier in this article, has been in biology, with the advent of molecular biology in the 1940s and 1950s (see Molecular Biology). In chemistry, meanwhile, the putative reduction of chemistry to physics has been both defended and challenged (see Chemistry, Philosophy of). In the social sciences, interlevel reductionism is usually called *methodological individualism*; it too has been both defended and challenged (see Methodological Individualism). Unexpectedly, within physics, interlevel reductionism has become controversial, with quantum entanglement often being interpreted as presenting a challenge to reductionism because it denies any standard individuation of the parts (see Jaeger and

Sarkar 2003) (see Locality; Quantum Mechanics). (For more extended discussions of reductionism in contemporary physics, see Shimony 1987 and Batterman 2002.)

Multiple Realizability and Supervenience

The concept of a level of organization figures centrally in two concepts that have played a central role in discussions of reduction and reducibility in philosophy of psychology, and the lessons one would draw from case studies in biology or physics run counter to the antireductionist views commonly asserted there. The very stability of higher levels of organization, relative to the more rapidly changing dynamics at the lower level, essentially guarantee multiple realizability. The lower level characteristically has many more variables and exponentially more possible states than the more macro level; so, there will inevitably be many-one mappings between microstate and macrostate.

But the relative stability of the macrostates generates something even stronger, which Wimsatt (1981) calls "dynamical autonomy," whereby the vast number of dynamical changes at the micro level map to the same or to neighboring macrostates, else macrostate fluctuations would be constantly observed—as they are in fact with "between-level" phenomena like Brownian motion (Wimsatt 1994; Sarkar 2000). This means that the most effective way of making a macro change in a system is virtually always to manipulate the macro variables. Thus macroscopic causality is safe in a reductionist world, and multiple realizability is no argument against reduction, but a very robust feature of the natural world.

But it is characteristic of any interlevel compositional reductionist explanations that there are exceptions to the upper-level regularities when these are mapped in the most natural ways to the micro level. (Note that there may be more than one such way.) For instance, fluctuation thermodynamics essentially concern continuous fluids, with exceptions having the right statistical characteristics. These upper-level exceptions may be negligible, as in statistical mechanics, or much more common, for instance, in relations between classical and molecular genetics. Even very reliable mechanisms have a fairly large and distinct number of ways of breaking down (which grows with increasing complexity and may also be open-ended). This open-endedness arises naturally in a mechanistic perspective, since it is easy to conceptualize in a systematic way how any proposed mechanistic intervention would act, but not possible to list systematically all possible mechanistic interventions.

Claims of supervenience, roughly, hold that there can be no change at the upper, or macro, levelf without a change at the lower, or micro, level but that there is no other systematic relationship between levels. The intuitions driving supervenience in interlevel cases are that there are multiply realizable macro-level properties that have no systematic micro-level account, in part because it is thought that a micro-level account would be wildly disjunctive and that there would be an open-ended class of exceptions to any attempt at a lower-level unity (see Supervenience). The characteristics of mechanisms described earlier appear to satisfy these conditions, but without denying reductionist explanations—and the dynamical autonomy of macro-level variables gives even more of what is needed to justify the autonomy of the "special sciences." It might be that the search for the apocalyptic completeness that would eliminate exceptions has motivated the idea of supervenience (see below), but this is an artifact of thinking of reductions in terms of laws rather than in terms of mechanisms—at least in the compositional sciences.

As used by philosophers of psychology, the formulation of supervenience (whether interpreted ontologically or epistemologically) commonly utilizes the relation between macro states and micro states *as revealed in terms of some future complete apocalyptic physics and psychology* (see Psychology, Philosophy of; Supervenience). But this leaves the discussion in an embarrassing situation: Can it now even be said that anything is supervenient? Not having the future sciences, it would appear that nothing can be said at all.

Interestingly, another concept and approach due to Levins renders something like supervenience immediately usable and widely applicable. Levins' (1966) concept of a "sufficient parameter" was originally elaborated for levels of organization but applies more broadly:

> It is an essential ingredient in the concept of levels of phenomena that there exists a set of what, by analogy with the sufficient statistic, we can call sufficient parameters defined on a given level ... which are very much fewer than the number of parameters on the lower level and which among them contain most of the important information about events on that level. (428–429)

He then provides an example of a causally potent robust property derived and caused in multiple ways. Levins' approach avoids dubious, in-principle arguments and assumptions about future physics and psychology and substitutes a thoroughly heuristic methodology tolerant of approximations and exceptions and better fitting actual scientific practice (Wimsatt 1994).

Aggregativity and Emergence

Opponents of reductionism usually fear what Dennett (1995) calls "greedy reductionism": explaining upper-level things in lower-level terms without intervening mechanisms mediating emergence of qualitatively different phenomena at higher levels (as Dennett says, "without cranes"). Greedy reductionism goes with "nothing but" talk, as reflected in Sperry's (1976) worries that to reductionists "eventually everything is held to be explainable in terms of essentially nothing." But this does not happen in an interlevel reduction. One moves to smaller and smaller parts in successive reductions, but in each transition, much of the explanatory weight is borne by the *organization* of those parts into a larger mechanism that explains the behavior of the higher-level system. (Those parts' properties explanatorily relevant to the behavior of the larger mechanism also provide the basis for judgments of multiple realizability and functional equivalence—roughly, any part[s] realizing those properties will do.)

But what if some properties of the parts were also manifested by the system—and were invariant no matter how one cuts up or rearranges its parts? For such properties, organization would not matter. Such properties are picked out by the most general conservation laws of physics but apparently not by any other scientific generalizations. These properties meet very restrictive conditions: For any decompositions of the system into parts, they are invariant over appropriate rearrangements, substitutions, and reaggregations, and their values scale appropriately under additions or subtractions to the system. For these aggregative properties, one is presumably willing to say that the mass of an animal considered as a subject by an artist for an anatomical drawing is nothing more than the mass of its parts. And the blame is on the artist—not vanished emergent interactions—for any shortfalls in what is represented in the drawing (Wimsatt 2000).

Essentially all other systemic properties partly depend upon the organization of parts, are in that sense emergent, and, according to reductionists, are mechanistically explainable. Within practical scientific contexts, emergence and reduction are not usually regarded as opposites; and nonadditive, organizational, and context-dependent interactions are the domain of emergence (see Emergence). Given these differences, it is not appropriate for interlevel reduction to be tarred with the ontologically corrosive reputation of aggregativity. Nevertheless, assumptions of aggregativity do play a role in initiating a reductionist research program; this may account for the confusion of aggregativity with reductionism.

Multiple conditions for aggregativity, each requiring invariance of system properties under different decompositions and operations on the system's parts, make aggregativity a degree property. In developing explanations, one starts with simple models. Simpler theories—ignoring higher-order interactions—typically look more aggregative. Few properties are aggregative in all respects for all decompositions, but many are aggregative or approximately so for some. Such decompositions are particularly simple and fruitful—more nearly factoring systems into modular parts with monadic, intrinsic, context-independent properties.

Such decompositions show varying success for different problems. Decompositions with more solutions get more attention, and it becomes tempting to accept "nothing but" statements that are really context bound and approximate, as if they were truly general. Bad decompositions for a problem produce functional localization fallacies, biases, and conceptual confusions (Bechtel and Richardson 1993). Powerful reductionist problem-solving heuristics can systematically lead to decisions to ignore or underestimate context dependence (Wimsatt 1980). This is one bias that epistemologically successful reductionist research must self-consciously guard against. Analyzing complex systems often requires simultaneous use of multiple decompositions, boundaries, and contexts—treacherous fields for functional errors: Reductionist heuristics must be deployed with special care in all such contexts (Wimsatt 1974 and 1994).

William C. Wimsatt
Sahotra Sarkar

References

Batterman, R. W. (2002), *The Devil in the Details*. Oxford: Oxford University Press.

Bechtel, W., and R. Richardson (1993), *Discovering Complexity: Decomposition and Localization as Scientific Research Strategies*. Princeton, NJ: Princeton UP.

Cartwright, N. (1983), *How the Laws of Physics Lie*. London: Oxford University Press.

Churchland, P. S. (1986), *Neurophilosophy*. Cambridge, MA: MIT Press.

Darden, L. (1991), *Theory Construction in Science: The Case of Genetics*. London: Oxford University Press.

Darden, L., and N. Maull (1977), "Interfield Theories," *Philosophy of Science* 44: 43–64.

Dennett, D. (1995), *Darwin's Dangerous Idea*. New York: Simon and Schuster.

Glennan, S. (1996), "Mechanism and the Nature of Causation," *Erkenntnis* 44: 49–71.

Hooker, C. (1981), "Towards a General Theory of Reduction," *Dialogue* 20: 38–59, 201–236, 496–529.

Jaeger, G., and S. Sarkar (2003), "Coherence, Entanglement, and Reductionist Explanation in Quantum Physics," in A. Ashtekar, R. S. Cohen, D. Howard, J. Renn, S. Sarkar, and A. Shimony (eds.), *Revisiting the Foundations of Relativistic Physics: Festschrift in Honor of John Stachel*. Dordrecht, Holland: Kluwer, 423–542.

Kauffman, S. (1972), "Articulation-of-Parts Explanation in Biology and the Rational Search for Them," in R. C. Buck and R. S. Cohen (eds.), *PSA-1970: Boston Studies in Philosophy of Science* 8: 257–272.

Kim, Jaegwon (1966), "On the Psycho-Physical Identity Thesis," *American Philosophical Quarterly* 3: 227–235.

Leggett, A. J. (1987), *The Problems of Physics*. Oxford: Oxford University Press.

Levins, R. (1966), "The Strategy of Model Building in Population Biology," *American Scientist* 54: 421–431.

Nagel, E. (1961), *The Structure of Science*. New York: Harcourt, Brace, Jovanovich.

Nickles, T. (1973), "Two Concepts of Inter-Theoretic Reduction," *Journal of Philosophy* 70: 181–201.

Ramsey, J. (1995), "Reduction by Construction," *Philosophy of Science* 62: 1–20.

Sarkar, S. (1992), "Models of Reduction and Categories of Reductionism," *Synthese* 91: 167–194.

——— (1998), *Genetics and Reductionism*. New York: Cambridge University Press.

——— (2000), "Physical Approximations and Stochastic Processes in Einstein's 1905 Paper on Brownian Motion," in D. Howard, and J. Stachel (eds.), *Einstein: The Formative Years, 1879–1909*. Boston: Birkhäuser, 203–229.

Schaffner, K. (1967), "Approaches to Reduction," *Philosophy of Science* 34: 137–147.

——— (1974), "The Peripherality of Reductionism in the Development of Molecular Biology," *Journal of the History of Biology* 7: 111–139.

Shimony, A. (1987), "The Methodology of Synthesis: Parts and Wholes in Low-Energy Physics," in R. Kargon and P. Achinstein (eds.), *Kelvin's Baltimore Lectures and Modern Theoretical Physics*. Cambridge, MA: MIT Press, 399–423.

Sperry, R. (1976), "Mental Phenomena as Causal Determinants in Brain Function," in G. Globus, G. Maxwell, and I. Savodnik (eds.), *Consciousness and the Brain*. New York: Plenum, 173–198.

von Bekesy, G. (1967), *Sensory Inhibition*. Princeton, NJ: Princeton University Press.

Wimsatt, W. C. (1974), "Complexity and Organization," *Boston Studies in the Philosophy of Science* 20: 67–86.

——— (1976), "Reductive Explanation: A Functional Account," *Boston Studies in the Philosophy of Science* 32: 671–710.

——— (1979), "Reduction and Reductionism," in P. D. Asquith and H. Kyburg, Jr. (eds.), *Current Research in Philosophy of Science*. East Lansing, MI: Philosophy of Science Association, 352–377.

——— (1980), "Reductionistic Research Strategies and Their Biases in the Units of Selection Controversy," in T. Nickles (ed.), *Scientific Discovery*, vol. 2: *Case Studies*. Dordrecht, Holland: Reidel, 213–259.

——— (1981), "Robustness, Reliability and Overdetermination," in M. Brewer and B. Collins (eds.), *Scientific Inquiry and the Social Sciences*. San Francisco: Jossey-Bass, 124–163.

——— (1987), "False Models as Means to Truer Theories," in M. Nitecki and A. Hoffman (eds.), *Neutral Modes in Biology*. Oxford: Oxford University Press, 23–55.

——— (1992), "Golden Generalities and Co-opted Anomalies: Haldane vs. Muller and the *Drosophila* Group on the Theory and Practice of Linkage Mapping," in S. Sarkar (ed.), *Founders of Evolutionary Genetics: A Centenary Reappraisal*. Dordrecht, Holland: Kluwer, 107–166.

——— (1994), "The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets," *Canadian Journal of Philosophy* 20: 207–274.

——— (2000), "Emergence as Non-Aggregativity and the Biases of Reductionisms," *Foundations of Science* 5: 269–297.

See also **Approximation; Biological Information; Chemistry, Philosophy of; Biology, Philosophy of; Emergence; Explanation; Locality; Mechanism; Methodological Individualism; Molecular Biology; Neurobiology; Psychology, Philosophy of; Quantum Mechanics; Social Sciences, Philosophy of the; Supervenience; Unity and Disunity of Science**

HANS REICHENBACH

(26 September 1891–9 April 1953)

Hans Reichenbach was born in Hamburg, Germany, a seaport and open-minded commercial town. He went to the *Oberrealschule,* where, chiefly, mathematics, natural sciences, and modern languages were taught. As a result of this schooling, in his later life he became more interested in science than in history. Reichenbach initially studied engineering in Stuttgart for a year, before studying mathematics, physics, and philosophy at Munich, Berlin, and Göttingen, respectively. Among his academic teachers were the physicists and mathematicians M. Planck, A. Sommerfeld, P. J. W. Debye, and D. Hilbert, and the philosophers E. von Aster, E. Cassirer, A. Riehl, G. Simmel, C. Stumpf, and E. Husserl. Von Aster and Cassirer impressed him the most (for details of Reichenbach's life, see Gerner 1997).

As a student Reichenbach was one of the leaders of the Freistudenten movement. Students who were active in this movement were influenced by the Wandervogel and other groups of the *Jugendbewegung* (youth movement). Members of this movement went hiking, sang, played musical instruments, did folk dances, and, in general, tried to find a new way of life. Reichenbach gave numerous talks and published many articles in student journals such as *Die freistudentische Idee: Ihr Inhalt als Einheit* (Reichenbach 1977–1999, vol. 1, 108–123). Reichenbach's noncognitivist ethics stem from the ideals he acquired from the Freistudenten movement (see below).

In March 1915 soon after the beginning of World War I, Reichenbach enlisted in the German army. Military service damaged his health, and in September 1917, he left the army for a job in the Gesellschaft für Funktelegraphie, which developed radio-telegraphy technology of military importance. In November 1918 a revolution shook Germany, and Reichenbach joined the Sozialistische Studentenpartei Berlin. He became one of the leaders of the organization and helped draft its program. In 1919 he became one of the first students to attend Einstein's lectures on the theory of general relativity, which strongly influenced his philosophical development. In 1920 Reichenbach became assistant, and later *Privatdozent*, for physics at the Technische Hochschule Stuttgart. There he developed the main ideas of his philosophy and wrote, in addition to important articles, his most ingenious work, which was later translated as *The Axiomatization of the Theory of Relativity* ([1924] 1969). Six years later, Reichenbach succeeded, with Einstein's help, in becoming an *außerordentlicher Professor* (roughly, an associate professor) for natural philosophy in the faculty of natural science. The job paid poorly, and Reichenbach wrote many articles, frequently on relativity for newspapers and popular journals, to supplement his income. Relativity was a popular subject for almost everyone at this time, and Reichenbach gave lectures on the radio, which later appeared as popular books, *From Copernicus to Einstein* ([1927] 1942) and *Atom and Cosmos* ([1930] 1932).

In 1923 Carnap, who had previously corresponded only with Reichenbach, organized the first meeting of "exact philosophers" at Erlangen. Besides Reichenbach, also participating were K. Lewin, H. Behmann, Paul Hertz, and others. There were many subsequent conferences and congresses of "Logical Empiricists," as they later called themselves. From 1923 until his death, Reichenbach remained in close contact with Carnap (see Carnap, Rudolf; Logical Empiricism). After the meeting, they both tried to found a journal for exact philosophy. It was, however, not before 1930 that Reichenbach was successful in taking over the *Annalen der Philosophie*, which became *Erkenntnis*, initially coedited with Carnap and Neurath (see Neurath, Otto).

The years in Berlin were Reichenbach's most productive. In 1928, he published what was later translated as *Philosophy of Space and Time* ([1928] 1958), but, after this, his work focused on probability theory. This eventually led to the publication of *Wahrscheinlichkeitslehre* ([1935] 1949). In 1929 Reichenbach replaced Joseph Petzold as president of the Gesellschaft für empirische Philosophie, which was very active in organizing talks and discussions. A smaller informal group, called the Berliner Kreis [Berlin Circle], was the counterpart of the Vienna Circle and included the philosophers Kurt Grelling and Walter Dubislav, the psychiatrist Alexander Herzberg, and the psychologist Kurt Lewin (see Vienna Circle).

In 1933, Hitler became chancellor of Germany, and some months later Jewish and so-called half-Jewish public servants, including professors, were dismissed from civil service. Reichenbach had to leave the university because his father was of Jewish descent. The same year, Kemal Atatürk refounded the University of Istanbul and replaced most of the Turkish professors with emigrants from Germany. Reichenbach had to stop lecturing in Berlin during the summer of 1933, but fortunately he could start teaching in Istanbul that October. During his lectures in Istanbul, Reichenbach was accompanied by an interpreter who translated the lectures, sentence by sentence, into Turkish. Reichenbach did not suffer as a result of exile as many others did. He was so inspired by logical empiricism that the uncomfortable political situation did not stop his research. Reichenbach remained editor of *Erkenntnis* for some years, until Felix Meiner, the publisher, was forced by the German authorities to discontinue the journal. During this period, Reichenbach met philosophical friends at many congresses: 1934 in Prague, and 1935 and 1937 in Paris. With the help of Charles W. Morris, Reichenbach received a chair at the University of California at Los Angeles (UCLA) in 1938. There, for the first time in his life, Reichenbach found ideal conditions for scientific work: He felt he was in the biblical "Land of Promise" far away from the dreadful war taking place in Europe. At UCLA, he wrote some of his most important books (Reichenbach 1944, 1947, [1935] 1949 [with about 50% new text], 1951, and 1954). On April 9, 1953, Reichenbach suffered a heart attack and died some hours later.

Space and Time

Reichenbach's reaction to Einstein's lectures motivated his first book, translated as *The Theory of Relativity and A Priori Knowledge* ([1920] 1965). The main idea of this work is that Kant's synthetic a priori principles are not uniquely determined by human understanding and pure intuition, but can be freely chosen or replaced by alternative principles: "Kant's concept of *a priori* has two different meanings. First it means 'necessarily true' or 'true for all times', and secondly 'constituting the concept of the object' " (48). Reichenbach rejected the first but retained the second meaning. Similar to Poincaré, throughout his life, Reichenbach held the view that knowledge is composed by empirical evidence and a contribution of reason, consisting of conventional principles, definitions, or rules (see Poincaré, Henri; Conventionalism).

Reichenbach called these conventional principles "coordinating principles." Examples include the principle of special relativity and the principle of the Euclidean character of space. The coordinating principles can become incompatible with the observed facts. It is a task of the theoretical physicist to find a set of coordinative principles that leads to a consistent picture of nature in agreement with the empirical data (see Kamlah 1985, 161–162). In private correspondence, however, Schlick convinced Reichenbach that he had left Kantianism by writing this book (see Gerner 1997, 54) (see Schlick, Moritz). From that time on, Reichenbach regarded himself as an anti-Kantian, even if he inherited more of Kant's tenets than he admitted.

Until 1928 Reichenbach studied primarily space and time. He wrote two books on that subject (Reichenbach [1924] 1969, essentially the mathematical theory of four-dimensional [light-ray or first signal] geometry; and [1928] 1958, which is less technical). In *Axiomatization*, Reichenbach axiomatized space-time geometry using three kinds of objects: space-time points or events, world-lines, and the relation of causal influenceability of one event by another, the *signal relation.* With these concepts Reichenbach formulated axioms aimed at showing that the inertial systems of special relativity and the corresponding coordinate systems can be constructed uniquely. The mathematical details of Reichenbach's light-ray geometry will not be discussed here, but there are interesting epistemological remarks that fill some of the crucial gaps later left by the *Philosophy of Space and Time.* The following discussion treats the two books together.

In the early 1920s, Reichenbach developed a theory of *coordinative definitions*, which replaced coordinating principles. Unfortunately, Reichenbach makes some confusing remarks about how these definitions endow signs with meaning. An expression is defined either by pointing at things ("This is a horse," "This is a bicycle," etc.) or by uttering a sentence that contains, together with other words, the expression to be defined. The first is a so-called ostensive definition, the second, a linguistic definition. Ostensive definitions are used to teach children some of their first words, while linguistic definitions presuppose an existing language. Reichenbach ([1924] 1969) writes: "Physical definitions, therefore, consist in the coordination of a mathematical definition to a 'piece of reality' " (8), which suggests that coordinative definitions are ostensive definitions for Reichenbach. With the exception of the definition of the meter by a standard rod in Paris, however, all definitions given by Reichenbach are linguistic definitions. Reichenbach's paradigm for linguistic definition is Einstein's definition of 'simultaneity' (45): "Let a first signal be sent from A at the time t_1; reflected at B, let it return to A at time t_3. Then its time of arrival at B will have the following value: $t_2 = \frac{1}{2}(t_3 - t_1)$." This definition has a linguistic character, but since its *definiens* refers to physical entities, it indirectly coordinates a concept to certain things.

The role of Reichenbach's coordinative definitions is essentially the same as the role of Poincaré's conventions (see Conventionalism). Poincaré distinguished "crude facts" from "scientific facts" (Poincaré 1958, 115–122). Crude facts are described in a language already known (a kind of observational language). With the aid of conventions, statements about scientific facts must be translated into the language that describes the crude facts. Analogously, Reichenbach's coordinative definitions define the scientific terms in the language in which observations are described.

Reichenbach ([1924] 1969) calls these just "facts," but it is clear from section 1 of *Axiomatization* that he means *empirical facts.*

Reichenbach (1977–1999, vol. 2) considered the separation of factual statements and definitions as one of the most important tasks of philosophy of science:

> Coordinative definitions are used at many points in the study of physics. It is not always easy to recognize them as such and to distinguish them from assertions of facts, and some well-known scientific disputes stem from seeking empirical knowledge where definitions belong... . We have a certain freedom of employing definitions and facts. It is only when a definition is given in one place that another assertion becomes an assertion of fact; conversely, the second may be regarded as a definition, which makes the first into an assertion of fact. (161–162)

Reichenbach's treatment of physical geometry, measurement of length, and simultaneity illustrate this task, and he considered his axiomatized theory of light geometry as a model for the semantics of physics.

Since definitions are arbitrary, the semantic analysis of science can be performed in different ways by splitting a theory into definitions and descriptions of facts differently. This yields pairs of different *equivalent descriptions* that have the same truth value. Reichenbach ([1928] 1958) originally called this thesis the 'epistemological' or 'philosophical theory of relativity' (177) and later the 'theory of equivalent descriptions' (1951, 133). If it is assumed that the same language is used in all descriptions of the observational facts (Reichenbach's "facts" and Poincaré's "crude facts"), this restricts the arbitrariness of the choice of alternative descriptions. The nontrivial content of the theory of equivalent descriptions is that under this condition, different equivalent descriptions are possible. This position can be accurately labeled 'nontrivial conventionalism.'

Reichenbach's conventionalism arises from epistemological issues and differs from his ostensible follower Grünbaum (1973). If the observational language is extended by theoretical terms, the latter must be defined in a partly arbitrary way. Grünbaum demands beyond this that nonconventional terms are "intrinsic," which he does not identify with "observable." Since Reichenbach's followers no longer shared the historical presuppositions of Reichenbach's generation, they had difficulty recognizing that his *philosophical relativity* was something new and interesting. Reichenbach's analysis responded to the neo-Kantian and iconic realist positions. According to Kant, humans are equipped with a system of a priori concepts, categories, pure intuitions, and derived concepts, and science cannot be conceived outside of this conceptual frame. Thus, an arbitrary choice of any definition of simultaneity is impossible. According to iconic realism, there is a single true picture of the world, and any different image is false. True knowledge means to have this picture in one's mind. Seventeenth-century philosophers would have said that at least God sees the world as it really is, even if human beings are unable to obtain its true picture. Thus, there is no way to choose between two different equivalent descriptions of nature. At least one of them has to be false. In contrast to both neo-Kantianism and iconic realism, Reichenbach emphasized that arbitrary definitions are needed to describe the world. It is important to note that in Reichenbach ([1928] 1958) the "theory of relativity of geometry" appears in the context of the refutation of neo-Kantianism. Reichenbach's rejection of iconic realism will not be discussed here (but see Kamlah 1979).

Against Kant, Reichenbach first shows that geometry itself is conventional. One can always introduce physical forces, which are said to compress or expand measuring rods in such a way that the same rods, if undistorted, would lead to a given metric or geometry. As Einstein had already stated, Reichenbach emphasized the point that it is not geometry alone, but geometry plus physics that is empirically testable (Reichenbach 1953b). Yet, neo-Kantians could respond: "Since geometry is given by pure intuition, one must always choose physics in such a way that it agrees with geometry and with the empirical evidence." Reichenbach's response was to develop a theory of geometrical intuition that did not permit this response; this is one of his most interesting achievements ([1928] 1958, §9–13). Reichenbach distinguished *physical* and *mathematical visualization* (82). The former is the anticipation of a possible perception described by Helmholtz (1962): "By the much abused expression 'to represent' ... I understand ... that we are able to imagine the series of perceptions we would have, if something like it occurred in an individual case" (277). Reichenbach showed how a non-Euclidean world can be visualized, and therefore that intuition does not acknowledge only Euclidean geometry. For instance, consider a movie that shows how it is to live in a non-Euclidean environment. Such a movie is possible only if some physical assumptions are made: Light has to travel in straight lines and solid bodies must obey the condition that during transportation, their length adapts to the metric of the geometry that is simulated.

While watching the non-Euclidean movie, however, neo-Kantians may be unconvinced that these experiences are veridical. They could argue that solid bodies are shrinking or growing in a quite unusual way, while the true geometry is still Euclidean. Pure intuition forces the use of only Euclidean physical models for the world. Neo-Kantians are committed to *mathematical intuition*, which is used to produce a model or picture of a state of affairs in the mind as to how it really is and not how it appears. In Reichenbach's spirit one may show that, by using physical models, pictures may be used that are either much bigger than the portrayed objects (e.g., atoms) or much smaller (e.g., the planetary system). Hence, the principle of geometrical similarity, which was already used as an axiom by Wallis in the seventeenth century in physical modeling, is an example of the fact that the axioms of Euclidean geometry are unconsciously used as a constraint for visual thought experiments. Reichenbach ([1928] 1958) called this the *normative function of visualization* (42).

In spite of Reichenbach's frequent assertions that coordinative definitions are arbitrary and that, therefore, many equivalent descriptions of nature are possible, he admitted that, as a practical rule, a single description may be superior to all others. Reichenbach (1951) called this the *normal system* (137, 180). Thus, in geometry, physicists normally define congruence by the transport of rigid rods. Yet, in nature, only solid rods are found, and rigid rods are obtained by correcting for the influences of deformatory forces such as temperature change or elastic forces. Reichenbach ([1928] 1958) proposed a definition of rigid rods that he believed was unique. He first distinguished between *universal* and *differential forces* (13). Universal forces act on all bodies in the same way, while differential forces (like temperature change) act on them differently. Hence, differential forces can be measured by comparing bodies of different kinds, while universal forces depend on the assumed geometry. Utilizing universal forces can "explain" why solid bodies do not follow the metrics of that geometry. Therefore, if it is assumed that all universal forces disappear, the metric of the space can be determined. However, there is no guarantee that in each set of equivalent descriptions a normal system exists; if quantum mechanics is true, a *normal system* for microphysics cannot be found (see below).

Finally, Reichenbach's theory of *simplicity* constituted a progress compared with Mach and Poincaré (see Mach, Ernest; Poincaré, Henri). If two equivalent descriptions of a physical system differ in simplicity, this is irrelevant for their truth. They are either both true or both false, differing only in *descriptive simplicity*. If, however, they are not empirically equivalent, they differ in *inductive simplicity*, and the simpler description is more likely to be true than the other one (Reichenbach 1977–1999, vol. 2, 163–164).

Realism and Meaning

After leaving neo-Kantianism, Reichenbach claimed to be a realist for the remainder of his life, but it is unclear what kind of realist. There are many different epistemological theories that are called 'realism,' and the arguments used in any particular case must be studied carefully in order to find out what kind of realism is being defended by each proponent (see Realism).

For Reichenbach, the distinction between realism and logical empiricism was intimately connected with the explication of the concept of meaning. Meaning can be defined in different ways. According to Reichenbach (1938), there is the *truth theory of meaning* and the *probability theory of meaning* (among others not mentioned here). The first says:

> Two sentences have the same meaning, if they obtain the same determination as true or false by every possible observation. (3)

The second says:

> Two sentences have the same meaning, if they obtain the same weight, or degree of probability, by every possible observation. (ibid)

It cannot be said beforehand which of these explications of meaning is correct. A decision between them must be made, which Reichenbach (1938) called a *volitional bifurcation* (10). This decision is not arbitrary, since it leads to different consequences in each case. It is not difficult to see that the *probability theory of meaning* allows more sentences to be confirmed than the *truth theory of meaning*. Reichenbach erroneously attributed the latter to the Vienna Circle (see Vienna Circle), because of an alleged commitment to the slogan: "The sense of a sentence is the method of its verification" (see Verifiability). Since most theoretical sentences cannot be strictly verified in the sense that they follow logically from observations, those people who decide to accept the truth theory of meaning are practically incapable of surviving, since they cannot empirically verify any important sentence about events in the future. By definition, all these sentences will have the same meaning as a contradiction (70–71). Reichenbach preferred the *probability theory of meaning*, and he called those who took the same position "realists." Thus, probability is indispensable to knowledge. Reichenbach believed that his *realism* differed from *positivism*, since, for the latter, knowledge came from observation and logic alone. For Reichenbach, probability and induction are additionally needed.

Reichenbach defended this theory until about 1938. Later, in *Philosophical Foundations of Quantum Mechanics*, he adopts yet another kind of realism: realism by convention, which claims that humans are normally committed to the following principle:

> The laws of nature are the same whether or not the objects are observed. (Reichenbach 1944, 19)

Reichenbach (1977–1999, vol. 2) called this convention an 'extension rule' (249). If it is not accepted, a tree that an observer looks at may double if the observer's eyes close or the observer looks in another direction before returning to the original state of attention. Similar to other conventions, realism for Reichenbach (1953b) depended on a "volitional decision" (§II.8).

Probability and Causality

Probability remained central to Reichenbach's work from 1915, when he wrote his dissertation, until at least 1949, when the English edition of *Wahrscheinlichkeitslehre* was published. Reichenbach's dissertation defended the frequency theory of probability at a time when, at least in Germany, textbooks still used Laplace's *Théorie de la probabilité* as their model. Two years later, in his *Staatsexamensarbeit*, Reichenbach discussed the possibility that fundamental physical laws were statistical (Kamlah 1998, 35). Thus, Kant's concept of causality must be replaced by a statistical relation between cause and effect. In 1923 Reichenbach (1977–1999, Vol. 2) tried to work out in detail a theory of statistical causality (345–371). It may be the case that even with complete knowledge of all boundary conditions of a physical process, single events can be predicted only with a probability, $p < 1$. In this case causality is irreducibly statistical. Probability, however, does not only appear in natural laws. It also plays a predominant role in the process of corroborating scientific theories and in any kind of induction. Without a probabilistic theory of induction, natural science is not possible. Logic and observation alone are not sufficient.

Until 1933 Reichenbach thought that logical empiricism could not justify scientific knowledge, which consists predominantly of general sentences. At that time Reichenbach was a *logical probabilistic empiricist:* The theory of probability and induction was the third pillar on which knowledge rested, besides empirical evidence and logic (1977–1999, Vol. 5, 435–436; see also Kamlah 1985, 166–167). Every kind of induction rests directly or indirectly on the *principle* or *rule of induction*, which makes estimation of a probability from a series of events f^n possible. Reichenbach (1953a) formulated this rule in the same year:

> If in a finite section [of a series of events] given we have observed a certain frequency f^n [of events having a certain property *A*], we posit that the frequency on further continuation, will converge towards a limit f^n (more precisely: within the interval $f^n \pm d$). We posit this; we do not say that this is true. (466)

Reichenbach considered this rule to be metaphysical and, later, analytical in some sense. However, in the winter of 1932–1933, the idea of a pragmatic justification of induction suddenly came to him: It cannot be shown that the rule will lead to success, but there is nothing to lose by applying it. Either the rule will help, or nothing else will do. Therefore, the rule is treated as a posit or a wager, and every conclusion that is logically dependent on it will be a posit. He called those posits that are direct results of the rule *blind posits.* Reichenbach (1953a) illustrated their nature by analogy to decisions made in ordinary life:

> We are often confronted by similar situations in daily life. We want to reach a certain aim and we know of a necessary step, which we shall have to take in order to attain this aim, but we do not know whether this step is sufficient. The businessman who keeps his store well stocked, so that he can sell something when a customer comes in, ... the shipwrecked man who climbs a cliff, although he does not know whether a rescue ship will spot him—all these persons find themselves in an analogous situation; they satisfy the *necessary* conditions of reaching an aim without knowing whether the *sufficient* conditions are satisfied. (472)

However, in practical situations, the application of this rule is rather the exception. In most cases pre-existing knowledge is combined with the inductive appraisal of empirical evidence. Reichenbach (1938) called the result of such a procedure an 'appraised posit,' in contrast to a *blind posit*, with which any experience has to start (352).

Reichenbach believed that this pragmatic justification solved Hume's problem (see Induction, Problem of). He presented this justification to the

Vienna Circle in February 1933, days after Hitler's seizure of power in Germany. It was typical for the members of the Vienna and the Berlin Circle that they continued to discuss their philosophical problems seemingly unaffected by the dangerous political situation at that time. Reichenbach wrote to E. von Aster:

> It is only now that I feel to be entitled to defend a radical empiricism, since I have shown that the principle of induction does not contain any synthetic a priori components, and since I have been successful by applying probability logic and the concept of posit to give a satisfactory theory of statements about the future. (Quoted in Kamlah 1994, 191)

Reichenbach now no longer needed a metaphysical principle to justify induction, but the philosophical cost was high. Strictly speaking, there was no longer room for scientific knowledge in Reichenbach's epistemology, only a system of posits that could be used as directives for action.

The Direction of Time

After an ingenious attempt to solve the problem of the source of the directionality of time in 1925, with interesting but half-finished results, Reichenbach left work on this issue until late in his life. What interested him most was the difference between the past and the future. When Reichenbach died, he left an almost completed book that was later edited for publication by his wife, Maria Reichenbach. The book contains philosophical observations on subjects such as traces left by physical processes, human memory, the order of playing cards in a deck, registering instruments, the measure of information, and other processes that undergo temporal change.

The following discussion treats, though only qualitatively, the *principle of common cause*, whereas Reichenbach (1956) used the theory of probability to treat the matter exactly. He wrote:

> Suppose two geysers which are not far apart spout regularly, but throw up their columns of water always at the same time. The existence of a subterranean connection of the two geysers with a common reservoir of hot water is practically certain. (158)

He subsumed this example under the following rule:

> If an improbable coincidence has occurred, there must exist a common cause. (157)

A more precise formulation is obtained by considering two events or series of events *A* and *B* that are both improbable alone and very similar to each other at two different places and at the same time. There is a high probability that a third earlier event or series of events *C* is the common cause of both (see Figure 1, case 1). Reichenbach (1956, Ch. 4) then asked whether the time-mirrored counterpart also exists, the common effect *D* (see Figure 1, case 2). The answer is yes only if *A* and *B* share a common cause *C* (see Figure 1, case 3). Hence, there is another asymmetry between the past and the future besides those generally arising from physics.

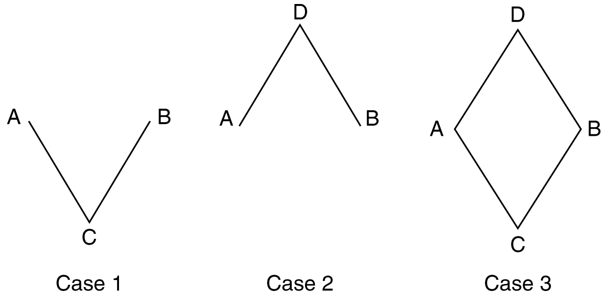


Fig. 1. Common cause and effect in a space-time diagram.

Chapter 4 of the *The Direction of Time* is full of ingenious but raw ideas waiting to be worked out by future philosophers; it is not a precise presentation of philosophical results. It is this character of the book, and of many other writings of Reichenbach, that makes his work valuable.

Logics

When Reichenbach taught logic, he looked for examples in natural language as exercises. In this way he reconstructed pronouns and tenses in first-order (predicate) logic. His theories of tenses and pronouns are well known as first attempts at formalizing natural languages with the aid of formal logics (see Reichenbach 1947, Ch. 7).

Quantum Mechanics

Twenty years after Einstein's special theory of relativity, a second revolution took place in physics. In 1925 Heisenberg's first paper on matrix mechanics was published, followed by Schrödinger's paper on wave mechanics in 1926 (see Quantum Mechanics). Reichenbach tried to understand this new microphysics, now called quantum mechanics. After 1938, he became intensively occupied with it and, in 1944, published *Philosophic Foundations of Quantum Mechanics*, which drew the attention of both physicists and philosophers.

Reichenbach (1944) did not significantly address the standard problem of the philosophy of quantum mechanics—the measurement problem (see Quantum Measurement Problem). Instead, he tried to integrate quantum mechanics into his *theory of equivalent descriptions.* Reichenbach thought that in quantum mechanics, a normal system (see "Space and Time" above) can no longer be found. Instead, the *wave* or the *particle interpretation* of quantum mechanics can be used (§8) (see Complementarity). The first construes Schrödinger's wave function as a real entity. Reichenbach does not clearly explain the second: Quantum systems are described as particles, but a formalism that would describe the quantum process in detail is not given. For Reichenbach, neither of the two descriptions can be considered a normal system. Only if one pays the price of accepting causal anomalies do either of the two interpretations describe the complete process.

To understand this point, consider the famous double slit thought experiment (Reichenbach 1944, 30). A particle starts from a source, passes through a diaphragm with two slits, and finally hits a screen or a photographic plate. If it passes slit *A* (see Figure 2), it will hit certain points on the screen with a certain probability. If the experiment is repeated many times, a certain density distribution of hits on the screen can be calculated using quantum mechanics. Now if slit *B* is also open, the density distribution will be different from that of an experiment in which both slits are open at different times (with the same duration). This causal process seems to violate the principle of action by contact. Slit *B* influences the particle in a nonlocal way (see Locality). The microprocess can thus be explained only if a *causal anomaly* is accepted.

Yet, electrons, protons, and other particles may behave as wave packets. A wave can pass both slits at once and interfere if both slits are open. If only one slit is open, the wave that leaves the diaphragm is different. At the moment the wave hits the screen, it suddenly shrinks to a point from which a new wave starts. Again, the principle of action by contact is violated. The breakdown of the wave is an instantaneous process. With infinite velocity, the wave is concentrated at one single point.

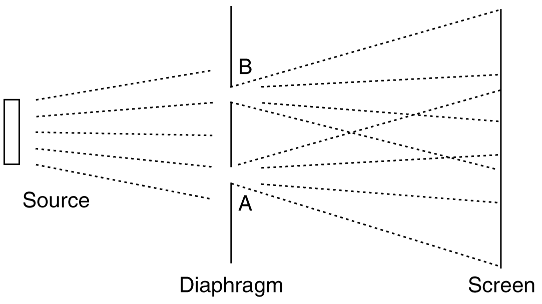


Fig. 2. The double slit experiment.

Reichenbach concluded that a choice must be made between two equivalent descriptions in which causal anomalies occur: the particle or the wave description. Neither of these can be called a normal description, since in both cases an important principle of everyday physics is violated. Reichenbach (1944) claims that Bohr and Heisenberg try to avoid this problem by using a third *restrictive interpretation* (33). They do not consider the *interphenomena*—those which happen between the measurements—as real physical processes. Therefore, statements about these *interphenomena* are meaningless for them. According to Reichenbach, they regard the quantum-mechanical formalism simply as an instrument for calculating predictions on the basis of past observations. Only observations really exist. It is also clear that this *restrictive interpretation* cannot avoid causal anomalies.

There is a fourth possibility that also agrees with the experimental findings. The microphysical *interphenomena* can be described in a language with a three-valued logic. Besides the truth values *true* and *false,* a third value, *indeterminate*, is introduced (Reichenbach 1944, 42). If a particle passes the diaphragm with two open slits *A* and *B*, it is *indeterminate* whether it goes through slit *A.* If a measurement of the position at the moment when the particle passes the slit is made and it is found in slit *A*, it is *true* that it is in *A.* Reichenbach's own interpretation of the third value, *indeterminate*, does not agree with the way he applied it, which suggested another explication (which, however, applies only if the quantum system is in a so-called pure state):

> Let U be the possible result of a measurement, then one defines: If the probability of the result that U will be measured as true is $p = 1$, U is indeed *true*, if $p = 0$, U is *false*, and if $0 < p < 1$, U is *indeterminate.*

This interpretation agrees more with the approach to this topic by many contemporary authors (see Reichenbach 1977–1999, vol. 5. 401–8). Even if Reichenbach's interpretation of quantum phenomena in terms of three-valued logics is problematic, it avoids causal anomalies with this logic. When the wave function changes instantaneously during the measuring process, many statements that have been *true* before now become *indeterminate,* and others that were *indeterminate* become *true* or *false.* There is no true statement in such cases, which does not accord with the *principle of action by contact.*

Ethics

Reichenbach, Carnap, and some other logical empiricists did not believe that a philosopher could establish a system of ethical rules that could be held universally obligatory. How can Reichenbach's radical position be understood? As mentioned before, Reichenbach was a student leader of the Freistudenten movement. The movement was characterized by anti-authoritarianism, and it is unsurprising that Reichenbach (1977–1999, Vol. 2) wrote:

> The individual may give his life whatever form he finds to be of value and may set for himself particular goals.... The individual may do whatever he considers to be right. Indeed he ought to do it; in general, we consider as immoral nothing but an inconsistency between goal and action. (110)

Later Reichenbach sent his children to the Montessori school in Berlin, where the main educational principle was that students should find their own way: Learn something if they want to do that and organize common activities themselves.

Reichenbach believed that his values were identical to those of many people in contemporary civilization. Thus, he says in *The Rise of Scientific Philosophy* (1951, 295): "I even have some fundamental moral directives, which, I think, are not so very different from yours." Reichenbach's belief that human beings will develop ethical directives useful for social life even if they are not taught to do so was later supplemented by logical analysis of ethical utterances. For him, these were not statements that can be true or false, but rather commands or directives, utterances of human will. These sentiments cannot be logically derived from empirical statements or from a priori principles. If a person commits a crime, that person will certainly be imprisoned if caught. The murderer's will conflicts with the will of the majority, but it *cannot be shown that the murderer is wrong.* Reichenbach was blind to the fact that in society there is vivid rational discourse on ethical questions. He cannot even discuss questions of justice based on his own standpoint, which, however, philosophers and others frequently do.

Reichenbach published only one discussion of ethics (1951, Ch. 17). Nevertheless, it is not possible to ignore this issue altogether. Reichenbach's writings frequently contain expressions like "arbitrary decision" and "volitional decision" in an epistemological context. These decision-indicating conventionalist terms have something to do with a deep-rooted desire for freedom, which is also found in Reichenbach's noncognitivist treatment of morals. Reichenbach had been a conventionalist in ethics even before he attended Einstein's lectures in 1919 and became a conventionalist in epistemology.

Andreas Kamlah

References

Gerner, K. (1997), *Hans Reichenbach. Sein Leben und Wirken*. Osnabrück: Phoebe Autorenpress.

Grünbaum, A. (1973), *Philosophical Problems of Space and Time* (2nd ed.). Boston, MA: Reidel.

Helmholtz, H. (1962), *Popular Scientific Lectures*. New York: Dover

Kamlah, A. (1979), "Hans Reichenbach's Relativity of Geometry," in W. Salmon (ed.), *Hans Reichenbach: Logical Empiricist*. Dordrecht, Netherlands: Reidel, 251–265.

——— (1985), "The Neo-Kantian Origin of Hans Reichenbach's Principle of Induction," in N. Rescher (ed.), *The Heritage of Logical Positivism*. London: University Press of America, 157–167.

——— (1994), "Hinweise des Nachlasses von Hans Reichenbach auf sein Menschenbild, auf Motive und Quellen seiner Philosophie," in L Danneberg, A. Kamlah, and L. Schäfer (eds.), *Hans Reichenbach und die Berliner Gruppe*. Braunschweig and Wiesbaden, Germany: Vieweg, 183–200.

——— (1998), "Die Analyse der Kausalrelation, Reichenbachs zweites philosophisches Hauptproblem," in H. Poser and U. Dirks (eds.), *Hans Reichenbach. Philosophie im Umkreis der Physik*. Berlin: Akademie Verlag, 33–53.

Poincaré, Henri (1958), *The Value of Science*. New York: Dover.

Reichenbach, H. ([1920] 1965), *The Theory of Relativity and A Priori Knowledge*. Berkeley and Los Angeles: University of California Press.

——— ([1924] 1969), *Axiomatization of the Theory of Relativity*. Berkeley and Los Angeles: University of California Press.

——— ([1927] 1942), *From Copernicus to Einstein*. New York: Philosophy Library.

——— ([1928] 1958), *The Philosophy of Space and Time*. New York: Dover.

——— ([1930] 1932), *Atom and Cosmos*. London: George Allen & Unwin.

——— ([1935] 1949), *The Theory of Probability* (2nd ed.). Berkeley and Los Angeles: University of California Press.

——— (1938), *Experience and Prediction*. Chicago: University of Chicago Press.

——— (1944), *Philosophic Foundations of Quantum Mechanics*. Berkeley and Los Angeles: University of California Press.

——— (1947), *Elements of Symbolic Logic*. New York: Macmillan.

——— (1951), *The Rise of Scientific Philosophy*. Berkeley and Los Angeles: University of California Press.

——— (1953a), "The Logical Foundations of the Concept of Probability," in H. Feigl and M. Brodbeck (eds.), *Readings in the Philosophy of Science*. New York: Appleton Century Crofts, 456–474.

——— (1953b), "The Verifiability Theory of Meaning," *Proceedings of American Academy of Arts and Sciences* 80: 46–60.

——— (1954), *Nomological Statements and Admissible Operations*. Amsterdam: North-Holland Publishing Company.

——— (1956), *The Direction of Time*. Berkeley and Los Angeles: University of California Press.

——— (1977–1999), *Gesammelte Werke in 9 Bänden* (Vols. 1–7). Edited by A. Kamlah and M. Reichenbach. Braunschweig and Wiesbaden, Germany: Vieweg.

——— (1978), *Selected Writings, 1909–1953* (Vol. 1–2). Edited by R. Cohen and M. Reichenbach. Dordrecht, Netherlands: Reidel.

See also **Carnap, Rudolf; Causality; Conventionalism; Logical Empiricism; Poincaré, Henri; Probability; Quantum Mechanics; Realism; Space-Time; Time; Vienna Circle**

REFERENCE

See **Putnam, Hilary; Quine, Willard Van; Russell, Bertrand**

RELATIVISM

See **Incommensurability; Kuhn, Thomas; Social Constructionism; Underdetermination of Theories**

RESEARCH PROGRAMS

Popper claimed that a theory is scientific if and only if it makes predications that can be tested by experiment or observation. Thus (Popper argued), genuine sciences, like physics, may be distinguished from pseudo-sciences, like astrology. On this basis Popper developed a cyclic picture of the scientific process: A theory is proposed; its empirical consequences are deduced; these consequences are empirically tested; some of these consequences are found to be false; a new theory is proposed that does not make these false predictions; this new theory is tested empirically; and so on. However, as it stands, this cyclic pattern allows scientists to repair refuted theories in an ad hoc fashion. One ad hoc strategy is to supplement a refuted generalization with a list of its known exceptions. Another is to hedge the theory with vague conditions such as 'normally' or 'other things being equal.' Popper needed to prevent such ad hoc repairs. To this end he specified a rule governing the replacement of an old, refuted theory by a new one. This rule employed the technical concept of *empirical content*. Popper defined the empirical content of a theory as the range of possible circumstances that the theory rules out. A typical scientific law has infinitely many such "potential falsifiers," that is, infinitely many ways in which it might be refuted. With this definition, Popper could specify the required rule: When a new theory replaces an old one, the new theory should have at least as much empirical content (so defined) as its predecessor.

In sum, a scientific discipline, for Popper, is one that does not use ad hoc maneuvers to preserve its theories from refutation at the expense of their empirical content. This apparently innocuous development permanently changed the character of normative philosophy of science. First, it required philosophers to consider the temporal sequence of theories. Popper's immediate predecessors had conceived the connection between a theory and the relevant evidence as an *atemporal* logical relation, like that between a mathematical theorem and the premises of its proof. In Popper's view, however, the reasons for accepting a theory T_n should include the fact that T_n is a content-increasing successor to T_{n-1}. This requirement, that philosophers should consider sequences of theories (T_{n-1}, T_n, T_{n+1}, etc.), helped pave the way for later collaboration between philosophers and historians of science. Second, Popper's original question, "Which theories are scientific?" has become "Which *sequences of* theories are scientific?" In other words, what makes for scientific respectability is the manner in which one develops and nurtures a cluster of thoughts over time, rather than any purely logical feature of an individual theory. Nevertheless, Popper remained convinced that individual theories may be assessed for verisimilitude. Thus in Popper's view the question of scientific status is put to sequences of theories, while the unit of epistemological appraisal remains the individual theory.

Imre Lakatos was a colleague of Popper's at the London School of Economics (see Lakatos, Imre). His early philosophical work was in the philosophy of mathematics, but in the late 1960s he turned to the philosophy of science. His target was the alleged relativism and irrationalism of Thomas Kuhn's *Structure of Scientific Revolutions* (see Kuhn, Thomas). Kuhn's model of scientific activity posed a problem for rationalist philosophers such as Popper and Lakatos because it seemed to offer a more historically accurate account of science than did any of its rationalist or realist competitors. Lakatos' aim was to offer a normative philosophy of science that combined Popper's rationalism with Kuhn's historical sense.

Lakatos attempted to solve the problem with his *methodology of scientific research programs*. One of the difficulties with Popper's view is that, in Lakatos' phrase, all theories are born into a sea of anomalies. That is, all scientific theories are, strictly speaking, false, if only because they are approximate. Moreover, most important scientific theories have "open problems" that, if they were to remain unsolved, would constitute refutations. But if all theories are false, how is one to know which to adopt as the current scientific orthodoxy? Lakatos' solution was to recognize sequences of theories (T_{n-1}, T_n, T_{n+1}, etc.) as the units of appraisal, rather than individual theories. All theories are false, but some sequences of theories seem to be heading toward the truth, while others seem to go nowhere. One should accept and support the former and abandon the latter. To use Lakatos' language, some sequences *progress* while others *degenerate*. His task was to specify criteria by which progress and degeneration could be distinguished. However, the Popperian notion of a sequence of contingently related theories is not adequate for this purpose, as it does not capture the thought that the same leading idea informs each successive theory of the sequence. In other words, one requires a unit of scientific appraisal that can maintain its identity as it changes.

Lakatos' suggestion for this role is the "research program." In his sense, a research program is the sum of the various stages through which a "leading idea" passes. This leading idea provides the "hard core" of the research program, that is, a set of commitments that cannot be abandoned without abandoning the research program altogether. Lakatos offers as an example the three laws of motion and the law of gravitation as the hard core of Newton's research program (Lakatos 1978a, 179). In addition to the hard core, a research program must have a "heuristic." A program's heuristic is its collection of characteristic problem-solving techniques. To continue with Lakatos' favorite example, the heuristic of Newton's program chiefly consisted in its mathematical apparatus: the differential calculus, the theory of convergence, and differential and integral equations.

What, then, is it for a program to progress? For Lakatos, change comes to a progressive program from its own inner logic, whereas a degenerating program changes in response to external criticism.

Newton first worked out his program for a planetary system with a fixed point–like sun and one single point–like planet. It was in this model that he derived his inverse square law for Kepler's ellipse. But this model was forbidden by Newton's own third law of dynamics; therefore the model had to be replaced by one in which both sun and planet revolved round their common center of gravity:

> Then he worked out the programme for more planets as if there were only heliocentric but no interplanetary forces. Then he worked out the case where the sun and planets were not mass-points but mass-*balls* . . . . Having solved this 'puzzle', he started work on *spinning balls* and their wobbles. Then he admitted interplanetary forces and started work on *perturbations*. (Lakatos 1978a, 50)

This is the paradigmatic research program: a sequence of theories representing stages in the development of a central idea. The important point is that the successive modifications were not forced on Newton by awkward empirical facts. Indeed, Newton scarcely glanced at the facts at all. Rather, the modifications were prompted by internal logical problems (such as the inadmissibility of infinite density in Newton's system), which were solved using the program's heuristic (i.e., Newtonian mathematics). Newton was able to ignore the fact that none of these models was empirically adequate because the program held out the promise that if he continued in the same vein, he would eventually produce a model that would not only respect the empirical evidence but also explain it.

This leads to the third and final element of the research program: a "protective belt" of auxiliary hypotheses. If empirical evidence requires changes to the program, they should be made here rather than to the hard core. In Newton's program, for example, this role was played by (amongst other things) geometrical optics and Newton's theory of atmospheric refraction (Lakatos 1978a, 179). Moreover, such changes must be "in the spirit of the heuristic" (ibid)—otherwise they would be ad hoc.

Thus, in a progressive research program, the central idea is developed and refined using the resources of the heuristic. Anomalies can and should be ignored in the hope that they will be accommodated and explained by a later stage of the program. That hope evaporates when the heuristic encounters problems that it cannot solve. Then, the program enters a degenerating phase, marked by ad hoc efforts to protect the hard core from criticism with the aid of devices external to the program. So long as successive changes to the protective belt of auxiliary hypotheses are in the spirit of the heuristic, the program is said to make *heuristic progress*.

Heuristic progress alone is not enough, however. A heuristically progressive program may ignore its anomalies, but it must also make *empirical progress*. Empirically progressive programs predict facts that are not only new but undreamt of. Lakatos' favorite example is Einstein's theory predicting that the distance measured between two stars would vary according to the time of day. Einstein's program gets as much credit for suggesting the experiment as it does for getting the prediction right (Lakatos 1978a, 5). In general, a program is *theoretically progressive* if it makes novel predictions, and it is *empirically progressive* if some of these are corroborated. Notice that the order of events matters: An empirically progressive program successfully predicts future empirical results, while *the same changes to the program* would signal degeneration if they came about after the results emerged.

This, in outline, is Lakatos' solution to the demarcation problem. Notice that the definition of heuristic progress is in fact a schema that gives a different specification for each program. Modifications to a program must be driven by its heuristic—but the meaning of this formula depends on the program in hand. Thus, the heuristic of a program plays a double role: It provides a logic of discovery for the scientists working on the program and at the same time sets the standard by which the program is to be evaluated (since fidelity to the spirit of its heuristic is a criterion of progress in a program). Thus, for Lakatos, the history of science is one of extended wars of attrition between research programs, some of which are progressing, while others are degenerating. A discipline is scientific so long as progressive programs triumph over degenerating ones.

The most effective critic of the methodology of scientific research programs was Paul Feyerabend, who claimed that any rigid account of the scientific method would eventually inhibit scientific progress (Feyerabend 1993, 14). Lakatos agreed, but he argued that his methodology was sufficiently flexible to escape this criticism. However, he had earlier argued that a program may recover from a degenerating phase to become progressive again *and* that degenerating programs should be starved of funds and support. Feyerabend had to point out only that Lakatos' methodology, if enacted as policy, would prevent a degenerating program from entering a new, progressive phase by killing it off too soon. The methodology of scientific research programs had another problem: Empirical progress requires the prediction of novel facts, but the

Copernican revolution made no such predictions. Rather, the heliocentric theory offered better explanations of already existing data. Thus, Lakatos was unable to explain why Copernicus' program was an improvement on Ptolemy. Lakatos was able to fix this problem by adopting an insightful suggestion by Elie Zahar. Zahar suggested that a program can achieve dramatic empirical success by explaining a fact that, though previously known, was not among the phenomena that the program was designed to explain. However, Zahar's suggestion, though insightful, was ad hoc with respect to Lakatos' philosophical program. By its own criteria, the methodology of scientific research programs fell into degeneration. This, together with the general decline of normative philosophy of science, may explain why it now has few adherents. Nevertheless, it will remain of interest so long as the question, "What is science?" retains its cultural and intellectual importance.

Brendan Larvor

References

Feyerabend, Paul (1993), *Against Method*, 3rd ed. London: Verso.

Lakatos, I. (1961), "Essays in the Logic of Mathematical Discovery," Ph.D thesis, Cambridge University.

——— (1976), *Proofs and Refutations*. Edited by J. Worrall and E. Zahar. Cambridge: Cambridge UP.

Lakatos, Imre (1978a), *The Methodology of Scientific Research Programmes*, in J. Worrall and G. Currie (eds), *Philosophical Papers*, vol. 1. Cambridge: Cambridge UP.

——— (1978b), *Mathematics, Science and Epistemology*, in J. Worrall and G. Currie (eds.), *Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press.

See also **Lakatos, Imre; Popper, Karl**

BERTRAND RUSSELL

(18 May 1872–2 February 1970)

"Almost everything that distinguishes the modern world from earlier centuries is attributable to science." With these words, Bertrand Arthur William Russell (1945) begins his chapter on "The Rise of Science" in his expansive *History of Western Philosophy*; and using this insight, he helped revolutionize not only twentieth-century logic, but philosophy itself.

Although known to many outside the academy primarily as a social critic, Russell contributed significantly to philosophy of science, in both a narrow and a broad sense. In the narrow sense, along with Gottlob Frege, Alfred North Whitehead, David Hilbert, and Kurt Gödel, Russell is responsible for groundbreaking work in two important subfields within twentieth-century philosophy of science: symbolic logic and philosophy of mathematics (see Hilbert, David). In the more broad sense, Russell is among those who originally championed connections among modern logic, mathematics, and science and between what he called the "scientific outlook" (Russell 1931) or "scientific temper" (Russell [1922] 1928, 172) and knowledge more generally.

Russell's most significant contributions include his refining and popularizing of Frege's predicate logic (upon which so many insights in philosophy of science have been based), his discovery of the paradox that bears his name, his introduction of the theory of types (his way of avoiding paradoxes), and his defense of logicism (the view that mathematics is in some important sense reducible to logic). Equally significantly, Russell, along with G. E. Moore, developed what would eventually become known as analytic philosophy, encouraging the view that science, mathematics, logic, and philosophy are all interconnected and all have something to learn from one another. Both in theory and in practice, Russell championed the view that progress in every intellectual sphere (science, philosophy, politics, education, etc.) depends on the scientific outlook, or as he also sometimes put it, "the liberal outlook" (Russell [1947] 1950, 15–16).

Born in 1872, Russell was the second grandson of Lord John Russell, the reform politician who twice served as prime minister to Queen Victoria. Following the death of his mother (in 1874) and of his father (in 1876), Russell and his brother went to live with their grandparents. Although Russell's father had granted custody of his sons to two atheists in order that they not be raised within the church, Russell's grandparents had little difficulty in getting his will overturned. Following the death of his grandfather (in 1878), Russell was raised primarily by his grandmother, Lady Russell. Educated at first privately and later at Trinity College, Cambridge, Russell obtained first-class degrees in both mathematics (1893) and the moral sciences (1894).

While at Cambridge, and like other later-to-be-famous intellectuals such as John Maynard Keynes, G. E. Moore, Henry Sidgwick, Alfred Lord Tennyson, Alfred North Whitehead, and Ludwig Wittgenstein, Russell became a member of the secret undergraduate association, the Society of the Apostles. As Sidgwick described it, "The essential value of the Society is a belief that we can learn, and a determination that we will learn, from people of the most opposite opinions" (Deacon 1985, 43).

Although appointed a Fellow of Trinity College in 1895 and elected to the Royal Society in 1908, Russell's career at Trinity appeared to come to an end in 1916 when he was convicted and fined for antiwar activities. He was dismissed from the College as a result of the conviction. The details of the dismissal are recounted in G. H. Hardy's book, *Bertrand Russell and Trinity* (Hardy 1942). The British government also refused to issue him a passport, which would have enabled him to lecture at Harvard. Two years later he was convicted a second time. This time he spent six months in prison. It was while in prison that Russell (1919) wrote his well-received *Introduction to Mathematical Philosophy* and began work on *The Analysis of Mind* (Russell 1921). In 1920 he traveled to both the Soviet Union and China. Although sympathetic to the socialist goals of the Bolsheviks, in the Soviet Union he became convinced that little good could come from the communist revolution. As he put it, "I see no reason whatever to expect equality or freedom to result from such a system, except reasons derived from a false psychology and a mistaken analysis of the sources of political power" (Russell [1920] 1962, 78–79). He did not return to Trinity until 1944, supporting himself and his family in the interim through a string of temporary lectureships and as a writer. Married four times and notorious for his many affairs, Russell also ran unsuccessfully for Parliament in 1907, 1922, and 1923. Together with his second wife, Dora Russell, he opened and ran an experimental school during the late 1920s and early 1930s. He became the third Earl Russell upon the death of his brother in 1931.

While teaching in California in the late 1930s, Russell was offered an appointment at City College, New York. The appointment was revoked following a large number of public protests and a judicial decision in 1940 stating that he was morally unfit to teach at the College. Nine years later, he was awarded the Order of Merit. He received the Nobel Prize for Literature in 1950. In making the award, the Nobel Committee explained that it was its intention to honor Russell "as one of our time's brilliant spokesmen of rationality and humanity, as a fearless champion of free speech and free thought" (Frenz 1969, 451). In 1953, perhaps recalling the 1940 New York court decision, Russell was also elected an Honorary Associate of the New York National Institute of Arts and Letters.

During the 1950s and 1960s, Russell spent much of his time encouraging scientists to participate in various antiwar and antinuclear activities. As a result, he became something of an inspiration to large numbers of idealistic youth around the world. Together with Albert Einstein, he released the Russell-Einstein Manifesto in 1955, calling for the curtailment of nuclear weapons. In 1957, he was a prime organizer of the first Pugwash Conference, which brought together scientists concerned about the proliferation of nuclear weapons. He became the founding president of the Campaign for Nuclear Disarmament in 1958 and was once again imprisoned, this time in connection with antinuclear protests, in 1961. Upon appeal, his two-month prison sentence was reduced to one week in the prison hospital. In 1960, he announced the launching of two foundations: the Bertrand Russell Peace Foundation and the Atlantic Peace Foundation. He remained a prominent public figure until his death 10 years later at the age of 97.

Russell's Contributions to Logic and the Philosophy of Mathematics

"Bertrand Russell's claim to be remembered by history," says his obituary in *The Times* of London, "rests securely on his work in mathematical and symbolic logic and in philosophy, on which his influence was pervasive and profound. The story of symbolic logic and of philosophy of mathematics in the twentieth century is the story of the expansion of the edifice which Russell and Frege founded. There have been major reconstructions, but they are reconstructions from within" ("Earl Russell" 1970).

This assessment is no doubt correct. In fact, when discussing Russell's intellectual life, it is useful to distinguish between events leading up to the publication of *Principia Mathematica*—Whitehead and Russell's (1910–1913) main contribution to logic and philosophy of mathematics—and those that came afterward. As Russell (1967–1969) himself has remarked in his *Autobiography*, completion of this project served as something of a turning point for him intellectually: "At the age of eleven, I began Euclid, with my brother as my tutor. This was one of the great events of my life, as dazzling as first love. I had not imagined there was anything so delicious in the world. ... From that moment until I was thirty-eight [the year the first volume of *Principia* appeared], mathematics was my chief interest, and my chief source of happiness" (37–38).

Principia Mathematica thus served as the culmination of years of work in both logic and mathematics. Along the way, Russell had accepted and then rejected idealism as the proper approach to mathematics; discovered what has since become known as Russell's paradox; introduced and then modified several versions of his theory of types; defended a new, robust version of logicism, thereby improving upon the work of Leibniz, Dedekind, and Frege; and refined and popularized Frege's newly invented quantificational logic.

His move to realism, not just in mathematics but generally, occurred only after first being exposed to the absolute idealism of J. M. E. McTaggart and F. H. Bradley at Cambridge. As Russell saw it, absolute idealism depended crucially upon the doctrine of internal relations, which holds that any relational fact (e.g., that x is related to y) is really a fact about the natures of the related terms. In this view, if x is greater than y, then being greater than y is a part of the nature of x. Object y is thus, in some sense, a part of x, and x is similarly a part of y. Given the complexity of relations in the world, it then turns out that all objects will be related to all objects. Hence there exists a single, all-encompassing unity. Further, since, if one is aware of x, x must also (in this view) be a part of one's mind, it follows that everything conceivable is a part of consciousness.

Initially, Russell used this idealist framework in his approach to mathematics (e.g., Russell 1897). However, he became disenchanted with idealism once he realized how incompatible the view was with his developing view of mathematics. In Russell's emerging view, geometrical points, for example, could be individuated only by their relations. But according to absolute idealism, relations depended in turn upon the individual, intrinsic natures of their relata. Forced to choose between idealism and mathematics, Russell unhesitatingly chose mathematics, replacing idealism with realism and replacing the doctrine of internal relations with a new doctrine of what he called "external relations," the view that relations, like objects, have a reality independent of the objects they relate. The change would lead, in the short term, to his next major work (Russell 1903) and, in the long term, to his commitment to pluralism, antipsychologism, the importance of science, and the emergence of analytic philosophy as a replacement for idealism.

It was in this context that Russell discovered, in the spring of 1901, the paradox that bears his name. In 1900, he had attended the Mathematical Congress in Paris and had been impressed by the work in mathematical logic by Giuseppe Peano and his students. As a result, he began studying both Peano's and Frege's logic in detail. A year later, he stumbled across a puzzling contradiction. The contradiction arose in connection with the set of all sets that are not members of themselves. Such a set (or collection), if it exists, will be a member of itself if and only if it is not a member of itself. To see this, it helps first to observe that some sets (such as the set of all Englishmen) are not members of themselves. In contrast, other sets (such as the set of all non-Englishmen) are members of themselves. It follows that if any property (or predicate) is sufficient to pick out a set in this way, then a new set, R, can be picked out using the property of not-being-a-member-of-itself. Immediately it follows that if R is a member of itself, then by definition it cannot be a member of itself; and if R is not a member of itself, then by definition it must be a member of itself.

The significance of the paradox follows, because in classical logic, all sentences are entailed by a contradiction. (For example, assuming both P and $\neg P$, one can prove any arbitrary proposition, Q, as follows: From P one can obtain $P \vee Q$ by the rule of addition; then from $P \vee Q$ and $\neg P$ one can obtain Q by the rule of disjunctive syllogism.) In the eyes of many mathematicians, including David Hilbert and Luitzen Brouwer, it thus appeared that no proof could be trusted. Once it was discovered that the logic and set theory apparently underlying all mathematics was contradictory, no mathematical claim was secure. A large amount of work throughout the early part of the twentieth century in logic, set theory, and the philosophy and foundations of mathematics was thus prompted. Russell wrote to Frege with news of his paradox on June 16, 1902. The paradox was relevant to Frege's logical work, because in effect, it showed that the axioms Frege was using to formalize logic were inconsistent. Specifically, Frege's rule 5, which states that two sets are equal if and only if their corresponding characteristic functions coincide in values for all possible arguments, requires that a propositional function such as $f(x)$ be considered both a function of the argument f and a function of the argument x. In effect, it was this ambiguity that allowed Russell to construct R in such a way that it could both be and not be a member of itself.

Russell's letter arrived just as the second volume of Frege's ([1903] 1964) *Grundgesetze der Arithmetik* was in press. Immediately appreciating the difficulty the paradox posed, Frege hastily added an appendix to the *Grundgesetze* to discuss Russell's discovery. In this appendix Frege ([1903] 1964) observes that the consequences of Russell's paradox are not immediately clear. For example, "Is it always permissible to speak of the extension of a concept, of a class? And if not, how do we recognize the exceptional cases? Can we always infer from the extension of one concept's coinciding with that of a second, that every object that falls under the first concept also falls under the second? These are questions," Frege notes, "raised by Mr Russell's communication" (127).

Because of these kinds of worries, Frege eventually felt forced to abandon many of his logical and mathematical views. Russell himself was also concerned about the paradox, and so, like Frege, he hastily composed an appendix for his soon to be released *Principles of Mathematics*. Entitled "Appendix B: The Doctrine of Types," the appendix represents Russell's first attempt at developing a workable theory of types. Russell's basic idea was that reference to troublesome sets (such as R) could be avoided by arranging all sentences into a hierarchy, beginning with sentences about individuals at the lowest level, sentences about sets of individuals at the next lowest level, sentences about sets of sets of individuals at the next lowest level, etc. Then, using the "vicious circle principle" (VCP) introduced by Henri Poincaré, together with his so-called "no class" theory of classes (Whitehead and Russell 1927, 37f, 187f), Russell was able to explain why propositional functions, such as the function "x is a set," should not be applied to themselves: because self-application would involve a vicious circle (see Poincaré, Henri). In other words, it is possible to refer to a collection of objects for which a given condition (or predicate) holds only if they are all at the same level or of the same "type."

Although first introduced by Russell in his *Principles*, his theory of types eventually found its mature expression in his 1908 article "Mathematical Logic as Based on the Theory of Types" and in *Principia Mathematica*. Thus, in its details, the theory admits of two versions, the "simple theory" and the "ramified theory." Both versions of the theory later came under attack. For some they were too weak, since they failed to resolve all of the known paradoxes. For others they were too strong, since they disallowed many mathematical definitions that, although consistent, violated the VCP. Russell's response to the second of these objections was to introduce, within the ramified theory, the axiom of reducibility. Although the axiom successfully lessened the VCP's scope of application, many claimed that it was simply too ad hoc to be justified philosophically (see Ramsey, Frank Plumpton).

Other responses to Russell's paradox include those of David Hilbert and the formalists (whose basic idea was to allow the use of only finite, well-defined, and constructible objects, together with rules of inference that were deemed to be absolutely certain) and of Luitzen Brouwer and the intuitionists (whose basic idea was that one cannot assert the existence of a mathematical object unless one can also indicate how to go about constructing it). Yet a fourth response was contained in Ernest Zermelo's 1908 axiomatization of set theory, **Z**. **ZF**, the axiomatization generally used today, is a modification of Zermelo's theory developed primarily by Abraham Fraenkel. All four responses helped logicians develop an explicit awareness of the nature of formal systems and of the kinds of metalogical results that are today commonly associated with them. Of equal significance during this same period was Russell's (1903) defense of logicism, the theory that mathematics is in some important sense reducible to logic. As he put it: "The fact that all Mathematics is Symbolic Logic is one of the greatest discoveries of our age; and when this fact has been established, the remainder of the principles of mathematics consists in the analysis of Symbolic Logic itself" (5). Initially, this meant showing that all mathematics was derivable from symbolic logic. It also meant discovering, so far as possible, the principles of symbolic logic itself. Later, in *Principia Mathematica*, it became clear that logicism would have to include two more precise theses. The first is that all mathematical truths can be translated into logical truths or, in other words, that the vocabulary of mathematics constitutes a proper subset of that of logic. The second is that all mathematical proofs can be recast as logical proofs or, in other words, that the theorems of mathematics constitute a proper subset of those of logic. Like Frege, Russell's basic idea for defending logicism was that numbers were to be identified with classes of classes and that number-theoretic statements were to be explained in terms of quantifiers and identity. Thus the number 1 would be identified with the class of all unit classes, the number 2 with the class of all two-membered classes, and so on. Statements such as "There are two books" would be recast as "There is a book, *x*, and there is a book, *y*, and *x* is not identical to *y*." It followed that number-theoretic operations could be explained in terms of set-theoretic operations such as intersection, union, and the like. In *Principia Mathematica*, Whitehead and Russell were then able to provide detailed derivations of many major theorems in set theory, finite and transfinite arithmetic, and elementary measure theory. A fourth volume on geometry was planned but never completed.

However, unlike Frege, Russell drew a quite different philosophical moral from the logicist reduction. As Frege saw it, if the principles of logic were understood to be self-evident, and if the laws of arithmetic can be shown to be derivable from them, arithmetic will have become epistemologically justified. According to this type of reduction, arithmetic would become just as certain as logic itself. In contrast, in Russell's view, the order of epistemic justification is exactly the reverse, as explained in the introduction to the first volume of *Principia* (Whitehead and Russell 1910–1913):

> But in fact self-evidence is never more than a part of the reason for accepting an axiom, and is never indispensable. The reason for accepting an axiom, as for accepting any other proposition, is always largely inductive, namely that many propositions which are nearly indubitable can be deduced from it, and that no equally plausible way is known by which these propositions could be true if the axiom were false, and nothing which is probably false can be deduced from it. If the axiom is apparently self-evident, that only means, practically, that it is nearly indubitable; for things have been thought to be self-evident and have yet turned out to be false. And if the axiom itself is nearly indubitable, that merely adds to the inductive evidence derived from the fact that its consequences are nearly indubitable: it does not provide new evidence of a radically different kind. Infallibility is never attainable, and therefore some element of doubt should always attach to every axiom and to all its consequences. In formal logic, the element of doubt is less than in most sciences, but it is not absent, as appears from the fact that the paradoxes followed from premisses which were not previously known to require limitations. (62)

Thus, like any of the other sciences, logic and mathematics are justified more by induction and coherence than by a *priori* reason. In the introduction to *Principia*'s second edition, Russell's comments are to much the same effect when he mentions the "purely pragmatic justification" of the axiom of reducibility (Whitehead and Russell 1927, xiv). This view of the relation between science and mathematics stayed with Russell throughout his career. For example, in his essay *Logical Atomism*, Russell (1924) again explains his position as follows:

> When pure mathematics is organized as a deductive system . . . it becomes obvious that, if we are to believe in the truth of pure mathematics, it cannot be solely because we believe in the truth of the set of premises. Some of the premises are much less obvious than some of their consequences, and are believed chiefly because of their consequences. This will be found to be always the case when a science is arranged as a deductive system. It is not the logically simplest propositions of the system that are the most obvious, or that provide the chief part of our reasons for believing in the system. With the empirical sciences this is evident. Electrodynamics, for example, can be concentrated into Maxwell's equations, but these equations are believed because of the observed truth of certain of their logical consequences. Exactly the same thing happens in the pure realm of logic; the logically first principles of logic—at least some of them—are to be believed, not on their own account, but on account of their consequences. The epistemological question: 'Why should I believe this set of propositions?' is quite different from the logical question: 'What is the smallest and logically simplest group of propositions from which this set of propositions can be deduced?' Our reasons for believing logic and pure mathematics are, in part, only inductive and probable, in spite of the fact that, in their logical order, the propositions of logic and pure mathematics follow from the premises of logic by pure deduction. I think this point important, since errors are liable to arise from assimilating the logical to the epistemological order, and also, conversely, from assimilating the epistemological to the logical order. (361–362)

According to the mature Russell, logic and mathematics differ from the natural sciences only as a matter of degree, not as a matter of kind.

Russell's Contributions to Philosophy of Science

In much the same way that Russell used logic in an attempt to clarify issues in the foundations of mathematics, he also used logic in an attempt to clarify issues in philosophy more generally. These attempts, together with his attempts to connect science to logic and philosophy, took center-stage following the appearance of the first volume of *Principia* in 1910.

As one of the founders of analytic philosophy, Russell made significant contributions to a wide variety of areas, including metaphysics, epistemology, ethics, and political theory, as well as to the philosophy of science. In the philosophy of science, he advanced discussion in a large number of areas, putting forward often-novel theories of mind, perception, causality, scientific method, and laws of nature. On the issue of causality, for example, Russell recognized that science demands more than simple observations of the form "*A* caused *B*" (see Causality). Instead, science regularly states causal laws of the form "*A* causes *B*." However, when it comes to applying this conjecture to real-world cases, the scientist does not mean that *A* always causes *B*, since there are always potential factors that may override *A*. For example, Newton's first law does not state that every object in a state of uniform motion will always remain in that state of motion. Rather, it states that every object in a state of uniform motion will remain in that state of motion unless it is acted upon by an external force. But how is the scientist to know whether such a law obtains if it is always possible to postulate such countervailing factors? As Russell (1948) puts it:

> We cannot take account of all the infinite complexity of the world, and we cannot tell, except through previous causal knowledge, which among possible circumstances would prevent *B*. Our law therefore becomes: "*A* will cause *B* if nothing happens to prevent *B*." Or, more simply: "*A* will cause *B* unless it doesn't." (474–475)

But as Russell points out, "This is a poor sort of law, and not very useful as a basis for scientific knowledge" (475) (see Laws of Nature). Russell's own solution was to rely upon a combination of scientific instruments (such as differential equations and statistical regularities) and metaphysical assumptions (concerning the quasi-permanence of logically constructed physical entities).

In books such as *The Scientific Outlook*, Russell (1931) also was one of the first to discuss the many social influences of modern developments in science and technology. Underlying these various projects lay not only Russell's use of logical analysis, but also his long-standing goal of discovering whether, and to what extent, knowledge is possible. Thus, Russell (1912) began the opening chapter of his *Problems of Philosophy* with the question "Is there any knowledge in the world which is so certain that no reasonable man could doubt it?" For Russell, this was the most central of all philosophical questions. Russell's various contributions to philosophy were thus unified by his views concerning both the centrality of scientific knowledge and the importance of an underlying scientific methodology common to both philosophy and science (see Carnap, Rudolf; Logical Empiricism). In the case of philosophy, this methodology expressed itself through Russell's use of logical analysis. So central was this methodology that Russell often claimed that he had more confidence in his methodology than in any particular philosophical conclusion.

As Russell explained it, his philosophical methodology consisted of the making and testing of hypotheses through the weighing of evidence (hence Russell's comment that he wished to emphasize the scientific method in philosophy), together with a rigorous analysis of problematic concepts using the machinery of first-order logic. For example, even in the application of elementary arithmetical concepts, one might come across difficult or marginal cases. As Russell (1942) explains,

> Two dogs and two dogs are certainly four dogs, but cases arise in which you are doubtful whether two of them are dogs. "Well, at any rate there are four animals," you may say. But there are microorganisms concerning which it is doubtful whether they are animals or plants. "Well, then living organisms," you say. But there are things of which it is doubtful whether they are living organisms or not. You will be driven into saying: "Two entities and two entities are four entities." When you have told me what you mean by "entity," we will resume the argument. (39)

In short, it was Russell's belief that conceptual analysis, together with the new logic of his day, would help philosophy exhibit the underlying "logical form" of natural language statements. A statement's logical form, in turn, would help philosophers resolve problems of reference associated with the ambiguity and vagueness of natural language. Thus, just as one can distinguish three separate senses of 'is' (the *is* of predication, the *is* of identity, and the *is* of existence) and exhibit these three senses by using three separate logical notations (Px, $x = y$, and $\exists x$, respectively), one will also discover other ontologically significant distinctions by being made aware of a sentence's correct logical form. In Russell's view, the subject matter of philosophy is then distinguished from that of the sciences only by the generality of philosophical statements, not by the underlying methodology of the discipline. In philosophy, as in mathematics, Russell believed that it was by applying logical machinery and scientific insights that advances would be made. Russell's most famous example of his analytic method concerns denoting phrases. In his realist *Principles of Mathematics*, Russell (1903) had adopted the view that every denoting phrase (for example, 'pi,' Sherlock Holmes, the number two, the golden mountain) denoted, or referred to, an existing entity. By the time his landmark article "On Denoting" appeared two years later, Russell (1905) had modified this extreme realism and had instead become convinced that denoting phrases need not possess a single theoretical unity.

According to this new view, logically proper names (words such as 'this' and 'that,' which refer to sensations of which an agent is immediately aware) always have referents associated with them. In contrast, descriptive phrases (such as 'the smallest number less than pi') should be viewed as a collection of quantifiers (such as 'all' and 'some') and propositional functions (such as 'x is a number'), which may or may not succeed in referring. As such, they are not to be viewed as referring terms but, rather, as "incomplete symbols." In other words, they should be viewed as symbols that take on meaning only within appropriate contexts but that in isolation remain meaningless. For example, consider the sentence 'The number between six and eight is prime.' Here the definite description 'the number between six and eight' plays a role quite different from that of a proper name such as 'seven' in the sentence 'Seven is prime.' Letting S abbreviate the predicate 'is a number between six and eight' and P abbreviate the predicate 'is prime,' Russell assigns the sentence 'The number between six and eight is prime' the logical form

There is an x such that
(i) Sx,
(ii) for any y, if Sy then $y = x$, and
(iii) Px.

Alternatively, in the notation of the predicate calculus, one has:

$$\exists x[(Sx \wedge \forall y(Sy \rightarrow y = x)) \wedge Px].$$

In contrast, by allowing s to abbreviate the name 'seven,' Russell assigns the sentence 'Seven is prime' the very different logical form Ps.

This distinction between logical forms allows Russell to explain three important puzzles. The first puzzle relates to true negative existential claims, such as 'The number postulated between seven and eight does not exist.' Here, by treating definite descriptions as having a logical form separate from that of proper names, Russell is able to give an account of how a speaker may be committed to the truth of a negative existential without also being committed to the belief that the subject term has reference. In other words, a claim such as 'Seven does not exist' is false, since the proposition $\neg\exists x(x = s)$ is self-contradictory. This is so because there must exist at least one thing that is identical to s, since it is presumably a logical truth that s is identical to itself. In contrast, the claim 'The number postulated between seven and eight does not exist' can be true because by allowing B to abbreviate the predicate 'is postulated between' and e to abbreviate 'eight,' there is nothing contradictory about the proposition $\neg\exists x(Bxse)$.

The second puzzle concerns the operation of the *law of excluded middle* and how this law relates to denoting terms. According to one reading of the law, it must be the case that either 'The largest prime is even' is true or 'The largest prime is not even' is true. But if so, since both sentences appear to entail the existence of a largest prime, this will clearly yield an undesirable result. Russell's analysis shows how such a result can be avoided. By appealing to his earlier analysis, it follows that there is a way to deny 'The largest prime is even' without being committed to the existence of a largest prime: by analyzing 'The largest prime' as a definite description rather than assuming it is equivalent to a proper name. Thus, the negation of 'The largest prime is even' becomes 'It is not the case that there exists a largest prime which is even,' rather than 'The largest prime is not even,' and the truth of the former can be asserted without controversy.

The third puzzle concerns the *law of identity* as it operates in so-called opaque contexts. For example, even though 'The positive square root of nine is three' is true, it does not follow that the two referring terms 'The positive square root of nine' and 'three' are interchangeable in every linguistic context. Thus although 'Alfred wanted to know whether the positive square root of nine was three' may be true, the similar sentence 'Alfred wanted to know whether three was three' is, presumably, false. Russell's distinction between the logical forms associated with the use of proper names and definite descriptions shows why this is so.

To see this, let t abbreviate the name 'three,' let n abbreviate the name 'nine,' and let P abbreviate the two-place predicate 'is the positive square root of.' It then follows that the sentence $t = t$ is not at all equivalent to the sentence

$$\exists x[Pxn \wedge \forall y(Pyn \rightarrow y = x) \wedge x = t].$$

Russell's emphasis upon logical analysis also had consequences for his metaphysics. In response to the traditional problem of the external world, which, it is claimed, arises because the external world can be known only by inference, Russell developed his famous distinction between "knowledge by acquaintance and knowledge by description" (Russell 1910). He then went on, in his lectures on logical atomism (Russell 1918, 1919), to argue that the world itself consists solely of a complex of logical atoms (such as "little patches of colour") and their properties. Together they form the atomic facts that, in turn, are combined to form logically complex objects. What are normally thought to be inferred entities (e.g., enduring physical objects) are then understood to be "logical constructions" formed from the immediately given entities of sensation with which people are directly acquainted, *viz.,* "sensibilia" (Russell 1914a and 1914b). It is only these latter entities that are known noninferentially and with certainty. Similar constructions allowed Russell to reduce points and instants to ordered classes of volumes and events, and classes to propositional functions.

According to Russell, a large part of the philosopher's job is to discover a logically ideal language. This logically ideal language will exhibit the true nature of the world in such a way that the speaker will not be misled by the casual surface structure of natural language. Just as atomic facts (the association of properties and relations with an appropriate number of individuals) may be combined into molecular facts in the world itself, such a language would allow for the description of such combinations using logical connectives such as 'and' and 'or.' In addition to atomic and molecular facts, Russell also held that general facts (about all members of a given class) were needed to complete the picture of the world. Famously, he vacillated on whether negative facts were also required.

Unlike Leibniz, Russell believed there was more to scientific, mathematical, and philosophical knowledge than getting the concepts right. Even so, just as Leibniz had wanted to develop a logically ideal language (or *characteristica universalis*), together with an instrument of universal deductive reasoning (his *calculus ratiocinator*), Russell believed that modern logic would help us represent the world as it actually is and reason about it without fear of error. As Russell himself translates one of Leibniz's most famous quotations:

> We should be able to reason in metaphysics and morals in much the same way as in geometry and analysis. . . . If controversies were to arise, there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates, and to say to each other (with a friend as witness, if they liked): Let us calculate. (Russell 1900, 169–170)

Conceptual analysis may not have the air of empirical discovery about it, but it is just as crucial to the scientific enterprise as observation and experiment.

With this observation in mind, it is thus worth recalling that after reviewing the history of the interplay between science and philosophy, Russell (1945) ends his *History of Philosophy* with these words:

> In the welter of conflicting fanaticisms, one of the few unifying forces is scientific truthfulness, by which I mean the habit of basing our beliefs upon observations and inferences as impersonal, and as much divested of local and temperamental bias, as is possible for human beings. To have insisted upon the introduction of this virtue into philosophy, and to have invented a powerful method by which it can be rendered fruitful, are the chief merits of the philosophical school of which I am a member. The habit of careful veracity acquired in the practice of this philosophical method can be extended to the whole sphere of human activity, producing, wherever it exists, a lessening of fanaticism with an increasing capacity of sympathy and mutual understanding. In abandoning a part of its dogmatic pretensions, philosophy does not cease to suggest and inspire a way of life. (836)

Andrew Irvine

References

Deacon, Richard (1985), *The Cambridge Apostles.* London: Robert Royce Ltd.

"Earl Russell, OM FRS" (1970, February 3). *The Times* (London) .

Frege, Gottlob ([1903] 1964), *The Basic Laws of Arithmetic.* Translated by M. Furth. Berkeley and Los Angeles, CA: University of California Press.

Frenz, Horst (1969), *Nobel Lectures: Literature 1901–1967.* Amsterdam: Elsevier Publishing Company.

Hardy, G. H. (1942), *Bertrand Russell and Trinity.* Cambridge: Cambridge University Press.

Russell, Bertrand (1897), *An Essay on the Foundations of Geometry.* Cambridge: Cambridge University Press.

——— (1900), *A Critical Exposition of the Philosophy of Leibniz.* Cambridge: Cambridge University Press.

——— (1903), *The Principles of Mathematics.* Cambridge: Cambridge University Press.

——— (1905), "On Denoting," *Mind* 14: 479–493. Reprinted in Russell, *Essays in Analysis*, London: George Allen and Unwin, 1973, 103–119.

——— (1908), "Mathematical Logic as Based on the Theory of Types," *American Journal of Mathematics* 30: 222–262. Reprinted in Russell, *Logic and Knowledge*, London: George Allen and Unwin, 1956, 59–102, and in Jean van Heijenoort, *From Frege to Gödel*, Cambridge, MA: Harvard University Press, 1967, 152–182.

——— (1910), “Knowledge by Acquaintance and Knowledge by Description,” *Proceedings of the Aristotelian Society* 11: 108–128. Reprinted in Bertrand Russell, *Mysticism and Logic*, London: George Allen and Unwin, 1963, 152–167.

——— (1912), *The Problems of Philosophy*. London: Williams and Norgate; New York: Henry Holt and Company.

——— (1914a), “The Relation of Sense-Data to Physics,” in *The Philosophy of Logical Atomism and Other Essays, 1914–19* (*The Collected Papers of Bertrand Russell*, vol. 8). London: George Allen and Unwin, 5–26.

——— (1914b), *Our Knowledge of the External World.* Chicago and London: Open Court Publishing Co.

——— (1918, 1919), “The Philosophy of Logical Atomism,” *Monist* 28: 495–527; 29, 32–63, 190–222, 345–380. Reprinted in Russell, *Logic and Knowledge*, London: George Allen and Unwin, 1956, 177–281.

——— (1919), *Introduction to Mathematical Philosophy.* London: George Allen and Unwin; New York: Macmillan.

——— ([1920] 1962), *The Practice and Theory of Bolshevism.* London: George Allen and Unwin.

——— (1921), *The Analysis of Mind.* London: George Allen and Unwin; New York: Macmillan.

——— ([1922] 1928), “Free Thought and Official Propaganda,” in *Sceptical Essays.* New York: W. W. Norton, 149–172.

——— (1924), “Logical Atomism,” in J. H. Muirhead, *Contemporary British Philosophers.* London: George Allen and Unwin, 356–383. Reprinted in Russell, *Logic and Knowledge*, London: George Allen and Unwin, 1956, 323–343.

——— (1931), *The Scientific Outlook.* London: George Allen and Unwin; New York: W. W. Norton.

——— (1942), *How to Become a Philosopher, How to Become a Logician, How to Become a Mathematician.* Girard, KS: Haldeman-Julius Publications.

——— (1945), *A History of Western Philosophy.* New York: Simon and Schuster.

——— ([1947] 1950), “Philosophy and Politics,” in *Unpopular Essays.* New York: Simon and Schuster, 1–20.

——— (1948), *Human Knowledge: Its Scope and Limits.* London: George Allen and Unwin; New York: Simon and Schuster.

——— (1967–1969), *The Autobiography of Bertrand Russell* (3 vols.), London: George Allen and Unwin.

Whitehead, Alfred North, and Bertrand Russell (1910–1913), *Principia Mathematica* (3 vols.). Cambridge: Cambridge University Press.

——— (1927), *Principia Mathematica* (2nd ed.). Cambridge: Cambridge University Press.

See also **Carnap, Rudolf; Causality; Hilbert, David; Logical Empiricism; Turing, Alan**

S

MORITZ SCHLICK

(14 April 1882–22 June 1936)

Friedrich Albert Moritz Schlick was born the third and youngest son of Protestant parents. His father, who owned a factory, descended from a Bohemian family of noble lineage, and his mother from the family of the poet Ernest Moritz Arndt, after whom Schlick was named. He attended primary and secondary school in Berlin. As a sickly schoolchild, Schlick was interested in philosophy, art, and poetry at the Luisenstädter Gymnasium, but he went on to study natural science and mathematics at universities in Heidelberg, Lausanne, and Berlin. In 1904, he completed his Ph.D. under Max Planck, who regarded him as one of his favorite students, with the thesis *Über die Reflexion des Lichtes in einer inhomogenen Schicht* [The Reflection of Light in an Inhomogeneous Layer] in mathematical physics. Schlick spent the following three years doing scientific research at the universities in Göttingen, Heidelberg, and Berlin. After the appearance of his first book, whose title is translated as *Life Wisdom* (Schlick 1908), Schlick spent two years studying psychology in Zurich. In 1907, he married Blanche Hardy and, after a short sojourn in Berlin, completed his habilitation in 1911 at the University of Rostock with a study "Das Wesen der Warheit nach der modernen Logik" (Schlick 1910).

During his next ten years of academic activity, Schlick worked on the reform of traditional philosophy against the backdrop of the revolution in natural science. He became friends with Einstein and was one of the first to study the theory of relativity from a philosophical perspective. During World War I, he served two years at a military airport.

In 1917, Schlick began teaching at Rostock, and was granted the official title of associate professor, with a teaching position in ethics and natural philosophy, in 1921. During the Weimar Republic, Schlick backed university reform as a member of the Union of Progressive Academics and completed his major study, *Allgemeine Erkenntnislehre* (Schlick [1918] 1974). In the summer of 1921 Schlick moved to the University of Kiel as full professor, where he taught for one year.

After the Rostock and Kiel periods, the then 40-year-old Schlick was appointed in 1922 to the chair for natural philosophy (philosophy of the inductive sciences) in Vienna, in the tradition of Mach and

Boltzmann (see Mach, Ernest). The mathematician Hans Hahn, who was the teacher of Karl Menger and Kurt Gödel, was mainly responsible for this innovative step (see Hahn, Hans). This move also represented an attempt to provide an institutional platform and an intellectual leader for the further development of the scientific philosophy of the so-called "first Vienna Circle" (with Frank, Hahn, and Neurath) (Uebel 2000). In Vienna, in 1924, at the suggestion of his students Herbert Feigl and Friedrich Waismann, Schlick began organizing a regular discussion group that first met privately, and then in the rear building of the Institute of Mathematics in Vienna.

This forum remained in existence up until Schlick's death and later went down in the history of philosophy and science as the "Vienna Circle." This institutionalization of the Circle between 1924 and 1929 was characterized by the discussion and encounter with Wittgenstein's early philosophy and finally with Rudolf Carnap's ([1928] 1969) *Scheinprobleme in der Philosophie,* which was inspired by neo-Kantianism, Gestalt theory, and set theory, based on Mach and Russell (see Carnap, Rudolf). The formation of the circle was most lucidly described by Philipp Frank, Einstein's successor in Prague (Frank 1949) (see Vienna Circle). In addition to his extensive research and academic teaching duties, Schlick was also active in adult education as a member of the Ethical Society and, most importantly, from 1928 to 1934, as chairman of the Verein Ernest Mach. In spite of his efforts, it was dissolved February 12, 1934, for political reasons.

Beginning in 1926, Schlick came in personal contact with Ludwig Wittgenstein, who influenced him significantly. At his students' request, in 1929 Schlick refused an attractive job offer at Bonn. He spent several months in California as a visiting professor in Stanford and later (1931–1932) in Berkeley. As a professor in Vienna, Schlick's publications and lectures led to relationships with scientific communities in Berlin, Prague, Göttingen, Warsaw, England, and the United States. Together with Philipp Frank, he published the series *Schriften zur wissenschaftlichen Weltauffassung* from 1929 to 1937.

At the apogee of his influential life as a scholar, on June 22, 1936, Moritz Schlick was murdered on the steps of the University of Vienna by a former student. The student received early release by the Nazis and lived as a free citizen in Austria after 1945. Schlick's murder marked the definitive demise of the Vienna Circle, whose remaining members were forced to emigrate after the German anschluss of Austria in 1938.

Between Natural and Cultural Philosophy

Schlick thought philosophy had an important and independent function in relation to the natural and social sciences. He also embodied the prototype of a liberal, cosmopolitan intellectual in the midst of a "national-socialist revolution." Intellectually, it is difficult to accurately characterize Schlick, who clearly sympathized with American pragmatism and its commonsense thinking. Archival evidence indicates an early desire to liberate himself from traditional philosophy and dedicate himself to the exact sciences. Equally important in his work was an attempt to deal with the natural and cultural world (ethics) simultaneously.

A revealing entry by Schlick in the *Philosophen-Lexikon* begins with the following programmatic claim: "Schlick attempts to justify and construct a consistent and entirely pure empiricism." However, unlike earlier forms of empiricism, Schlick thought that the doctrine could be justified only by applying the techniques of modern mathematics and logic to reality. The entry continues:

> From there, and with the help of an analysis of the process of knowledge, the 'General Theory of Knowledge' arrives first at a clear distinction between the rational and the empirical, the conceptual and the intuitive. Concepts are mere symbols that are attributed to the world in question; they appear in 'statements' ordered in a very particular way, by which these are able to 'express' certain structures of reality. Every statement is the expression of a fact and represents knowledge insofar as it describes a new fact with the help of old signs—in other words, with a new combination of terms which have already been used in other regards. The ordering of reality . . . is determined solely by experience, for which reason there exists only empirical knowledge. The so-called rational truths, then, purely abstract statements such as the logical-mathematical ones, . . . are nothing more than rules of signs which determine the syntax of the language (L. Wittgenstein) which we use to speak about the world. They are of purely analytic-tautological character and therefore contain no knowledge; they say nothing about reality, but it is for precisely this reason that they can be applied to any given fact in the world. Thus, knowledge is essentially a reproduction of the order, of the structure of the world; the material or content belonging to this structure cannot enter it; for the expression is, after all, not the thing itself which is being expressed. Therefore, it would be senseless to attempt to express the 'content' itself. Herein lays the condemnation of every variety of metaphysics; for it is precisely this that metaphysics has always wanted, in having as its goal the cognizing of the actual 'essence of being.' (Schlick 1950)

This short text has its origins in Schlick's Viennese period, as the reference to Wittgenstein indicates. It also represents the essence of his most important work on epistemology, his *General Theory of Knowledge* (Schlick [1918] 1974), which manifests a specific sort of interaction between philosophy and the sciences. As Schlick (1950, 463) concludes:

> Philosophy is not a science, even though it pervades all sciences. Because while these latter consist of systems of true assertions and contain knowledge, philosophy consists in the search for the meaning of the statements and creates understanding, which leads to wisdom.

These statements do not entail any relation of priority between nature and culture, or between theoretical and moral philosophy. According to Schlick (1950), no priority exists because ethics and aesthetics can be done in congruence with his concept of "consistent empiricism." Hence, "It makes no sense to speak of 'absolute' values; only the evaluative behaviors actually practiced by human beings can be the object of study. Based on this standpoint arises a new justification for a kind of eudaimonism, which moral principle reads more or less so: Increase your happiness!" (463).

This ethical claim clearly distinguishes Schlick's view of ethics from Wittgenstein's philosophy of the ineffable, and from those members of the Vienna Circle and the Berlin Group who regarded moral-philosophic issues as unimportant. Schlick's view of ethics is elaborated in detail in *Lebensweisheit*, *Fragen der Ethik* (Schlick 1908 and 1930), and numerous autobiographical fragments (up to 1922) scattered throughout his writings. Moreover, throughout his life, Schlick was interested in the idea of a lifelong balance between nature, culture, and art.

Schlick claimed that even early in his life, he had begun "to philosophize before I even knew that philosophy existed. All sorts of doubts eventually drove me to metaphysics, but after having read Kant, I was calmed and done with metaphysics" (Schlick 1900, 1). Later, Schlick began studying the human sciences, which he "always thought to be less of a philosophy than a *natural* science" (ibid). These early claims anticipate Schlick's later view that the Cartesian unity of nature and humanity is manifested in modern natural science. Schlick had adopted a sort of monist conception of the world. Precisely this conception of nature led him to believe that the ultimate goal and principle of ethics was "happiness." It is unsurprising, then, that he juxtaposed Kantian ethics with the obligation of a eudaimonist "ethics of charity" and proposed the mindless play of youth as a paradigm for an everyday morality.

Returning to epistemology, Schlick described his intellectual development prior to Rostock in an unpublished manuscript (Schlick CV): "The endeavor towards universality and the most general principles of knowledge and science," he wrote, "distinguishes philosophy from the special sciences. Searching for answers to the great questions of life, not just working the theoretical fields of knowledge, are the tasks of scientific philosophy." Schlick does not claim to be a philosopher by virtue of his work on the epistemological foundations of inductive science. Rather, his dual interests in knowledge of nature and of the personal life remained constant and significant throughout his career. There is here a parallel to Ernest Mach (see Mach, Ernest).

Central to Schlick's early conception of philosophy was a belief that metaphysics was trivial and a distrust of pure speculation. Schlick thought a peaceful and friendly coexistence between philosophy and natural science was possible, and a passion and ambition for this ideal formed the foundation of his academic career. He initially elaborated this ideal in a pragmatic conception of truth: "that truth is merely another name for the utility of a judgment, and nothing more; in other words, that the utility of a judgment is not a consequence of its truth—it is rather the very essence of the truth" (Schlick CV, 15). This pragmatic view is also displayed in Schlick's fondness for American philosophy (which prompted him to visit Stanford and Berkeley).

Later, however, he rejected the pragmatic theory of truth (Schlick 1910). This rejection depended upon a distinction between content and form, which Schlick developed in detail during a lecture series at the University of London in 1932. The purely formal structure of the system of facts is the same as that of a system of (logical) judgment. The contents of a system reflected the empirical world. The domains of reality and thought are thus separated and can be correlated only by the assignment of content to "sign" of the formal system. This was neither rationalism nor extreme empiricism, and therefore rejected the neo-Kantian reconciliation of the two around the turn of the century.

Schlick began exchanging letters with Einstein on a regular basis in 1915. Initially, Einstein thought Schlick was *the* philosophical interpreter of the theory of relativity. Later, Einstein himself was seen, along with Russell and Wittgenstein, as a model of the *Wissenschaftliche Weltauffassung* ([1929] 1973), particularly regarding his claim that "[i]nsofar as the statements of mathematics refer to reality, they are not certain, and insofar as they are certain, they do not refer to reality" (Einstein 1921, 119f).

Schlick produced one of the most important early philosophical interpretations of Einstein's work, *Space and Time in Contemporary Physics* ([1917] 1920), but began disagreeing with Einstein after turning to Wittgenstein's philosophy of language in the 1920s. In the early 1920s, Schlick read Wittgenstein's *Tractatus* in seminars held by the mathematicians Hans Hahn and Hans Reidemeister. He also studied Russell's philosophy of logical atomism and neutral monism (McGuinness 1985). Wittgenstein and Schlick's later differences concerned realism, conventionalism, and causality (Schlick 1932–1933). Wittgenstein's philosophy, focused primarily on linguistic analysis, viewed the problem of reality as a pseudo-problem in the sense of Carnap ([1928] 1969).

By 1924, the intellectual and institutional foundations of the Schlick Circle (that is, the later Vienna Circle) had been laid. Before Schlick's (1925) publication of the second edition of *Allgemeine Erkenntnislehre*, he was committed to a form of critical realism. After the 1930s, furthermore, he and Waismann constituted the wing of the Circle inspired by Wittgenstein. Schlick continued to conceive of philosophy as a system for expressing the most general principles inherent in the sciences. However, philosophy was seen as a clarifying activity, involving the logical analysis of statements within particular sciences that would overcome metaphysics and provide a clearer account of meaning. Schlick initially endorsed a realistic position, via verificationism, but later he adopted a more liberal position, but still based on a correspondence theory of truth. This view distanced Schlick from Neurath's "nonphilosophical" physicalism based on a coherence theory of truth, and from Neurath's project of an encyclopedia of unified science, which emerged from the controversial debate about protocol sentences.

Last Years

With the rise of fascism and the threat of National Socialism in Vienna, Schlick was prompted, like the mathematician Karl Menger, to focus on the threatening intellectual situation of his time. Within such an environment, Schlick developed what he called a "consistent individualism" as a correlate to consistent empiricism. The omnipotent state was to be dissolved in favor of a league of nations to promote the happiness of all people, but not in the guise of Social Darwinism. Accordingly, Schlick criticized nationalism as destructive. Specifically, he saw the National Socialist state as contrary to liberal democracy (Schlick [1930] 1962, 44ff). Schlick also reflected on the concept of the state. He developed a conception of a nonterritorial state as a concrete utopia that would be defined by the population that was to become citizens of different states by their own free will. Peaceful coexistence of several "invisible states" would be possible in the same territory; such an organization would make absurd any nationalist social organization (Schleichert 2003). Schlick argued that the only basis for a union of individuals into states being reliable are ethical qualities, such as character, and the peaceful voluntary nature of individual political and religious affiliations (Schlick [1930] 1962, 107).

Schlick worked on a book, *Natur und Kultur*, which was not completed due to his untimely death. It was published in a posthumous redaction, and in it Schlick (1952) said that in the last years of his life, he had been interested mainly in problems of cultural philosophy and ethics, on which he had lectured. The central issue of the book was suffering caused by culture, particularly existential need, tribulations of love, and the mind. Only the first part was completed.

Schlick's last works show that he was attempting a programmatic synthesis of nature and culture. It was a modern variant of a monistic world view that tries to situate the realm of facts and values in a humanist and cosmopolitan context, or in Schlick's words: "Art is a desire for nature. Culture is a bridge on both ends of which nature rests" (Schlick A 110).

Friedrich K. Stadler

References

Carnap, Rudolf ([1928] 1969), *The Logical Structure of the World and Pseudoproblems in Philosophy*. Berkeley and Los Angeles: University of California Press 1969. Original published as: *Scheinprobleme in der Philosophie. Das Fremdpsychische und der Realismusstrei; Der logische Aufbau der Welt*. Berlin-Schlachtensee: Weltkreis-Verlag.

Einstein, Albert (1921), *Geometrie und Erfahrung* [Geometry and Experience]. Berlin: Springer.

Frank, Philipp (1949), *Modern Science and its Philosophy*. Cambridge, MA: Harvard University Press. (Especially the Introduction and Historical Background)

Gadol, Eugene (ed.) (1982), *Rationality and Science: A Memorial Volume for Moritz Schlick in Celebration of His Birthday*. Vienna and New York: Springer.

Haller, Rudolf (ed.) (1982), *Schlick und Neurath. Ein Symposium* Amsterdam: Rodopi.

Mach, Ernest (1883), *Die Mechanik in ihrer Entwickelung historisch-kritisch dargestellt*. Leipzig: F. A. Brockhaus. English edition: *The Science of Mechanics*. Translated by Thomas J. McCormack. Chicago and London: Open Court Company, 1893.

McGuinness, Brian (ed.) (1985), *Zurück zu Schlick. Eine Neubewertung von Werk und Wirkung*. Vienna: Hölder-Pichler-Tempsky.

Schleichert, Hubert (2003). "Moritz Schlick's Idea of Non-Territorial States," in F. Stadler (ed.), *The Vienna Circle and Logical Empiricism: Re-Evaluation and Future Perspectives*. Dordrecht, Holland: Kluwer, 49–62.

Schlick, Moritz (1900), Curriculum Vitae 1900, C 1b. Moritz Schlick Archives: Rijksarchief Noord Holland, Netherlands.

——— (CV). Curriculum Vitae (undated), C 26. Moritz Schlick Archives: Rijksarchief Noord Holland, Netherlands.

——— (A 110), *Staat und Kultur*. Moritz Schlick Archives: Rijksarchief Noord Holland, Netherlands.

——— (1908). *Lebensweisheit. Versuch einer Glückseligkeitslehre* München: C. H. Beck'sche Verlagsbuchhandlung Oscar Beck.

——— (1911), "Das Wesen der Wahrheit nach der modernen Logik," *Vierteljahrsschrift für wissenschaftliche Philosophie und Soziologie* 34: 386–477.

——— ([1917] 1920), *Space and Time in Contemporary Physics*. Translated by Henry L. Brose. Oxford and New York: Clarendon. Originally published as "Raum und Zeit in der gegenwärtigen Physik. Zur Einführung in das Verständnis der allgemeinen Relativitätstheorie," in *Die Naturwissenschaften.* Berlin: Julins Springer. 5: 161–167 and 177–186.

——— ([1918] 1974). *General Theory of Knowledge*. Translated by Albert E. Blumberg. Vienna and New York: Springer.

——— (1925), *Allgemeine Erkenntnislehre* (2nd ed.). Berlin: J. Springer.

——— ([1930] 1962). *Problems of Ethics*. Translated by David Rynin. New York: Prentice-Hall.

——— (1932–33), "Positivismus und Realismus," *Erkenntnis* 3: 1–31. English translation: "Positivism and Realism," *Synthese* 7, 478–505.

——— (1950), "Schlick, Moritz," in *Philosophen-Lexikon. Handwörterbuch der Philosophie nach Personen*. Edited by Werner Ziegenfuss und Gertrud Jung. Berlin: de Gruyter, 462ff.

——— (1952), *Natur und Kultur*. From the Schlick estate, edited by Josef Rauscher. Vienna: Gerold.

——— (1979), *Philosophical Papers* (2 vols.). Edited by Henk L. Mulder and Barbara van de Velde-Schlick. Translated by Peter Heath. Dordrecht, Holland: Reidel.

——— (1987). *The Problems of Philosophy in Their Interconnection* Winter Semester Lectures, 1933–1934. Edited by Henk L. Mulder, A. J. Kox, and Rainer Hegselmann. Translated by Peter Heath. Dordrecht, Holland: Reidel.

Stadler, Friedrich (2001), *The Vienna Circle: Studies in the Origins, Development, and Influence of Logical Empiricism*. Vienna and New York: Springer.

Uebel, Thomas (2000), *Vernunftkritik und Wissenschaft. Otto Neurath und der erste Wiener Kreis*. Vienna and New York: Springer.

Wissenschaftliche Weltauffassung. Der Wiener Kreis ([1929] 1973). English translation in Otto Neurath, *Empiricism and Sociology*, edited by Marie Neurath and Robert S. Cohen. Dordrecht and Boston: Reidel.

SCIENTIFIC CHANGE

Epistemology, as the theory of knowledge, may refer either to the process of knowledge acquisition in an individual or to the growth of knowledge in general (see Epistemology). If scientific knowledge is accepted as a valid form of knowledge, then scientific change is central to the second aspect of epistemology. Scientific change is the process by which scientific knowledge is accumulated, transformed, and, possibly, lost. Understanding scientific change then becomes part of epistemology. To the extent that scientific knowledge is privileged over other forms of knowledge, no epistemology would be complete without an account of scientific change. Explicit interest in scientific change, as distinct from epistemology in general, emerged with empiricism, and especially with inductivism and interest in the scientific method (see Induction, Problem of). By the time of the Enlightenment, scientific change was widely taken to be progressive, resulting in the cumulative and monotonic growth of knowledge. It was also viewed as a collective enterprise with its methods and results available for public scrutiny. Implicit theories of scientific change were widespread by the late nineteenth century—for instance, in Comte's positivism, as well as in British natural philosophy, associated with figures such as Mill and Pearson. For Peirce, abduction provided the mechanism of scientific change (see Abduction). The twentieth century saw both philosophical theories of scientific change and many attempts to model the process. Many of the articles in this encyclopedia deal with particular models and theories of scientific change—this essay will provide a brief overview and a guide to the others.

Traditional Objectivism

With respect to scientific change, the logical empiricists distinguished between contexts of discovery and contexts of justification (Reichenbach 1938, 6–7). For justification, the logical empiricists hoped to find a systematic theory that, in Carnap's case, became inductive logic or the logic of confirmation (see Carnap, Rudolf; Confirmation Theory; Inductive Logic). The project of inductive logic remains incomplete. During roughly the same period, Popper and Reichenbach produced their own variants of frameworks for justifying theories (see Corroboration; Popper, Karl Raimund; Reichenbach, Hans). One feature that unites all these approaches is that the process of justification is *supposed* to be guided by a priori principles; thus, they reject a naturalized epistemology, at least with respect to justification. This option allows the straightforward introduction of normative principles into theory acceptance. Meanwhile, during the last few decades, going beyond the work of Popper and the logical empiricists, there have been more promising attempts to develop justificatory frameworks more consonant with the standard use of statistical reasoning within science—for instance, classical and Bayesian inference (see Bayesianism).

Beyond justification, is there a logic of scientific discovery? The use of "logic" suggests that scientific discovery is a rational process. While explicit claims of the existence of such a logic were rare in the first half of the twentieth century, several mechanisms of theory change (e.g., theory reduction and unification) can be seen as rational procedures that form part of such a logic (see Reductionism; Unity and Disunity of Science). However, for most logical empiricists, as also for Popper, discovery was a psychological category for which there could be no philosophical theory. The process of discovery was thus at best orthogonal to the aims of epistemology. Whether or not it is rational is simply not interesting from this perspective.

Irrationality and Relativism

Even philosophical theories of scientific change are about actual historical episodes and must partly appeal to empirical features for their justification. Thus, detailed case studies of episodes in the history of science become central to the study of scientific change. However, until the 1950s, most philosophically oriented histories of science paid little attention to the sociological and institutional contexts of scientific work. That situation changed in the 1960s, especially with the publication of Kuhn's (1962) *Structure of Scientific Revolutions*. Over the years, different commentators have taken different lessons from Kuhn's work (see Kuhn, Thomas). What is uncontroversial about it is that according to Kuhn's model, science does not proceed by cumulative progressive changes. There may be drastic shifts of perspective, with participants from divergent positions sometimes being unable to communicate with each other (see Feyerabend, Paul; Incommensurability; Hanson, Norwood Russell).

Whether or not Kuhn intended it, this view of science challenges many traditional assumptions about the rationality of the scientific process. What is critical here is that such rationality is being challenged in the context not only of discovery but also of justification. Meanwhile, Feyerabend and others denied the validity of the distinction between the contexts of discovery and justification, strengthening the claim of irrationalism.

Kuhn's work has been seen as suggesting an evolutionary model for scientific growth. Partly independently of Kuhn, many others, including Popper (1972) and Toulmin (1972), also argued for such a model; Hull (1988) offers what is probably its most detailed application to an actual example of scientific change (see Evolutionary Epistemology). Kuhn's work is also associated with the view that scientific "truths" are relative, not only to the empirical evidence, but to social ideologies (see Feminist Philosophy of Science; Social Constructionism). Extreme versions of this view even deny that scientific knowledge is privileged *qua* knowledge over other cultural products. And he is seen as denying that the history of science can be viewed as one of cumulative progress. This claim challenges the most central tenet of beliefs about scientific change since the Enlightenment (see Scientific Progress).

Each of these positions continues to be debated and developed to the present day. (For Kuhn's own doubts about such interpretations of his work, see Kuhn 2000; Hacking 1981.) Lakatos and Musgrave (1970) and Laudan (1977) made important early responses, each developing novel accounts of scientific change that at least partly preserved traditional assumptions of rationality and progress (see Lakatos, Imre; Research Programs).

Modeling Discovery

The best antidote to claims of irrationalism, whether it be with respect to the discovery or the justification of scientific claims, is to model these processes explicitly from rational principles. One strategy is to use detailed historical case studies, as Kuhn and his followers had done, but with explicit attention to the presumed mechanisms of change. Donovan, Laudan, and Laudan (1988) collect many such cases from chemistry, geology, and physics from the seventeenth to the twentieth centuries. They find wide agreement on a variety of theses, connected with issues of theory acceptance, the presence of anomalies, and innovation and discovery. This work seems to justify many traditional ideas, such as the role of predictive success and the resolution of anomalies in generating acceptance of theories. Lakatos' work on research programs can also be viewed as being of similar intent and more general in scope, including considerations of both acceptance and discovery of theories (see Research Programs). However, what remains unclear from such case studies is whether particular principles gleaned from them ever adequately address the normative aims of epistemology: When *should* scientific claims be accepted or rejected?

An even more powerful method to study scientific change, especially discovery, is to model it computationally. This approach to scientific discovery was pioneered by Simon (1977; see Buchanan 1982). Science was viewed as a problem-solving activity using heuristics that could be equally applied to historical and contemporary situations. Techniques from artificial intelligence were brought to bear on the problem of generating explanatory hypotheses given a body of evidence (see Artificial Intelligence). The first expert system to be designed for this purpose was DENDRAL (Lindsay et al. 1980); it hypothesized the presence of specific chemical compounds from mass spectrographic data. Of only slightly later vintage, a system devised by the group around Simon, BACON, sought patterns in numerical empirical data. Going beyond traditional curve-fitting, BACON was based on the (metaphysical) assumption that an entity's relational properties were caused by its intrinsic properties (Langley et al. 1987). STAHL, also designed by the same group, used theory-driven discovery. Another product, KEKADA, successfully modeled the patterns of reasoning in the work of the biochemist Krebs (Kulkarni and Simon 1988). Taken as a whole, this work helped dispel that view that scientific discovery was a mysterious process, relying on serendipity or some special inspiration. It also made plausible the case that with improving heuristics, there would be a discernible logic of scientific discovery. In a recent review of this collective work, Langley (1998) argues that the best results are obtained when there is interaction between an expert system and human agents (that is, automation is not total). Throughout, the problem that has generated most controversy within the field is the choice of representation of the domain, that is, how the basic conceptual entities should be modeled. This is where human agents may play a crucial role. Critics of this body of work include Gillies (1996), who has argued that researchers have focused on laws and regularities already known, which may have artificially inflated the success of the approaches.

The last twenty years have seen many other computational frameworks entrenched in scientific contexts. Some were spawned from these pioneering frameworks. Glymour (e.g., 2001) and others have advocated the use of Bayesian causal networks to model cognitive processes, including those of scientific innovations. Many of the new frameworks draw on general learning and search algorithms (including metaheuristic algorithms such as genetic algorithms and neural networks) as well as data mining methods (see Fayyad et al. 1996). These have been especially valuable in the biological and environmental sciences. In these fields, the large quantities of data that are generated (e.g., from DNA sequencing projects, from remote sensing via satellites) require automated analysis (see Molecular Biology). What remains to be seen is whether such automated analysis can lead to genuine scientific innovation.

What has also become unclear, though, is whether these techniques of "discovery" are, in any sense, plausibly similar to the methods used by human agents for scientific discovery. (The contrast here is with the earlier methods from the 1970s and 1980s, which typically used heuristics gleaned from human reasoning.) However, from a normative perspective, the question of similarity is irrelevant: if the automated methods generate correct results, they are the ones that should be used irrespective of whether they capture human reasoning. Thus, while it may be the case that no new understanding of the historical processes of scientific discovery are being achieved by these new methods, they may well become staples of the process of such discovery in the future. The situation calls for more normative analysis of automated reasoning.

Sahotra Sarkar

References

Buchanan, B. G. (1982), "Mechanizing the Search for Explanatory Hypotheses," in P. Asquith and T. Nickles (eds.), *PSA 1982: Proceedings of the Biennial Meetings of the Philosophy of Science Association*, vol. 2. East Lansing, MI: Philosophy of Science Association, 129–146.

Donovan, A., L. Laudan, and R. Laudan (eds.) (1988), *Scrutinizing Science: Empirical Studies of Scientific Change*. Dordrecht, Netherlands: Kluwer.

Fayyad, U. M., G. Piatesky-Shapiro, P. Smyth, and R. Uthurusamy (eds.) (1996), *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA: AAAI Press.

Gillies, D. (1996), *Artificial Intelligence and Scientific Method*. Oxford: Oxford University Press.

Glymour, C. (2001), *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. Cambridge, MA: MIT Press.

Hacking, I. (ed.) (1981), *Scientific Revolutions*. Oxford: Oxford University Press.

Hull, D. L. (1988), *Science as a Process: Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.

Kuhn, T. S. (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

——— (2000), *The Road since Structure: Philosophical Essays, 1970–1993*. Chicago: University of Chicago Press.

Kulkarni, D., and H. Simon (1988), "The Process of Scientific Discovery: The Strategy of Experimentation," *Cognitive Science* 12: 139–175.

Lakatos, I., and Musgrave, A. (eds.) (1970), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press.

Langley, P. (1998), "The Computer-Aided Discovery of Scientific Knowledge," *Lecture Notes in Computer Science* 1532: 25–39.

Langley, P., H. A. Simon, G. L. Bradshaw, and J. M. Zytkow (1987), *Scientific Discovery: Computational Explorations of the Creative Process*. Cambridge, MA: MIT Press.

Laudan, L. (1977), *Progress and Its Problems: Toward a Theory of Scientific Growth*. London: Routledge and Kegan Paul.

Lindsay, R. K., B. G. Buchanan, E. A. Feigenbaum, and J. Lederberg (1980), *Applications of Artificial Intelligence to Organic Chemistry: The DENDRAL Project*. New York: McGraw-Hill.

Popper, K. R. (1972), *Objective Knowledge: An Evolutionary Approach*. Oxford: Oxford University Press.

Reichenbach, H. (1938), *Experience and Prediction: An Analysis of the Foundations and Structure of Knowledge*. Chicago: University of Chicago Press.

Simon, H. A. (1977), *Models of Discovery*. Dordrecht, Netherlands: Reidel.

Toulmin, S. (1972), *Human Understanding*. Oxford: Clarendon Press.

See also **Abduction; Artificial Intelligence; Carnap, Rudolf; Confirmation Theory; Corroboration; Epistemology; Evolutionary Epistemology; Feyerabend, Paul; Incommensurability; Inductive Logic; Hanson, Norwood Russell; Kuhn, Thomas; Lakatos, Imre; Logical Empiricism; Popper, Karl Raimund; Reductionism; Reichenbach, Hans; Research Programs; Scientific Progress; Scientific Revolutions; Social Constructionism; Unity and Disunity of Science**

SCIENTIFIC DISCOVERY

See **Scientific Change**

SCIENTIFIC DOMAINS

According to empiricism, all our concepts and beliefs are based on observation, which is equated with sense perception. This doctrine has been widely criticized on two grounds: first, that it fails to specify what counts as an observation, free of any presupposition that should, according to it, be based on observation rather than presupposed by observation; and second, that it fails to delineate the sense in which all our nonobservational concepts are "based on" observation. However, deeper defects exist, showing that even if these two objections were overcome, empiricism, so conceived, would still have distorted or ignored aspects of science that are arguably far more central thereto than those on which empiricists focused. Some of these defects stem not from the difficulty of delineating the class of observation terms but rather from the fact that most empiricists classified all the leftovers—all nonobservational concepts—indiscriminately as "theoretical." When coupled with the later (logical) empiricist emphasis on meaning and logic, this meant that nearly every issue examined concerned definitional and logical relations between the two classes, and the functions, if any, of theoretical concepts.

Consider accounts of observational and theoretical terms in twentieth-century logical empiricism. Theoretical terms were first conceived as definable and therefore eliminable via observational terms. Where meaningful at all, theoretical terms served only some practical function, perhaps as tools for economically organizing a mass of empirical data, or as suggestive scaffolding, useful in seeking empirical correlations but to be discarded once those relations were found. In conformity with the antimetaphysical roots of empiricism, they had nothing to do with the world; science was said to be concerned, fundamentally, only with correlating observations. Later, when this thesis of definability failed, theoretical terms were seen relative to their place in the logical structure of a theory, but still only as links between observation terms, which were correlated with the world as theoretical terms were not.

It was precisely here that the empiricist program, as applied to science, revealed its drastically restricted scope. For increasingly since science became a critically developing tradition, deep changes have occurred in what scientists spoke and wrote of as the subject matters of their investigations. In the seventeenth and eighteenth centuries, scientists widely engaged in studying and attempting to explain such subject matters, or portions thereof, as heat, light, electricity, magnetism, motion, and the properties of physical matter. By the end of the twentieth century, the ensuing sequence of developments resulted in scientists' specializing in studies of various levels of specificity and delineation of such areas, or parts thereof, as nuclear physics, weak interactions, dark matter, neutrino astrophysics, quantum gravity, genomic mapping, archaea, and mantle convection.

An example of one type of such change is the development of the concept of atoms in the nineteenth and twentieth centuries. Dalton proposed that atoms could account for the laws governing the proportions in which elements enter into compounds. After much resistance to such unobservable entities, Einstein's predictions about Brownian motion, and their experimental verification by Perrin, led to scientific acceptance of atoms, and they themselves became objects of study. The atom was found to consist of a nucleus with "orbiting" electrons, and quantum mechanics made it necessary to describe these entities (if such they could be called) in terms far removed from those of sense experience. The nucleus in turn became a subject for study, as did particles outside the nucleus; these particles in turn came to be seen as excitations of a quantum field rather than fundamental particles, and as having all the quantum idiosyncrasies.

In this sequence of reconceptions, what was originally the focal "theoretical" idea in an explanation of a subject matter (elemental proportions in compounds) took on a new status, becoming itself a subject matter for investigation and explanation. Insofar as terms like 'fact' and 'observation' refer to what is accepted and sometimes requires explanation, then at a certain point (roughly, Perrin's experiments) the theory of atoms passed from being "theoretical" to assuming this characteristic of the factual or observational, as having to do with a subject matter requiring study and explanation.

Such transitions occurred repeatedly in this specific case and in general throughout science. Thus, within the class of theoretical terms or concepts there exists a considerable degree of structure and function; and as a result of a change of status, these terms sometimes take on features usually associated with fact or observation. The structure is dynamic, subject to alteration. Indeed, scientists say that, at least as a general rule, the changes in status, description, and organization of subject matters are brought about for scientific reasons; and they are able, as a general rule, to detail those reasons in particular transitions.

Such facets of science were far beyond the central concerns of logical empiricism and its antecedents. The source of this neglect lay in the very conception of the empiricist program, in terms of the distinction between the observational and the theoretical, and how to analyze each. Much of the reasoning that occurs in science was masked by the empiricists' focus on this distinction, their indiscriminate classification of all nonobservational concepts as theoretical, and their attempt to treat all theoretical concepts solely in terms of definitional or logical relations to observational concepts. In effect, this combination suggested that scientists are not really, and cannot be, studying what they think and say they are studying. The real work of science is with observations (sense perceptions) and their interrelations, delusions stemming from the use of theoretical terms notwithstanding. Thus empiricist doctrine discouraged any analysis of the reasoning involved in the changes of status, description, and organization of theoretical terms. The idea that, in these respects, concepts in the theoretical category were in flux, and that changes might be made for reasons, went unconsidered.

A few philosophers did propose that the "observational-theoretical" distinction was oversimple, that the distinctions of theoretical from nontheoretical and observational from nonobservational were not coextensive, or that some terms are both theoretical and observational, depending on context (Achinstein 1968; Maxwell 1962; Putnam 1962). However, these proposals remained primarily critical of how things were being done and did not generate systematic attempts to carry out the proposals constructively. Most important, they did not encourage examination of the reasons why the two alleged classes overlap or the more fundamental ways in which the status of theoretical terms depends on context.

Critics of empiricism such as Hanson (1958), Kuhn (1970), and Feyerabend (1978) did little better. Despite offering some important insights, their emphasis on the "theory-ladenness" of observation perpetuated the focus on the observational/theoretical distinction. They departed from empiricism chiefly in regard to theory as being prior to observation rather than based thereon. This, however, appeared to deny the basic empiricist insight that if prior beliefs influence observation, objectivity is violated. Attention therefore tended to concentrate on issues stemming from this denial, such as whether the view was committed to relativism and incommensurability of alternative sets of presuppositions.

Thus, for both empiricists and their critics, the question of changes in subject matter in science was obscured by the distinction between observational and theoretical concepts. In light of this neglect, it is appropriate to separate questions about subject matters and their changes from the observational/theoretical distinction by using a new term, 'domain,' to refer to what is being studied by scientists at a given stage. This term focuses attention on how subject matters of inquiry are described and conceived, the reasons for studying them, the ways and reasons domains become reconceived and reorganized, and the implications of those changes for understanding science.

As presented here, the motivation behind the domain concept resides solely in the fact that important issues about variety and change in theoretical terms have been ignored, and this neglect is traceable largely to the standard formulation of the empiricist program. It is independent of the fact that transitions in domain description and organization often proceed from the sensory to the nonsensory, from the familiar to the unfamiliar, or from the descriptive to the explanatory. Nevertheless, analysis of domains and domain change does have important implications for a number of philosophical issues. For example, analysis of domain change is concerned with what are taken to be reasons for such changes and can generate an account of why those considerations are justified as reasons. (This is what distinguishes the inquiry as philosophical rather than only exclusively historical.) Since the topic is concerned with what is seen as requiring explanation, understanding of the reasons given for such a requirement can illuminate problems about scientific explanation. And finally, since the process of altering descriptions and functions of domains builds on what has been learned through inquiry regarding domains, light can be thrown on issues of realism and reference.

An Example of a Domain, Seventeenth Through Early Nineteenth Centuries

Among the major domains investigated widely from the seventeenth through the early nineteenth centuries were the motion of bodies, light, heat, magnetism, electricity, "airs" (gases), and various types of material substances. Light provides a good example, partly because the items included in it are readily presentable but also because it was the subject of Newton's (1952) *Opticks*, which became an early model for investigation of domains. The items of this domain included rectilinear propagation of light from source to recipient, the laws of reflection and refraction, the colors produced when light passes through a prism, diffraction bands outside the geometrical shadow of thin objects, the finite velocity of light, double refraction in Iceland spar crystals, and the colors of thin films. Some of these properties had been known since ancient times; others were discovered later. They were considered to be related, as having to do with a particular experientially identifiable aspect of visual experience, light, along with color.

Light as a domain played two important roles: first, as a domain of *investigation* to be studied, and second, as a domain of *responsibility* for any explanatory theory of the domain. Regarding the latter role, although in the seventeenth century most workers were not familiar with every one of these items of the domain, the expectation grew that a satisfactory theory of light must account for all of them.

Chief Characteristics of Domains

Four properties of domains are discussed here:

1. *Domains could be studied in isolation from other subject matters, or at least it was assumed that they could.* This fundamental tenet of the approach could in principle have been false, as is brought out by an analogy with gravity. If gravity remained constant at all distances, the paths of planets and projectiles could not be calculated, as the effects of all other bodies in the universe would enter equally into the calculations. If the same had proved true of the factors ignored in studying individual domains, treatment of them as isolated subject matters would have been impossible.
2. *Domains were to be studied for the sake of "understanding"("explaining") them.* That is, they were not to be studied for their everyday uses and purposes. Much debate in the seventeenth century was shaped by earlier doctrines regarding fundamental explanation. For ancient philosophers influenced by Parmenides, fundamental explanations must be in terms of what is unchanging. Newton's atoms were of this nature, changing not in their intrinsic properties but only in their positions and velocities. In contrast, for Leibniz (as for Anaximenes, Heraclitus, and the Stoics) a fundamental explanation must be in terms of the intrinsically active, and therefore space, time, and matter cannot explain. Initially, then, expectations about explanation were quite general and were relevant chiefly to fundamental explanation. As science developed, requirements for explanation became increasingly more specific in the light of inquiry.
3. *Domains were derived primarily from classifications made in everyday life and language, mainly for practical ("applied") purposes, and descriptions of the domain and its items were given in terms of everyday language.* In the domain of light, for example, descriptions of properties were generally given in everyday sensory terms. (These descriptions were thus independent of the received views of explanation.) This claim depends not on there being a universal and unchangeable "observation language" but only on there being descriptive language common enough for mutual understanding within a particular community existing at a particular time and place, with regard to the most common objects, events, and properties. Thus the properties of light were described in terms of paths of travel, speeds, order of colors, alternating bands of light and darkness, appearances of sticks partially submerged in water, etc. Some properties, like the periodicities manifested in Newton's rings, could be experienced but, not being immediately apparent to the senses, had to be drawn out by close investigation. But even those were describable, naturally and unproblematically, in terms of the common language for talking about familiar objects, events, and properties in the everyday experience and exchanges of a given community. This is not surprising: At that early stage of concerted study of nature, what else could description be? This feature probably served as a model for empiricist philosophers, although they misconstrued such description as some fundamental and universal "given" in experience and, further, failed to take into account the departures from familiar description that characterized subsequent scientific change.

   From the beginning, there were presumptive exceptions. Even with light, geometry appeared in descriptions of reflection and refraction, and perhaps such terms do not qualify as having to do with "everyday sensory" properties. Other ideas, such as Newton's force of gravity, seemed "occult," and Newton himself often portrayed this force in purely sensory terms, as the manner in which two bodies move in each other's presence. But over the following centuries, scientific departures from common everyday descriptions became increasingly radical. This brings us to the fourth characteristic of domains.
4. *In the course of inquiry, the classification of items under domains and the description of the domains and their items are subject to change.* Such change can be of several types, as discussed below.

Types of Domain Change

That domains are dynamic entities, undergoing various types of change, is their most important property for illuminating the knowledge-seeking enterprise. It is what primarily distinguishes domains from ordinary classifications.

Certain types of change tend to occur early in the development of a science and are often generated by close study of the domain independently of whether an explanation of it has been proposed.

These include changes in ways of describing a domain to avoid the vagueness and ambiguity of everyday language. Sometimes preexplanatory investigation of a subject matter brings about revision of the descriptive language, either by adding to it (diffraction and double refraction), by abandoning other such language (as with the gradual replacement of alchemical vocabulary), or by modifying the concepts without altering the language of description (as with distinguishing the motion of falling bodies from that of projectiles). Other changes result from discoveries of new items of the domain, as with the double refraction of Iceland spar. Close study of a domain occasionally leads to dividing it into separate domains; for instance, hazy patches in the sky came to be viewed not as one sort of thing but as being either clouds of gas or galaxies, henceforth to be studied as separate specialties. Conversely, two domains can be joined as a result of closer study, as when preexplanatory experiments by several workers, culminating with Faraday, showed that what had seemed to be different sorts of electricity were all of the same type.

Other types of change are brought about by proposed or accepted explanations of a domain. The language for dealing with a particular domain can be substantially revised by an explanation, the most remarkable instance being the revision of the language of chemistry by Lavoisier and his associates in the light of his proposed explanatory theory. Occasionally, older descriptions survive, divorced from their original contexts. Thus early explanations of heat and electricity as fluids introduced a descriptive vocabulary ("flow," "current") into descriptions in those domains. When the fluid theory of heat, for example, was replaced by the kinetic theory, the fluid-based language was largely retained as descriptive of the domain but not as explanatory—the explanation lay in a different theory. Explanations can also lead to splitting of domains (bacteria versus archaea).

Unifications and Transformations of Science

All of the above types of changes in description and organization of domains, and their explanations, were concerned with single domains, studied and explained in isolation from one another. In that process, once a domain is given an acceptable explanation, the conceptual or descriptive reform can lead to that explanation's being given a deeper explanation. What was explanatory thus itself becomes a domain, to be studied and explained. Sometimes such deeper explanations result in unification of previously separated domains.

Although domain unifications had been achieved earlier (e.g., Newton's fusion of terrestrial and celestial physics), the process of unification became more and more prominent in the nineteenth and twentieth centuries (see e.g., Shapere 1991). In one degree and manner or another, unifications and other interrelationships of previously distinct areas of inquiry profoundly transformed the ways in which scientific domains are studied and explained. The transformations can be understood as follows.

There is an appearance of relativity about what counts as a domain. Even in the seventeenth century, some investigators concentrated on the study of different items of a particular domain. At an opposite extreme were those (e.g., advocates of the "mechanical philosophy of nature") who sought grand theories unifying all domains. The scope and focus of investigation, even of single domains, could vary. Despite this, and despite the fact that whether some item belonged in a particular domain could be debated, domains maintained a recognized objectivity as subject matters of investigation and explanatory responsibility: Researchers could be identified as concerned with agreed-on domains like electricity or light. With the establishment of interrelationships between domains and their explanations, new approaches to inquiry arose. A particular subject matter could be examined from any one or more of several different perspectives; conversely, one perspective could conceive and organize its subject matter in ways different from other approaches to roughly the same subject matter. Thinking in terms of domains of investigation and domains of responsibility remains instructive. But the deepening interrelationships among explanatory theories and among the subject matters investigated make the concept of a domain vastly more flexible and raise philosophical issues going far beyond the traditional empiricist focus on relations between observation and theory.

Dudley Shapere

References

Achinstein, Peter (1968), *Concepts of Science*. Baltimore, MD: Johns Hopkins Press.

Carnap, Rudolf (1954), *Testability and Meaning*. New Haven, CT: Whitlock's.

——— (1956), "The Methodological Character of Theoretical Terms," in Herbert Feigl and Michael Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 1: *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*. Minneapolis: University of Minnesota Press, 38–76.

Carswell, Donald S. L. (1971), *From Watt to Clausius: The Rise of Thermodynamics in the Early Industrial Age*. London: Heinemann.

Crosland, Maurice P. (1962), *The Language of Chemistry*. Cambridge, MA: Harvard University Press.

Cohen, I. Bernard (1956), *Franklin and Newton: An Inquiry into Speculative Newtonian Experimental Science and Franklin's Work in Electricity as an Example Thereof*. Philadelphia: American Philosophical Society.

Feyerabend, Paul (1978), *Against Method*. London: Verso.

Fox, Robert (1971), *The Caloric Theory of Gases from Lavoisier to Regnault*. Oxford: Clarendon.

Hanson, Norwood Russell (1958), *Patterns of Discovery*. Cambridge: Cambridge University Press.

Heilbron, John (1979), *Electricity in the Seventeenth and Eighteenth Centuries*. Berkeley and Los Angeles: University of California Press.

Hempel, Carl (1958), "The Theoretician's Dilemma: A Study in the Logic of Theory Construction," in Herbert Feigl, Michael Scriven, and Grover Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 2: *Concepts, Theories, and the Mind Body Problem*. Minneapolis: University of Minnesota Press, 37–98.

Kuhn, Thomas S. (1970), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Maxwell, Grover (1962), "The Ontological Status of Theoretical Entities," in Herbert Feigl and Grover Maxwell (ed.), *Minnesota Studies in the Philosophy of Science*, vol. 3, *Scientific Explanation, Space, and Time*. Minneapolis: University of Minnesota Press, 3–27.

Newton, Isaac (1952), *Opticks*. New York: Dover.

Nye, Mary Jo (1972), *Molecular Reality*. New York: Elsevier.

Putnam, Hilary (1962), "What Theories Are Not," in Ernest Nagel, Patrick Suppes, and Alfred Tarski (eds.), *Logic, Methodology, and Philosophy of Science*. Palo Alto, CA: Stanford University Press, 240–251.

Sabra, A. I. (1967), *Theories of Light from Descartes to Newton*. London: Oldbourne.

Shapere, D. (1977), "Scientific Theories and Their Domains," in Frederick Suppe (ed.), *The Structure of Scientific Theories*. Urbana: University of Illinois Press, 518–565.

——— (1991), "The Universe of Modern Science and Its Philosophical Exploration," in Alberto Cordero and Evandro Agazzi (eds.), *Philosophy and the Origin and Evolution of the Universe*, Dordrecht, Netherlands: Kluwer, 1991, pp. 87–202.

See also **Einstein, Albert; Empiricism; Feyerabend, Paul; Hanson, Norwood Russell; Kuhn, Thomas; Logical Empiricism; Observation; Particle Physics; Quantum Mechanics**

SCIENTIFIC METAPHORS

For centuries, in both the philosophical and the scientific literatures the value of the use of metaphors has been repeatedly denied, discouraged, or dismissed. Bacon, Hobbes, Locke, and Berkeley established the so-called literal-truth paradigm: Philosophers and scientists should abstain from metaphors altogether. In the twentieth century it did not help matters that the philosophical orthodoxy turned to language both ordinary and scientific. The so-called linguistic turn has proved in this respect too narrow an approach to understanding the world and how it is known, let alone a broader range of scientific practices. In the philosophy of science, the formalist orthodoxy that gave logical empiricism its backbone placed special emphasis on language as a symbolic calculus of logical relations and on observational language as the main source of cognitive significance. Within that framework, metaphors were considered as, at best, paraphrases of structural analogies within or between theories.

It was with the emergence of post–logical empiricist philosophies of science that the relations between language, cognition, and the world were understood differently and metaphors were given more attention and found to play a more important role. Renewed attention to history acknowledged the value metaphors were given by scientists such as Darwin, Maxwell, Einstein, and Bohr. Views of the role of metaphors changed accordingly.

In particular, Black (1962) rejected the comparison view that reduced metaphors to similarities and replaced it with an interaction view: Instead of formulating a preexisting similarity, metaphors allow one system to be "seen" through the frame of the second, and thereby similarities are instead created.

In her classic *Models and Analogies in Science* and subsequent works, Hesse (1966, 1993, and 1995) is concerned with the relation between models, metaphors, and truth and concludes that metaphors are explanatory redescriptions of phenomena. Hesse argues that models and analogies contribute to the interpretation and testability of theoretical hypotheses. Just as she sees models as interpretations of explanatory theories of phenomena in her account of models and analogies, in her account of metaphors, Hesse argues that a term in an explanatory model shifts its meaning to describe a new system that is to be explained. This borrows from Black's interaction account of metaphors and replaces Hempel's deductive account of explanation. Meaning shift involves a change in referent as well as in use and associated ideas and is inextricable from the explanation. An advantage of this view is that it bypasses the difficulties associated with meaning invariance and derivability of corresponding rules that plague deductive accounts of explanation (see Explanation). Another advantage, according to Hesse, is that it preserves a certain form of realism, or truth value, within the context dependence of language use (see Realism).

Finally, Hesse defends the rationality of the role of metaphors in science, characterizing rationality as the continuous adaptation of language and beliefs to a continually expanding world. Rationality rests on the prior creativity, or generativity, of metaphors; this is one of the most important features associated with the role of metaphors in science. This role can be viewed as a matter of method (a heuristic) or a matter of understanding (a cognitive standard such as concrete images or mechanical models). For Hesse, the analogy created by metaphors is linked to the heuristic fertility— the testability—of theories based on models.

Realism and explanation are also central to Boyd's (1993) approach to scientific metaphors. Boyd argues that there is an important distinction between pedagogical and theoretically relevant metaphors and that important metaphors are "theory-constitutive." In particular, according to Boyd, they are devices not just for introducing a new vocabulary but also for accommodating language to the causal structure of the world. They provide "epistemic access," introduce a causal mechanism for fixing the reference of scientific terms, and advance guiding explanatory hypotheses leading to causal theories. Hence, Boyd couples the pursuits of a causal theory of reference and a causal theory of the world. In his view, the progress of science leads inexorably to the true theory of nature in terms of a final set of natural kind terms.

Kuhn's (1993) view of metaphors constitutes a rejoinder to Boyd in the spirit of his own view of scientific language and evolution (see Kuhn, Thomas; Scientific Revolutions). Kuhn rejects the distinction between pedagogical and theory-constitutive metaphors and claims that both share the same metaphorical mechanism. Metaphors, according to Kuhn, are central to the understanding, development, and use of scientific theories. Scientific metaphors are like natural kind terms in that they bring out features of kinds of systems or phenomena but, because of their open-endedness, they also challenge and relativize the alleged inexorability of a rigid taxonomy of natural kinds. Like Hesse and Boyd, Kuhn accepts that metaphors establish links between language and the world, but this link is constitutive, in an almost Kantian sense, since there is no determined way in which the world is outside a language. This view falls in line with the post–logical empiricist theory-ladenness of observation and description. In the light of Feyerabend's and Kuhn's views of theory dependence of meanings and incommensurability, McCormach (1971) suggested that metaphors express the unknown new use (new semantically relevant properties) of an old term with known semantically relevant properties (such as 'length' in relativity theory). With a dual role of capturing familiar analogies and connotations and suggesting new theoretical possibilities, metaphors would bridge incommensurable gaps between old and new theories (see Feyerabend, Paul).

Along the lines of the idea of an underlying shared metaphorical mechanism, some authors have attempted to make sense of the role of metaphors by introducing considerations from the philosophy of rhetoric. It has been argued, for instance, that Black (and hence also Hesse) was initially wrong in assuming a reference, or extensional, theory of meaning. Metaphors organize the scientists' thoughts, but they do so by means of the sense, or intensional meaning, of the metaphorical term (closer to Black's [1993] later thoughts on the subject) (Soskice and Harre 1995). Ultimately, metaphors, suggested by models, act as probative tools that enable scientists to introduce meanings, and possibly reference to new entities, outside the conditions of experience in a given domain of phenomena. The educational, creative, and organizing view of metaphors has been espoused by Holton (1986). He sees metaphors also at the center of, or cutting across, "themata," or quasi-universal, *Gestalt*-like modes of scientific thought such as synthesis versus analysis, wave versus particles, and determinism versus indeterminism.

A large number of philosophers and especially historians of science have since the 1980s documented the use and role of metaphors in specific sciences, thereby illustrating in more detail, testing, and enriching general accounts. The outcomes range from the emphasis on concept formation and theoretical continuity to the extension of theory-ladenness to value-ladenness and the denunciation of ideological agendas. In the relation between the natural and social sciences the interaction has been shown to work both ways: Social theories developed out of mechanistic and physiological discourses in the natural sciences ("social hygine," "flow of capital," etc.), and biologists and physicists borrowed from political economy ("competition," "natural selection," "survival of the fittest," "minimum work," "least action," "value," "dissipation," "waste," etc.) (Mirowski 1989).

In biology, Keller (1985), for instance, has drawn attention to the fact that through metaphors the language of gender has carried into science certain norms and values that contribute to its shape and growth. From a feminist point of view, such norms and values are epistemologically limiting as well as ideologically unacceptable. Keller has noted also that different metaphors give rise to different cognitive perspectives, different aims, different questions, and even different methodological and explanatory preferences. She has illustrated this important point with examples from genetics, development, and evolutionary biology (e.g., "selfish gene," "gene action," "competition," "self-regulation"). Most recently, she has followed Hesse in emphasizing the role of models and metaphors in explanation (Keller 1995). Johnson and Tuana (1986) have discussed the case of Hans Selye's research in the 1920s on metabolic stress. Selye's shift from mechanical to organic metaphors for categorizing the body changed in a very definite manner the way in which his medical experience,expectations, theorizing, and treatment were structured as coherent units.

In cognitive psychology many theoretical terms are metaphorically borrowed from computer science: "information processing," "memory storage capacity," etc. Chemical language has been rife with metaphors from the time of alchemy to the modern naming of elements and their kinds, using terms such as "reaction," "chemical equation," and "chemical balance" (Radman 1995).

In physics the metaphorical character of concepts relating to time, space, matter, and causation has been discussed in general terms. References to "bodies," "work," "electric flow," "electron clouds," "electric tension," "black holes," "worm holes," and "strings" have become pervasive. The use of metaphors as central to physicists' researches has been documented in the works of, for instance, Kepler, Newton, and Maxwell (Cantor and Christie 1987). Maxwell's case is unique in having provided explicit discussion of, as well as having championed, the application of metaphors and analogies in physics connecting scientific method with experimental culture, logic, psychology, and philosophy of language (Cat 2001). Thus, projectile metaphors for light in the eighteenth century drove attempts to determine its mechanical momentum. In high-energy physics, anthropomorphic metaphors are superimposed onto theoretical language, describing detectors to enhance the expression and demarcation of the objective ontology of entities under study (by enhancing a subject/object distinction); detecting machines are described with behavioral, physiological, and moral metaphors such as "response," "reaction," "talk," "noise," "seeing," "blindness," "dead," "poisoning," "life expectancy," "misbehaving," "trustable" (Knorr-Cetina 1995). The use of nonliteral, nontechnical, or nonformal languages in science is also relevant in explaining its intersubjective and social nature. By the use of metaphors, trust and communication are enhanced within a particular community as well as across communities, such as between experimenters and theoreticians.

A general spirit behind the interest in scientific metaphors has been the perception that science is not an isolated intellectual part of human culture (Leatherdale 1974). An important consequence of the discovered value of metaphors in science is a more sophisticated understanding of language as well as the awareness of its theoretical and practical consequences. Actions, after all, are conceptualized and valued under specific descriptions. From a cognitive point of view, another consequence is that it leads to the recognition of the value of the role of the imagination in science. In any event, since the role of metaphors in scientific activity appears both important and practically inevitable, the recent historical and philosophical analyses should impress upon scientists a heightened recognition of metaphors and responsibility in their use.

JORDI CAT

References

Black, M. (1962), *Models and Metaphors*. Ithaca, NY: Cornell University Press.

——— (1993), "More about Metaphor," in A. Ortony (ed.), *Metaphor and Thought*. New York: Cambridge University Press, 19–41.

Boyd, R. (1993), "Metaphor and Theory Change: What Is 'Metaphor' a Metaphor For?" in A. Ortony (ed.), *Metaphor and Thought*. New York: Cambridge University Press, 481–532.

Cantor, G. N., and J. R. R. Christie (eds.) (1987), *The Figural and the Literal*. Manchester, UK: Manchester University Press.

Cat, J. (2001), "On Understanding: Maxwell and the Methods of Illustration and Scientific Metaphor," in *Studies in History and Philosophy of Modern Physics* 23: 295–441.

Hesse, M. (1966), *Models and Analogies in Science*. Notre Dame, IN: University of Notre Dame Press.

——— (1993), "Models, Metaphors and Truth," in F. R. Ankersmit and J. J. A. Mooji (eds.), *Knowledge and Language*. Dordrecht, Netherlands: Kluwer.

——— (1995), "Models, Metaphors and Truth," in Z. Radman (ed.), *From a Metaphorical Point of View: A Multidisciplinary Approach to the Cognitive Content of Method*. New York: Walter de Gruyter, 351–372.

Holton, G. (1986), *The Advancement of Science and Its Burdens*. New York: Cambridge University Press.

Johnson, M., and N. Tuana (1986), "The Rationality of Creativity in Science," in D. DeLuca (ed.), *Essays on Creativity and Science*. Honolulu: Hawaii Council of Teachers of English, 225–235.

Keller, E. F. (1985), *Reflections on Gender and Science*. New Haven, CT: Yale University Press.

——— (1995), *Refiguring Life*. New York: Columbia University Press.

Knorr-Cetina, K. (1995), "Metaphors in the Scientific Laboratory: Why Are They There and What Do They Do?" in Z. Radman (ed.), *From a Metaphorical Point of View: A Multidisciplinary Approach to the Cognitive Content of Method*. New York: Walter de Gruyter: 329–350.

Kuhn, T. S. (1993), "Metaphor in Science," in A. Ortony (ed.), *Metaphor and Thought*. New York: Cambridge University Press: 533–542.

Leatherdale, W. H. (1974), *The Role of Analogy, Model and Metaphor in Science*. Amsterdam: North-Holland Publishing Co.

McCormach, E. R. (1971), "Meaning Variance and Metaphor," *British Journal for Philosophy of Science* 22: 145–159.

Mirowski, P. (1989), *More Heat Than Light: Economy as Social Physics, Physics as Nature's Economics*. New York: Cambridge University Press.

Radman, Z. (ed.) (1995), *From a Metaphorical Point of View: A Multidisciplinary Approach to the Cognitive Content of Method*. New York: Walter de Gruyter.

Soskice, J. M., and R. Harre (1995), "Metaphor in Science," in Z. Radman (ed.), *From a Metaphorical Point of View: A Multidisciplinary Approach to the Cognitive Content of Method*. New York: Walter de Gruyter: 289–308.

See also **Empiricism; Explanation; Logical Empiricism; Realism; Scientific Models; Theories**

SCIENTIFIC MODELS

Models are of central importance in many scientific contexts. Cases in point are the roles played in their respective domains by the MIT bag model of the nucleon, the billiard ball model of a gas, the Bohr model of the atom, the Pauling model of chemical bonds, the Gaussian-chain model of a polymer, the Lorenz model of the atmosphere, the Lotka-Volterra model of predator/prey interaction, agent-based and evolutionary models of social interaction, and general equilibrium models of markets.

This importance has been increasingly recognized by philosophers. As a result, the philosophical literature on models has been growing rapidly over the last decades, and with it the number of different types of models that philosophers recognize. Some of the notions used as categories have created phenomenological models, computational models, developmental models, explanatory models, impoverished models, testing models, idealized models, theoretical models, scale models, heuristic models, caricature models, didactic models, fantasy models, toy models, imaginary models, mathematical models, substitute models, iconic models, formal models, analog models, and instrumental models. The key to coming to terms with this variety is to realize that these different categories pertain to these different issues that arise in connection with models:

1. Semantics: What is the representational function that models perform?
2. Ontology: What kind of things are models?
3. Epistemology: How does one learn with models?
4. Models and theory: How do models relate to theory?
5. Models and other debates in the philosophy of science:
    (a) Models and the realism versus antirealism debate
    (b) Models and reductionism
    (c) Models and laws of nature
    (d) Models and scientific explanation

Semantics: The Representational Functions of Models

Models can perform two fundamentally different representational functions. On the one hand, a model can be a representation of a selected part of the world (the "target system"). Depending on the nature of the target, such models are either models of phenomena or models of data. On the other hand, a model can represent a theory in the sense that it interprets the laws and axioms of that theory. These two notions are not mutually exclusive and scientific models can at once be representations in both senses.

Representational Models I: Models of Phenomena

Many scientific models represent a phenomenon, where 'phenomenon' is used as an umbrella term covering all relatively stable and general features of the world that are interesting from a scientific point of view. Well-known examples of models of this kind include the billiard ball model of a gas, the Bohr model of the atom, the double helix model of DNA, the scale model of a bridge, the Mundell-Fleming model of an open economy, and the Lorenz model of the atmosphere. The representational function of these models is widely acknowledged among philosophers; but despite the ubiquity of representation talk in the literature on models, the issue of scientific representation as regards models has barely been recognized, much less seriously discussed.

A first step toward a discussion of this issue is to realize that there is no such thing as the problem of scientific representation. Rather, there are different but related problems. It is not yet clear what the specific set of questions is that a theory of representation should come to terms with, but two problems in particular seem to occupy center stage in tackling the issue (Frigg 2003, chap. 1). The first problem is to explain in virtue of what a model is a representation of something else; or more formally: What fills the blank in '*M* represents *T* if and only if ______,' where *M* is a model and *T* a target system? Somewhat surprisingly, this question did not attract much attention in twentieth-century philosophy of science.

The second problem is concerned with representational styles (see Scientific Style). It is a commonplace that one can represent the same subject matter in different ways. Weizsäcker's liquid-drop model represents the nucleus of an atom in a manner very different from the shell model, and a scale model of the wing of an airplane represents the shape of the wing differently from how a mathematical model does. What representational styles are there in the sciences?

Although this question is not explicitly addressed in the literature on the so-called semantic view of theories (see Theories), two answers seem to emerge from its understanding of models. One version of the semantic view posits that a model and its target have to be isomorphic (Suppes 2002) or partially isomorphic (da Costa and French 2003) to each other. Another version drops isomorphism in favor of similarity (Giere 1988). This approach enjoys the advantage over the isomorphic view that it is less restrictive and also can account for cases of inexact and simplifying models.

Furthermore, one can understand the discussions about certain types of models as contributions to an investigation into representational styles.

Iconic Models An iconic model is supposed to be a naturalistic replica or a truthful mirror image of the target. Paradigm cases of iconic models are scale models such as wooden cars or model bridges, which are either enlarged or downsized copies of the original. More elaborate examples of iconic models can be found in the life sciences, where one particular organism (or group thereof) is investigated in order to find out something about the species to which it belongs. In a clinical trial, for instance, a certain number of patients are administered a drug, their reaction(s) to this drug is monitored, and the result is supposed to show how humans in general react to this drug.

What criteria does a model have to satisfy in order to qualify as an icon? Although there seem to be strong intuitions about how to answer this question in particular cases, no theory of iconicity for models has been formulated yet.

Idealized Models An idealization is a deliberate simplification of something complicated with the objective of making it more tractable. Most idealizations fall into either of two classes.

One class consists of cases in which idealization amounts to "stripping away" all properties from a concrete object that are believed not to be relevant to the problem at hand. This allows one to focus on a limited set of properties in isolation. An example from economics is the Philips curve, which specifies a relationship between inflation and unemployment, disregarding all other economic factors. This process of stripping away is often referred to as 'Aristotelian abstraction,' 'method of isolation,' or 'use of negligibility assumptions.'

The other class comprises idealizations that involve deliberate distortions. Physicists build models consisting of point masses moving on frictionless planes; economists assume that agents are perfectly rational; biologists study isolated populations, and so on. It was characteristic of Galileo's approach to science to use simplifications of this sort whenever a situation was too complicated to tackle. For this reason one can refer to this process as 'Galilean idealization' (cf. McMullin 1985).

Galilean idealizations are beset with riddles. What does a model involving distortions of this kind say about the world? How does one test its accuracy? In reply to these questions, Laymon (1991) has put forward a theory that understands idealizations as ideal limits: Imagine a series of experimental refinements of an actual situation that approach the postulated limit and then require that the closer the properties of a system come to the ideal limit, the closer its behavior has to come to the behavior of the ideal limit (monotonicity). But these conditions need not always hold, and it is not clear how to understand situations in which no ideal limit exists.

Galilean and Aristotelian idealizations are not mutually exclusive. On the contrary, they often come together. For instance, this happens in what is sometimes called 'caricature models,' which isolate a small number of main characteristics of a system and distort them into an extreme case.

Analogical Models Stock examples of analogical models include the hydraulic model of an economic system, the billiard ball model of a gas, the computer model of the mind, and the liquid-drop model of the nucleus. At the most basic level, two things are *analogous* if there are certain relevant similarities between them. Hesse (1963) distinguishes different types of analogies according to the kinds of similarity relations in which two objects enter. A simple type of analogy is based on shared properties. There is an analogy between the Earth and the moon based on the fact that both are large, solid, opaque, spherical bodies, receiving heat and light from the sun, revolving around their axes, and gravitating toward other bodies. But sameness of properties is not a necessary condition. An analogy between two objects can also be based on relevant similarities between their properties. In this more liberal sense, one can say that there is an analogy between sound and light because echoes are similar to reflections, loudness to brightness, pitch to color, detectability by the ear to detectability by the eye, and so forth.

Analogies can also be based on the sameness or resemblance of relations between parts of two systems rather than on their monadic properties. It is in this sense that some politicians assert that the relation of a parent to children is analogous to the relation of the state to citizens. The analogies mentioned so far have been what Hesse calls 'material analogies.' A more formal notion of analogy can be obtained by abstracting from the concrete features the systems possess and focusing only on their formal setup. What the analog model then shares with its target is not a set of features, but the same pattern of abstract relationships. This notion of analogy is closely related to what Hesse calls 'formal analogy.' Two items are related by formal analogy if they are both interpretations of the same formal calculus. For instance, there is a formal analogy between a swinging pendulum and an oscillating electric circuit because they are both described by the same mathematical equation.

A further distinction due to Hesse is among positive, negative, and neutral analogies. In comparing properties or relations between two items, positive analogies consist in those they share (both gas molecules and billiard balls have mass), while negative analogies consist in those they do not (billiard balls are colored, gas molecules are not). The *neutral analogy* comprises the properties not yet known to belong to either the positive or the negative analogy (do gas molecules obey Newton's laws of collision?). Neutral analogies play an important role in scientific research because they give rise to questions and suggest new hypotheses.

Phenomenological Models Phenomenological models have been defined in different, though related, ways. A standard definition takes them to be models that represent only observable properties of their targets and refrain from postulating hidden mechanisms and the like. Alternatively one can define phenomenological models as being independent of general theories. These two definitions, though not equivalent, often coincide in practice because hidden mechanisms or theoretical entities are commonly brought into a model via a general theory.

Each of these notions has its internal problems. But more pressing is the question of how the different notions relate to each other. Are analogies fundamentally different from idealizations, or do they occupy different areas on a continuous scale? How do icons differ from idealizations and analogies? At the present stage the answers to these questions are not known. What one needs is a systematic account of the different ways in which models can relate to the world and of how these ways compare with each other.

Representational Models II: Models of Data

Another kind of representational model is the *model of data* (Suppes 2002). A model of data is a corrected, rectified, regimented, and in many instances idealized version of the data gained from immediate observation, the so-called raw data. Characteristically, one first eliminates errors (e.g., removes points from the record that are due to faulty observation) and then presents the data in a "neat" way—for instance, by drawing a smooth curve through a set of points. These two steps are commonly referred to as data reduction and curve fitting. When investigating the trajectory of a certain planet, for instance, one first eliminates erroneous points from the observation records and then fits a smooth curve to the remaining ones. Models of data play a crucial role in confirming theories because it is the model and not the often messy and complex raw data that is compared with a theoretical prediction.

Both steps in the construction of a data model raise serious questions. How does one decide which points on the record need to be removed? And given a clean set of data, what curve can be fitted to it? The first question has been dealt with mainly within the context of the philosophy of experiment (see Experiment). At the heart of the latter question lies the so-called *curve fitting problem*, which is that the data themselves do not indicate what form the fitted curve should take. Traditional discussions of theory choice suggest that this issue is settled by background theory, considerations of simplicity, prior probabilities, or a combination of these. Forster and Sober (1994) point out that this formulation of the curve fitting problem is a slight overstatement because there is a theorem in statistics due to Akaike that shows (given certain assumptions) that the data themselves underwrite (though do not determine) an inference concerning the curve's shape if it is assumed that the fitted curve has to be chosen so that it strikes a balance between simplicity and goodness of fit in a way that maximizes predictive accuracy.

Models as the Thing Represented: Models of Theory

In modern logic, a model is a structure that makes all sentences of a theory true, where a *theory* is taken to be a set of sentences in a formal language, and a *structure* a set of objects along with the relations in which they enter. The structure represents the abstract theory in the sense that it interprets it and provides an object that embodies its essential features. As a simple example, consider Euclidean geometry, which consists of axioms (e.g., Any two points can be joined by a straight line) and the theorems that can be derived therefrom. Any structure of which all these statements are true is a model of Euclidean geometry.

Many models in science carry over from logic the idea of interpreting an abstract calculus. This is particularly pertinent in physics, where general laws—such as Newton's equation of motion—lie at the heart of a theory. These laws are applied to a particular system (e.g., a pendulum) by choosing a special force function, making assumptions about the mass distribution of the pendulum, etc. The resulting model, then, is an interpretation (or realization) of the general law.

Ontology: What Are Models?

Physical Objects

Some models are straightforward physical objects. These are commonly referred to as material models. The class of material models comprises anything that is a physical entity and that serves as a scientific representation of something else. Among the members of this class are wooden models of bridges, planes, and ships; analog models of neural systems resembling electric circuits or of an economy resembling lengths of pipe; and Watson and Crick's model of DNA. But material models also lend themselves to more cutting-edge cases, especially from the life sciences, where certain organisms are studied as stand-ins for others.

Material models do not give rise to any ontological difficulties over and above the well-known quibbles in connection with objects, which metaphysicians deal with (e.g., the nature of properties, the identity of objects, parts and wholes, and so on).

Fictional Objects

Many models are not material models. The Bohr model of the atom, a frictionless pendulum, or isolated populations are in the scientist's mind rather than in the laboratory, and they do not have to be physically realized and experimented upon to perform their representational function.

It seems natural to view them as fictional entities. This position can be traced back to the German neo-Kantian Vaihinger and has been advocated more recently by Giere (1988, Ch. 3), who calls them 'abstract entities.' The drawback of this suggestion is that fictional entities are notoriously beset with ontological riddles. This has led many philosophers, most prominently Quine, to argue that there are no such things as fictional entities and that apparent ontological commitments to them must be renounced (see Quine, Willard Van). This has resulted in a glaring neglect of fictional entities, in particular among philosophers of science.

Set-Theoretic Structures

An influential point of view takes models to be set-theoretic structures. This position can be traced back to Suppes' work in the 1960s and is now, with slight variants, held by most proponents of the semantic view of theories (see Theories).

This view of models has been criticized on different grounds. One pervasive criticism is that many types of models that play an important role in science are not structures and cannot be accommodated within the structuralist view of models, which can account neither for how these models are constructed nor for how they work in the context of investigation (Cartwright 1999; Morgan and Morrison 1999). Another charge held against the set-theoretic approach is that it is not possible to explain how structures represent a target system that forms part of the physical world without making assumptions that go beyond what the approach can afford (Frigg 2003, Chs. 2 and 3; Suárez 2003).

Descriptions

A time-honored position has it that what scientists display in scientific papers and textbooks when they present a model are more or less stylized descriptions of the relevant target systems.

This view has not been subject to explicit criticism. However, some of the criticisms that have been marshaled against the syntactic view of theories equally threaten a linguistic understanding of models. First, it is a commonplace that one can describe the same thing in different ways. But if one identifies a model with its description, then each new description yields a new model, which seems to be counterintuitive. Second, models have different properties than descriptions. On the one hand, one can say that the model of the solar system consists of spheres orbiting around a big mass or that the population in the model is isolated from its environment, but it does not seem to make sense to say this about a description. On the other hand, descriptions have properties that models do not have. A description can be written in English, consist of 517 words, be printed in red ink, and so on. None of this makes sense when said about a model.

Equations

Another group of things that are habitually referred to as models, in particular in economics, consists of equations (which are then termed 'mathematical models')—for instance, the Black-Scholes model of the stock market and the Mundell-Fleming model of an open economy.

The problem with this suggestion is that equations are syntactic items, and as such they face objections similar to the ones put forward against descriptions. First, one can describe the same situation using different coordinates and as a result obtain different equations; but one does not seem to obtain a different model. Second, the model has properties different from the equation. An oscillator is three-dimensional, but the equation describing its motion is not. Equally, an equation may be inhomogenous while the system it describes is not.

Gerrymandered Ontologies

The proposals discussed so far have tacitly assumed that a model belongs to one particular class of objects. But this assumption is not necessary. It might be the case that models are a mixture of elements belonging to different ontological categories.

Epistemology: Learning with Models

Models are vehicles for learning about the world. By studying a model one can discover features of the system the model stands for. This cognitive function of models has been widely acknowledged in the literature, and some even suggest that models give rise to a new style of reasoning, called 'model-based reasoning' (Magnani and Nersessian 2002). This leaves one with the question of how learning with a model is possible.

Hughes (1997) provides a general framework for discussing this question. According to his "DDI" account of modeling, learning takes place in three stages: *d*enotation, *d*emonstration, and *i*nterpretation. One begins by establishing a representation relation (denotation) between the model and the target. Then one investigates the features of the model in order to demonstrate certain theoretical claims about its internal constitution or mechanism; i.e., one learns about the model (demonstration). Finally, these findings have to be converted into claims about the target system; Hughes refers to this step as 'interpretation.' It is the latter two notions that are at stake here.

Learning About the Model: Experiments, Thought Experiments, and Simulation

Learning about a model happens at two places, in the construction and the manipulation of the model (Morgan and Morrison 1999). There are no fixed rules for model building, and so the very activity of figuring out what and how a model fits together affords an opportunity to learn about the model. Once the model is built, one learns about its properties not by looking at it, but by using and manipulating it to elicit its secrets.

Depending on what kind of model one is dealing with, building and manipulating a model employs different activities demanding a different methodology. Material models seem to be unproblematic, as they are commonly used in the kind of experimental contexts that have been discussed extensively by philosophers of science (the model of a car is put in the wind tunnel to measure its air resistance). This is not the case with fictional models. What constraints are there to the construction of fictional models, and how does one manipulate them? The natural response seems to obtain an answer to these questions by performing a thought experiment. Different authors have explored this line of argument but they have reached very different and often conflicting conclusions as to how thought experiments are performed and what the status of their outcomes is (Hitchcock 2004, Chs. 1 and 2).

An important class consists of mathematical models. In some cases it is possible to derive results or solve equations analytically. But quite often this is not the case. It is on this point that the invention of the computer has had a great impact, as it allows one to solve equations that are otherwise intractable by making a computer simulation. Many parts of current research in both the natural and social sciences rely on computer simulations. To mention only a few examples, computer simulations are used to explore the formation and development of stars and galaxies, the detailed dynamics of high-energy heavy-ion reactions, aspects of the intricate process of the evolution of life, and factors determining the outbreak of wars, the progression of an economy, decision procedures in an organization, and moral behavior.

What is a simulation? Simulations characteristically are used in connection with dynamic models, i.e., which involve time. The aim of a simulation is to solve the equations of motion of such a model, which is designed to represent the time evolution of its target system. So, one can say that a simulation represents one process by another process (Hartmann 1996; Humphreys 2004).

It has been claimed that computer simulations constitute a genuinely new methodology of science, or even a new scientific paradigm (Humphreys 2004). Although this contention may not meet with univocal consent, there is no doubt about the practical significance of computer simulations. In situations in which the underlying model is well confirmed and understood, computer experiments may even replace real experiments, which has economic advantages and minimizes risk (as, for example, in the case of the simulation of atomic explosions). Computer simulations are also heuristically important. They may suggest new theories, models, and hypotheses, for example, based on a systematic exploration of a model's parameter space.

But computer simulations also bear methodological perils, as they may provide misleading results. In many cases the relevant variables are continuous. But due to the discrete nature of the calculations carried out on a computer, they do not allow for an exploration of the full range of the variables, and therefore may not reveal certain important features of the model.

Converting Knowledge About the Model into Knowledge About the Target

Once knowledge about the model is available, it has to be "translated" into knowledge about the target system. It is at this point that the representational function of models becomes important again. Models can provide information about the nature of their target systems only if one assumes that (at least some of) the model's aspects have counterparts in the world. But if learning is tied to representation and if there are different kinds of representation (analogies, idealizations, etc.), then there are also different kinds of learning. If, for instance, one has a model that is taken to be a realistic depiction, the transfer of knowledge from the model to the target is accomplished in a different manner than when one deals with an analog model or a model that involves idealizing assumptions.

What are these different ways of learning? Although numerous case studies have been made of how certain specific models work, there do not seem to be any general accounts of how the transfer of knowledge from a model to its target is achieved (with the possible exception of theories of analogical reasoning; see references above). This is a difficult question, but it is one that deserves more attention than it has received so far.

Models and Theory

One of the most perplexing questions in connection with models is how they relate to theories. The separation between models and theory is a very hazy one, and in the jargon of many scientists it is often difficult, if not impossible, to draw a line. So the question is: Is there a distinction between models and theories, and if so, how do they relate to one another?

In common parlance, the 'model' and 'theory' are sometimes used to express someone's attitude toward a particular piece of science. The phrase "It's just a model" indicates that the hypothesis at stake is asserted only tentatively, while something is awarded the labeled 'theory' if it has acquired some degree of general acceptance. However, this way of drawing a line between models and theories is of no use to a systematic understanding of models.

The Two Extremes: The Syntactic and the Semantic View of Theories

The syntactic view of theories, which is an integral part of the logical empiricist picture of science, construes a theory as a set of sentences in an axiomatized system of first-order logic (see Theories). Within this approach, the 'model' is used in both a wider and a narrower sense. In the wider sense, a model is just a system of semantic rules that interpret the abstract calculus, and the study of a model amounts to scrutinizing the semantics of a scientific language. In the narrower sense, a model is an alternative interpretation of a certain calculus. If, for instance, one takes the mathematics used in the kinetic theory of gases and reinterprets the terms of this calculus so that they refer to billiard balls, the billiard balls are a model of the kinetic theory of gases. Proponents of the syntactic view believe such models to be irrelevant to science. Models, they hold, are superfluous additions that are at best of pedagogical, aesthetical, or psychological value (cf. Bailer-Jones 1999).

The semantic view of theories reverses this standpoint and declares that one should dispense with a formal calculus altogether and view a theory as a family of models (see Theories). Although different versions of the semantic view assume a different notion of model, they all agree that models are the central unit of scientific theorizing.

Models as Independent of Theories

One of the most perspicuous criticisms of the semantic view is that it mislocates the place of models in the scientific edifice. Models are relatively independent from theory, rather than being constitutive of them; or to use Morrison's (1998) phrase, they are "autonomous agents." This independence has two aspects: construction and functioning (Morgan and Morrison 1999).

A look at how models are constructed in actual science shows that they can be derived entirely from neither data nor theory. Theories do not provide algorithms for the construction of a model; model building is an art and not a mechanical procedure. The London model of superconductivity is a good example: The model's principal equation has no theoretical justification and is motivated solely on the basis of phenomenological considerations (Cartwright 1999).

The second aspect of the independence of models is that they perform functions that they could not perform if they were a part of, or strongly dependent on, theories.

Models as Complements of Theories A theory may be incompletely specified in the sense that it imposes only certain general constraints but remains silent about the details of concrete situations, which are provided by a model (Redhead 1980). A special case of this situation is if a qualitative theory is known and the model introduces quantitative measures. Redhead's example for a theory that is underdetermined in this way is axiomatic quantum field theory, which imposes only certain general constraints on quantum fields but does not provide an account of particular fields.

While Redhead and others seem to think of cases of this sort as somehow special, Cartwright (1983) has argued that they are the rule rather than the exception. In her view, fundamental theories such as classical mechanics and quantum mechanics do not represent anything at all, as they do not describe any real-world situation. Laws in such theories are schemata that need to be concretized and filled with the details of a specific situation, which is a task that is accomplished by a model.

Models Stepping in When Theories Are Too Complex to Handle Theories may be too complicated to handle. In such a case a simplified model may be employed that allows for a solution (Redhead 1980). Quantum chromodynamics, for instance, cannot easily be used to study the hadron structure of a nucleus, although it is the fundamental theory for this problem. To get around this difficulty, physicists construct a tractable phenomenological model (e.g., the MIT bag model) that effectively describes the relevant degrees of freedom of the system under consideration (Hartmann 1999). A more extreme case is the use of a model when there are no theories available at all—take Bohr's model of the atom at the time he proposed it. The models scientist then construct to tackle this situation are sometimes referred to as "substitute models."

Models as Preliminary Theories The notion of models as substitutes for theories is closely related to "developmental models," which consist of cases in which models are some sort of a preliminary exercise to theory. A closely related notion is that of probing models (also known as 'study models' or 'toy models'). These are models that do not perform a representational function and that are not expected to provide information about anything beyond the model itself. The purpose of these models is to test new theoretical tools that are used later to build representational models (cf. Wimsatt 1987).

Models and Other Debates in the Philosophy of Science

The debate about scientific models has important repercussions for other debates in the philosophy of science. The reason for this is that traditionally the debates about realism, reductionism, explanation, and laws were couched in terms of theories, because only theories were acknowledged as carriers of scientific knowledge. So the question is whether, and if so how, discussions of these matters change when shifting the focus from theories to models. Up to now, no comprehensive model-based accounts of any of these issues have been developed, but models did leave some traces in the discussions of these topics, and it is these traces that will be dealt with in this section.

Models and the Realism Versus Antirealism Debate

It has been claimed that the practice of model building favors antirealism over realism (see Instrumentalism; Realism). Antirealists point out that truth is not the main goal of scientific modeling. Cartwright (1983), for instance, presents several case studies illustrating that good models are often false. Realists reply that a good model, though not literally true, is usually at least approximately true. In this vein, it has been argued that by relaxing idealizations (de-idealization) the predictions of the model typically become better, which is taken to be evidence for realism (cf. McMullin 1985; Nowak 1979).

Apart from the usual complaints about the elusiveness of the notion of approximate truth, antirealists have criticized this reply as flawed for two related reasons. First, there is no in-principle reason to assume that one can always improve the model by adding de-idealizing corrections. Second, it seems that the outlined procedure is not in accordance with scientific practice, in which it is unusual for scientists to try to repeatedly de-idealize an existing model. Rather, they shift to a completely different modeling framework once the needed adjustments get too complicated. A further difficulty with de-idealization is that most idealizations are not "controlled." For example, it is not clear in which way one has to de-idealize the MIT bag model to eventually arrive at quantum chromodynamics, the supposedly correct underlying theory.

The antirealist "incompatible models argument" takes as its starting point the observation that scientists often use several incompatible models of one and the same target system for predictive purposes. There are, for example, numerous models of a gas or the atomic nucleus. These models seemingly contradict each other as they ascribe different properties to the target system. This seems to cause problems for realists, as they typically hold that there is a close connection between the predictive success of a model and its being at least approximately true. But if several theories of the same system are predictively successful, and if these theories are mutually inconsistent, they cannot all be true.

Realists can react to this argument in three ways—first, by challenging the claim that the models in question are indeed predictively successful; second, by defending a version of perspectival realism, according to which each model reveals one aspect of the phenomenon in question; and, finally, by denying that there is a problem in the first place, because scientific models, which strictly speaking are always false, are just the wrong vehicles to make a point about realism.

Models and Reductionism

The existence of a multiplicity of models raises the question of how different models are related. A simple picture of the organization of science along the lines of Nagel's model of reduction or Oppenheim and Putnam's pyramid picture does not seem to be compatible with the practice of modeling (see also Reductionism). But which picture of science is?

Cartwright (1999) and others have suggested a picture of science according to which there are no systematic relations between different theories and models. All theories and models are tightened together only because they apply to the same domain of phenomena but do not enter into any further relations (deductive or otherwise). One is confronted with a patchwork of theories and models, all of which hold, *ceteris paribus*, in their specific domains of applicability. Some argue that this picture is at least partially incorrect because there are various types of interesting relations that hold between different models or theories. These relations range from those of controlled approximations over a singular limit to rather loose relations. Some have even argued that, at least within the context of biology, models play an essential role in a reductionist enterprise (Sarkar 1998). These suggestions have been made on the basis of case studies, and it remains to be seen whether a more general account of these relations can be given and a deeper justification for them provided (e.g., in a Bayesian framework) (see Bayesianism).

Models and Laws of Nature

It is widely held that science aims at discovering laws of nature. Philosophers, in turn, have been faced with the challenge of explicating what laws of nature are (see Laws of Nature). According to the two currently dominant accounts—the best-systems approach and the universals approach—laws of nature are understood to be universal in scope, meaning that they apply to everything that there is in the world. This take on laws does not seem to square with a view that assigns models a center stage in scientific theorizing. What role do general laws play in science if models are what represent what is happening in the world?

One possible response is to argue that laws of nature govern entities and processes in a model rather than in the world. Fundamental laws, in this approach, do not state facts about the world but hold true of entities and processes in the model (cf. Cartwright 1983).

Models and Scientific Explanation

Laws of nature play an important role in many accounts of explanation, most prominently in the deductive-nomological model and the unification approach (see Explanation). Unfortunately, these accounts inherit the problems that beset the relationship between models and laws. This leaves two options. Either one can argue that laws can be dispensed with in explanations, an idea employed both in van Fraassen's pragmatic theory of explanation and in certain causal accounts of explanation. Or one can shift the explanatory burden on models. A positive suggestion along these lines is Cartwright's (1983) "simulacrum account of explanation," which suggests that one explains a phenomenon by constructing a model that fits the phenomenon into the basic framework of a grand theory (chap. 8). In this account, the model itself is the explanation that is sought. This squares well with basic scientific intuitions but leaves the question of what notion of explanation is at work. Other accounts of explanation do not seem to be more hospitable to models. Causal or mechanistic accounts of explanation (see Explanation, Mechanism) do not assign models an explanatory function and, at best, regard them as tools to find out about the causal relations that hold between certain parts of the world.

Conclusion

Models play an important role in science. But despite the fact that they have generated considerable interest among philosophers, there remain significant lacunae in the philosophical understanding of what models are and of how they work.

Roman Frigg
Stephan Hartmann

References

Bailer-Jones, Daniela (1999), "Tracing the Development of Models in the Philosophy of Science," in L. Magnani, N. J. Nersessian, and P. Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*. New York: Kluwer Academic/Plenum, 23–40.

Cartwright, Nancy (1983), *How the Laws of Physics Lie*. Oxford: Oxford University Press.

——— (1999), *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

da Costa, Newton, and Steven French (2003), *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford University Press.

Forster, Malcolm, and Elliot Sober (1994), "How to Tell When Simple, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions," *British Journal for the Philosophy of Science* 45: 1–35.

Frigg, Roman (2003), *Re-Presenting Scientific Representation*, Ph.D dissertation, University of London.

Giere, Ronald (1988), *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Hartmann, Stephan (1996), "The World as a Process: Simulations in the Natural and Social Sciences," in Rainer Hegselmann, Ulrich Müller, and Klaus Troitzsch (eds.), *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*. Theory and Decision Library. Dordrecht, Netherlands: Kluwer, 77–100.

Hartmann, Stephan (1999), "Models and Stories in Hadron Physics," in M. Morgan and M. Morrison (eds.), *Models as Mediators: Perspectives on Natural and Social Science*, 326–346.

Hesse, Mary (1963), *Models and Analogies in Science*. London: Sheed and Ward.

Hitchcock, Christopher (ed.) (2004), *Contemporary Debates in Philosophy of Science*. Malden, MA, and Oxford: Blackwell.

Horowitz, T., and G. T. Massey (eds.) (1991), *Thought Experiments in Science and Philosophy*. Savage, MD: Rowman and Littlefield.

Hughes, R. I. G. (1997), "Models and Representation," *Philosophy of Science* 64 (Proceedings): S325-S336.

Humphreys, Paul (2004), *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford: Oxford University Press.

Laymon, Ronald (1991), "Thought Experiments by Stevin, Mach and Gouy: Thought Experiments as Ideal Limits and Semantic Domains," in Horowitz and Massey, 167–191.

Magnani, Lorenzo, and Nancy Nersessian (eds.) (2002), *Model-Based Reasoning: Science, Technology, Values*. Dordrecht, Netherlands: Kluwer.

McMullin, Ernan (1985), "Galilean Idealization," *Studies in the History and Philosophy of Science* 16: 247–273.

Morgan, Mary, and Margaret Morrison (1999), "Models as Mediating Instruments," *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press, 10–37.

Morrison, Margaret (1998), "Modelling Nature: Between Physics and the Physical World," *Philosophia Naturalis* 35: 65–85.

Nowak, Leszek (1979), *The Structure of Idealization: Towards a Systematic Interpretation of the Marxian Idea of Science*. Dordrecht, Netherlands: Reidel.

Redhead, Michael (1980), "Models in Physics," *British Journal for the Philosophy of Science* 31: 145–163.

Sarkar, Sahotra (1998), *Genetics and Reductionism*. Cambridge: Cambridge University Press.

Suárez, Mauricio (2003), "Scientific Representation: Against Similarity and Isomorphism," *International Studies in the Philosophy of Science* 17: 225–244.

Suppes, Patrick (2002), *Representation and Invariance of Scientific Structures*. Stanford, CA: CSLI Publications.

Wimsatt, William (1987), "False Models as Means to Truer Theories," in N. Nitecki and A. Hoffman (eds.), *Neutral Models in Biology*. Oxford: Oxford University Press, 23–55.

See also **Laws of Nature; Theories**

SCIENTIFIC PROGRESS

The concept of 'progress' presupposes some aim or end against which progress can be measured. In order to talk meaningfully about the progress of science, one must have in mind some concept of the aim of science. What, then, is the aim of science? In Francis Bacon's famous phrase, "Knowledge is power." Insofar as that dictum was taken as the catchword of the emerging Scientific Revolution, it was understood to mean the power to accurately predict, manipulate, and control the natural world. There can be no question that in this sense more is now known than in past generations and that science, so understood, has progressed from its beginnings to the present. The future prospects appear bright as well, although perhaps not unlimited.

Predictive power is an important part of the story of scientific progress but it cannot be the whole story. Were there access to an infallible oracle, the power to predict would be unlimited, but one would still not be inclined to call it science. For one thing, there would be no understanding of why the oracle always got it right. So, scientific progress is tied up with the notion of increased understanding of how the world works. This understanding typically takes the form of some story, theory, or narrative. 'Understanding,' in turn, seems to be connected to the question of truth. Producing a coherent narrative about some sequence of events is not sufficient to guarantee understanding. One wants to know that the narrative is true or, at least, approximately so. Hence, another measure of scientific progress is an increase in or convergence to truth. For progress to occur, one demands that successor theories or narratives be truer than their predecessors. However, if two accounts appear to be equally true, the notion of explanatory power may be invoked. In this dimension, progress occurs when a less explanatory account is replaced by a more explanatory account. Many have argued not only that science progresses but that the progress is "rational." The most trenchant criticisms of Kuhn's (1973) analysis of scientific revolutions were directed toward the alleged "irrationality" of scientific change. The rationality of scientific progress, in turn, is taken by some to be grounded in a commitment to scientific realism (see Realism). So, for some realists, progress occurs in virtue of scientists' accounts getting closer to the truth about how the world is. The explanatory power and predictive accuracy of accounts is alleged to rest in the fact that these accounts are getting at the way the world is, more or less. When it comes time to spell out this picture in detail, however, things become problematic. In what follows, some of the major themes that have emerged in the twentieth-century discussion of the scope and limits of scientific progress will be sketched.

The Logical Positivists

The positivists are perhaps best remembered for their efforts to promote the unity of science, a project that culminated in the *International Encyclopedia of Unified Science* (Neurath, Carnap, and Morris 1952). The project promoted the idea that there was one language of science, a unity of method that became codified as the hypothetico-deductive method. This is often interpreted to imply that there is also a unity of laws that generated a reductionism with physics as the most basic science serving as a foundation for chemistry, biology, and psychology and the other social sciences (see Reductionism; Unity and Disunity of Science). Correlative with this reductionist worldview was a commitment to what became known as the covering-law model of scientific explanation (see Explanation). Associated with the covering-law model of explanation is a hierarchical model of explanatory connections that has been labeled the "layer cake" model. The basic idea is that there is an observational base that contains singular statements reporting observational facts, for example, about charge distributions, chemical reactions, embryonic developments, animal behaviors, social groups, and so on. At the next level up, there will be empirical generalizations about these phenomena. The third stage will include even more general statements connecting the generalizations at the first level. At some stage, theoretical constructs will be introduced. At some level, the sociological laws and theories will be subsumed under the psychological laws and theories, and they in turn will be subsumed under biological laws, and these in turn will be subsumed under laws of chemistry and finally under laws of physics. This procedure is conceived as continuing indefinitely. Each rise in level brings greater systematization to the body of established scientific knowledge and constitutes scientific progress.

In the Aristotelian model and in the rationalist tradition, this process of looking for more and more general laws would eventually end at some very general laws that are "self-evident." This is one aspect of what some logical empiricists have called the "search for certainty," which characterized traditional philosophy. For the modern empiricists, the highest-level laws always function as "unexplained explainers." Thus, the search for ever more general laws continues, in principle, indefinitely. Even if one was to arrive at some "final theory," whatever that might mean, the fallibilism of the empiricists would preclude one's ever recognizing it. The net effect is a view of endless progress (see Logical Empiricism).

The Popperian View

In the 1930s, Karl Popper developed an alternative to the positivist picture of science (see Popper, Karl Raimund). Popper accepted the positivists' view of the nature of theories and the importance of the hypothetico-deductive method but parted company with them on what he saw as the crucial question of how scientific claims are to be validated. The positivists adopted the view that theories were to be validated by being confirmed by evidence. Measures of confirmation were generally taken to be probability measures of some sort. In a choice between two theories, the rule was: choose the theory that is most probable given the evidence. Popper rejected this approach in favor of emphasizing the falsifiability of scientific claims. In this view, the job of the scientist is to subject hypotheses and conjectures to the severest possible tests and to provisionally accept those that passed the most severe (Popper 1961, 1968).

For Popper, the philosopher of science is a methodologist whose job it is to propose a series of methodological rules that will promote the growth of knowledge, that is, the discovery of general laws. These methodological choices are partially a matter of adopting certain conventions based on value assumptions about the aim and nature of science. For Popper, the appropriate choice is the methodology that maximizes the solution of interesting problems. The result encourages methodologies that involve bold conjectures and severe testing of hypotheses. The severe tests to which hypotheses are to be put is a reflection of a critical method that promotes rational change. Indeed, Popper identifies 'being rational' with 'being critical.' In this way, the progress of science, insofar as it is produced by a critical evaluation of hypotheses, is a rational endeavor.

Popper's evolutionary model of scientific change consists of conjectures and refutations. Problems give rise to hypotheses that are then subjected to severe tests. If the hypothesis fails to pass those tests, it is rejected and another conjectured takes its place. The cycle of testing and assessing begins anew. If the hypothesis passes one severe test, then there will always be others more severe yet. Theories or conjectures that pass such severe tests are said to be *corroborated*. The degree of corroboration of a theory is a measure of how well it has stood up to severe tests in the past. But it says nothing about the likelihood of the theory withstanding future tests. So, in what sense is the most highly corroborated theory the best choice in a given situation? Popper's argument for identifying the most highly corroborated theory as the best goes along lines like these:

1. The rational choice is the best choice.
2. It is always rational to choose that option that has been most severely tested and survived, because those theories are the ones that have been subjected to the severest criticism. (Recall that, for Popper, to be rational is to be critical.)
3. In the case of theories, it is rational, then, to choose that theory among the competing options that is most highly corroborated. Therefore,
4. The best theory is the most highly corroborated (see Corroboration).

The most highly corroborated theories also "track" the truth, in the sense that Popper argued that a measure of closeness to the truth that he called "verisimilitude" could be defined. With such a measure, it would be possible in principle to determine which of a set of alternative conjectures represented genuine scientific progress. This proposal generated an extensive literature designed to refine and validate such a measure. For the most part, such attempts to construct viable measures of verisimilitude have come to naught, although there are still defenders of this approach as the best method for gauging scientific progress (cf. Miller 1974; Niiniluoto 1999) (see Verisimilitude).

The Kuhnian View

Thomas Kuhn's (1962) influential *Structure of Scientific Revolutions* was seen at the time as a radical challenge to traditional views about the nature of science and the prospects of scientific progress. It is well to remember that *Structure* was in fact the last number in that canon of logical positivism, the *International Encyclopedia of Unified Science*. Nonetheless, it is fair to say that Kuhn in the 1960s saw himself as breaking with that tradition. Over the years, he came to see the continuity that existed between his views and those of the later positivists. He also came to refine and modify his views in the light of what he took to be the excessive lengths to which some of his readers took them. The central themes remain, however, much as he laid them out in the 1960s (see Kuhn, Thomas).

The growth of a scientific discipline, in Kuhn's view, follows a standard pattern of different stages. In his earliest formulation of this stance, it is possible to distinguish five stages that characterize the progress of science. These are:

1. Immature science
2. Mature (normal) science
3. Crisis science
4. Revolutionary science
5. Resolution; normal science resumed.

The cycle then repeats itself through stages 2–5 indefinitely. Central to Kuhn's view is the notoriously slippery concept of a "paradigm," or "disciplinary matrix." A disciplinary matrix is the overall collection of methods, formulae, rules, procedures, and commitments that govern scientific research. Kuhn came to distinguish four major components of disciplinary matrices: symbolic generalizations, models, values, and exemplars. These are the shared standard examples that give "content" to the abstract principles of the disciplinary matrix. The exemplars are the fundamental units in the matrix and are the basic tools by which the scientist working in a normal science tradition advances the range of phenomena rendered lawlike by the basic principles and theories of the tradition.

In *Structure*, immature science is characterized as preparadigmatic. Later, Kuhn argued that no research occurs in the absence of paradigms. Immature science is science that is characterized by paradigms, which for some reason or other, fail to generate a "puzzle-solving" tradition. The shift from immature to mature science is then seen as a shift from a paradigm that "leads nowhere" to one that provides a context of unsolved puzzles and problems such that some hope exists for their solution.

Mature (normal) science is thus a problem-solving tradition. According to Kuhn, most scientists spend most of their lives working in such traditions. The normal science tradition seeks to extend and entrench tried-and-true theories and practices. Progress at this stage consists in increasing the number of systems that can be understood in terms of the fundamental exemplars of the theory. However, no theory (or paradigm) successfully solves all its problems. Problems that resist assimilation to the techniques of the paradigm are labeled 'anomalies.' When the failure of a paradigm to reduce anomalies to lawfulness is perceived (by scientists) to be as great or greater than the power of the paradigm to force nature into its mold, a crisis results.

Crisis science is characterized by a general recognition that the ruling paradigm is no longer functioning effectively. This recognition may come from the community of scientists working within the tradition of the disabled paradigm, or it may come from without. With the paradigm breakdown comes a proliferation of new ideas, theories, methods, and alternative paradigms.

When one of the new paradigms begins to emerge as a contender for succession, the result is a conflict between the new and the old paradigm. Since paradigms are pervasive in determining worldviews, the ground rules for deciding among competing paradigms are not the same as those that operate within a single tradition. Kuhn argues that not only are the worldviews of different paradigmatic traditions different, but, in a sense, the *world itself is different* for the practitioners in different paradigmatic traditions.

The ultimate resolution of paradigm conflicts results in the emergence of a new normal science tradition. According to Kuhn, such resolutions involve something akin to a Gestalt shift. Critics were quick to conclude that, for Kuhn, scientific progress from one paradigm to another is a fundamentally irrational process, since the standards and values of one paradigm need not be shared by its successor. In responding to his critics, Kuhn sought to soften this implication by pointing out that there were certain values characteristic of science, such as simplicity, predictive accuracy, and requirements for consistency, that transcended particular disciplinary matrices. Thus, he argued, revolutionary science was both progressive and rational.

Post-Kuhnian Developments

Among the post-Kuhnians was Imre Lakatos, who developed a blend of Kuhnian and Popperian elements that he labeled the 'methodology of scientific research programs' (see Lakatos, Imre; Research Programs). He argued that the fundamental unit of assessment should be a research tradition such as Newtonian mechanics, rather than a single hypothesis or conjecture. These research programs consisted of a "hard core" of central symbolic generalizations, along with a "protective belt" of auxillary hypotheses. Programs were either progressive, when they made bold and stunning predictions, or regressive, when they focused on protecting the core from refutation. The resulting model of scientific development was claimed to be a sophisticated falsificationist version of Popper's program that offered a response to what Lakatos saw as Kuhn's irrationalist view of scientific change. (For critical assessments, see Dilworth 1994 and Laudan 1977.)

Larry Laudan proposes that scientific progress be measured in terms of the problem-solving capacity of a research tradition. His conception of a research tradition is, despite protestations, similar to Kuhn's conception of a disciplinary matrix. Laudan suggests that instead of understanding progress as a rational activity, one defines rationality in terms of progress. So, given two problem-solving research traditions, the rational choice is to choose that which is most progressive in terms of its problem-solving ability. Problems, on Laudan's view, are either empirical or conceptual. Empirical problems are either solved, unsolved, or anomalous. Conceptual problems are either internal or external to the tradition. The degree of progressiveness is taken to be a measure of the solved problems minus the unsolved and anomalous. This approach presumes that problems can be characterized more or less independently of the traditions that address them. This is certainly true for a large group of problems, but not for all. To the extent that problems are identified with the traditions within which they occur, the attempt to compare alternative traditions in terms of their problem-solving capacity is compromised. In any case, this analysis succeeds no better than any of the others in providing an effective procedure for determining scientific progress.

Philip Kitcher (1993) argues that the traditional positivist picture, which he dubbed "legend," while flawed, is basically correct in its picture of the progressive unifying nature of science. By exploring the multidimensional practices that constitute scientific inquiry, he paints a broadly realistic conception of science that allows for objective, rational progress.

Ilkka Niiniluoto opts for a "realist" measure of progress based on the notion of verisimilitude, or "nearness" to the truth (see Verisimilitude). He rejects Laudan's problem-solving criterion on the ground that no truth-independent criterion can be a suitable measure of progress, although it may serve as a truth indicator. But the argument for this appears to be a presumption that the aim of science is to arrive at the truth. Indeed, if the aim of science is to produce true theories, then no truth-independent criterion of success or progress will be adequate. But what does this mean to a defender of a criterion of pragmatic problem solving capacity for the success of science?

Does scientific progress entail convergence to the truth? Kuhn (1962) hinted that the growth of scientific knowledge should be viewed more along the lines of aimless biological evolution rather than as a teleologically directed process (cf. Laudan 1977). He subsequently backed off from the more radical implications of this claim, but a preliminary sketch of a similar view was given by Stephen Toulmin (1972). Unfortunately, the promised development of this sketch never appeared. At the same time, however, Donald Campbell was developing an evolutionary model of scientific change that he called "hypothetical realism" (Campbell 1974). David Hull (1988) presents a gripping picture of science progressing to the truth as a result of the struggle between scientists for credit and acclaim in the open market of ideas (see Evolutionary Epistemology).

The Price of Progress

For the most part, those who endorse the progressive character of science see it continuing forever. There are exceptions, but these are in the minority (e.g., Horgan 1996). The problem with postulating the end of science is that those who profess it are extrapolating current conceptual and technological capacities in a way that history belies. Since the last century, Big Science has come into its own, first in physics but now in other fields as well. Making discoveries costs money, and no extended principled discussion of who should pay and who should benefit has yet been undertaken (cf. Rescher 1978; Kitcher 1993). At the beginning of the last century, Poincaré argued that science for its own sake was justification enough (Poincaré 1958). In the light of the development of weapons of mass destruction and the social implications of advances in biotechnology, to name just two areas of concern, this is a conceit that the twenty-first century can ill afford to harbor.

Michael E. Bradie

References

Aronson, Jerrold L, Rom Harré, and Eileen Cornell Way (1995), *Realism Rescued: How Scientific Progress is Possible*. Chicago: Open Court.

Campbell, D. T. (1974), "Evolutionary Epistemology," in P. A. Schilpp (ed.), *The Philosophy of Karl Popper*. LaSalle, IL: Open Court.

Dilworth, Craig (1994), *Scientific Progress: A Study Concerning the Nature of the Relation between Successive Scientific Theories*, 3rd ed. Dordrecht, Netherlands: Kluwer Academic.

Giere, Ronald N. (1988), *Explaining Science*. Chicago: University of Chicago Press.

Harré, Rom (ed.) (1975), *Problems of Scientific Revolution: Progress and Obstacle to Progress in the Sciences*. Oxford: Clarendon Press.

Horgan, John (1996), *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age*. Reading, MA: Addison-Wesley.

Hull, David (1988), *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.

Kitcher, Philip (1993), *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.

Kuhn, Thomas S. (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, Imre (1978), *The Methodology of Scientific Research Program*. Cambridge: Cambridge University Press.

Laudan, L. (1977), *Progress and Its Problems: Towards a Theory of Scientific Growth*. Berkeley and Los Angeles: University of California Press.

Miller, David (1974), "Popper's Qualitative Theory of Verisimilitude," *British Journal for the Philosophy of Science* 25: 166–177.

Neurath, O., R. Carnap, and C. Morris (eds.) (1952), *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press.

Newton-Smith, W. H. (1981), *The Rationality of Science*. London: Routledge & Kegan Paul.

Niiniluoto, Ilkka (1999), *Critical Scientific Realism*. Oxford: Oxford University Press.

Poincaré, Henri (1958), *The Value of Science*. Translated by George Bruce Halsted. New York: Dover.

Popper, Karl (1961), *The Logic Of Scientific Discovery*. New York: Science Editions, Inc.

——— (1968), *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper.

——— (1972), *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.

Rescher, Nicholas (1978), *Scientific Progress: A Philosophical Essay on the Economics of Research in Natural Science*. Pittsburgh: University of Pittsburgh.

Toulmin, Stephen (1972), *Human Understanding: The Collective Use and Evolution of Concepts*. Princeton, NJ: Princeton University Press.

See also **Evolutionary Epistemology; Kuhn, Thomas; Logical Empiricism; Popper, Karl Raimund; Unity and Disunity of Science; Unity of Science Movement**

SCIENTIFIC REVOLUTIONS

Since an explicit definition of 'scientific revolution' in terms of necessary and sufficient conditions is out of the question, this article will instead characterize revolutions by contrasting them with normal science and evolutionary development, by considering exemplary cases of revolution, and by noting the history of the the term 'revolution.'

Origins of the Idea of Scientific Revolution

The concepts of political and scientific revolution have been intertwined almost from the beginning. Hatto (1949) and Cohen (1985) have assembled the most thorough collection of historical sources. The ancients and medievals had no term for political revolution in the modern sense. However, they already drew an analogy to revolution in the sense of celestial orbits (*orbis* = cycle, circle, wheel), since they sometimes supposed that a city or state cycled through the gamut of political constitutions, from tyranny to democracy. Thus a revolution was a return to a previous state rather than something new and progressive. Such cycling was often considered natural or even fated, by the turn of the wheel of fortune, and therefore outside of human control. Astrological beliefs abetted the linking of political and celestial turnings. In the Italian Renaissance, Machiavelli and a few others suggested that states were at least partly artificial and that revolutions could be humanly made. Meanwhile, the Vulgar Latin 'revolutio' came to designate celestial revolution, and an analogy to the wheels of the new clockwork machinery apparently helped propagate the term into political contexts.

The title of Copernicus's work of 1543, *De revolutionibus orbium coelestium*, implied a cyclic return, a sense retained in early modern political contexts, as when the Stuart monarchy replaced the Rump Parliament in Britain in the Restoration of 1660. But soon after, 'revolution' came to suggest a political-structural change away from the status quo, as in the Glorious Revolution of 1688, although this, and the American Revolution a century later, could still be considered a return to an imagined previous condition. The French Revolution stabilized the sense of rejection of an old regime for something new. This meaning in turn passed back, metaphorically, into science, where Newtonian mechanics became the paradigmatic revolution in replacing both Aristotelian and Cartesian philosophy. During the nineteenth century the new chemistry of Lavoisier and Dalton and, later, Darwin's theory of evolution inspired talk of revolution; but it was not until the twentieth century that 'scientific revolution' became common parlance.

The shocking double revolution in physics (relativity and quantum mechanics) as well as the Russian Revolution stimulated thinking about revolution early in the century. Meanwhile, Burtt (1932), Koyré ([1939] 1978 and later writings), and Butterfield (1949) integrated the work of Copernicus, Kepler, Galileo, Descartes, and Newton into a single, extended event, dubbed the 'Scientific Revolution,' a label that came to name a historical period, which in turn became the central locus of historical research. Butterfield's *Origins of Modern Science* was especially accessible to a general audience. He contended that the Scientific Revolution was the emergence of modern science itself, not simply a revolution within science; and, moreover, that it was the most significant event since the rise of Christianity. The new discipline of history of science was therefore fundamentally important. (See Lindberg and Westman 1990 and Cohen 1994 for critical accounts of the historiography of the Scientific Revolution.) In addition, Butterfield presented the history of modern science as a *series* of conceptual revolutions, including the revolution in chemistry (with Lavoisier as the central figure) and the Darwinian revolution. Historians also began to speak liberally of technological "revolutions," such as the agricultural revolution, the print revolution, and, of course, the Industrial Revolution.

Since 1950, with the maturation of history of science as a discipline, and especially since the publication of Thomas Kuhn's (1962) *The Structure of Scientific Revolutions*, talk of scientific revolutions has proliferated (see Kuhn, Thomas). Notoriously, Kuhn incorporated some aspects of political revolutions in his account of scientific revolutions, and so the interplay continues. Historians and historical philosophers soon discovered other revolutions, for example, the revolution in wave optics (Young and Fresnel), the ongoing revolution in heat theory (Carnot, Kelvin, Clausius, Maxwell, Boltzmann), the revolution in electromagnetic theory (Faraday, Maxwell), the nineteenth-century revolution in probabilistic and statistical techniques, the Second Scientific Revolution around the turn of the nineteenth century, and so on. Meanwhile, revolutions in geology (plate tectonics) and molecular biology were taking place. Social scientists sought their own revolutions, leading some to announce a Chomskyan revolution in psycholinguistics (see Chomsky, Noam).

Attitudes toward both scientific and political revolutions have changed historically. The Restoration was controversial, while the Glorious Revolution was generally welcomed. Both the French and Russian upheavals gave revolutions a bad reputation as excessively violent rejections of old forms of life and culture. This negative association carried over to talk of scientific revolution in the ensuing decades. Old-guard scientists sometimes sounded rather like Edmund Burke in his conservative reaction to the French Revolution. The relativity and quantum revolutions were especially shocking epistemologically, for they did not replace prescientific speculation. On the contrary, they overturned classical mechanics, the very paradigm of scientific knowledge. They thereby destroyed the conception of science as foundational, as yielding virtually certain conclusions, and as the one truly cumulative institution there is. Today, however, revolution enjoys a generally positive image associated with liberation and opportunity for dramatic improvement. At least in the West, where political stability is unquestioned, this positive attitude is reflected in popular slogans and advertising, such as "Challenge authority" and "We broke all the rules and created new rules."

Unlike 'rebellion,' 'revolution' is an achievement term rather than a process term. It connotes a community's success in freeing itself from an established order and instituting a new order. Thus 'revolution' implies significant innovation. This application of the term is modern, for there can be no genuine revolutions as long as a divinely established order makes radical, humanly instigated change impossible and as long as significant innovation of any kind seems beyond human capacity. As a rejection of an established constitution, revolution also implies a normative struggle. Revolutionaries typically appeal to reason and justice, arguing that the old order is arbitrary, irrational, and inefficient and that it gratuitously thwarts human freedom and blocks progress. Thus it is not surprising that the modern idea of revolution emerged during the Enlightenment. In simplest terms, Enlightenment thinkers set reason and autonomy in opposition to history and tradition. However, this Enlightenment understanding is challenged by later revolutions that undermine previous, clearly modern scientific and political paradigms themselves.

Hegel attempted to overcome the dualism of reason and blind historical forces by positing transcendental reason or the Absolute as an agency that subtly harnesses myriad contingent events in such a way that a rational pattern emerges. Hegel and Marx developed conflict theories of profound social change, where tensions arise from the old structure rather than from external forces. Modern and postmodern thinkers generally accept a dynamical conception of human history and of the history of particular fields. One mark of modernity is the idea of essential social change, especially deliberately instituted progressive change. One mark of postmodernity is the denial that deep historical change is indisputably progressive and rational and that it fits a single, coherent narrative. As indicated, historical change becomes a threat to the claim that the sciences cumulatively reveal the timeless and culturally neutral truth about the universe.

The positivist-empiricist model of scientific change is ahistorical and only minimally dynamic, for it conceives scientific change as cumulative empirical and theoretical change within a stable but expanding conceptual framework. Conflict between established results is to be avoided at all costs. By contrast, evolutionary models postulate continuous but potentially transformative change over time. In them the conflict tends to be widely distributed and less intense than in revolutionary models, which work best when there exists a central core of ideas and practices to overthrow. Thus postmodern critics' decentralization of community power structures problematizes revolutions of the classical kind, yet the postmodern sympathy for disunity also inspires talk of discontinuity or rupture, that is, revolution.

The following sections survey how historians and philosophers of science have dealt with the problem of understanding deep scientific change.

Revolution as a Unit for History of Science

There are several additional reasons why revolution has become a popular topic, some of which raise the question as to what extent revolutions are artifacts or illusions of professional historical and philosophical research. These reasons include the following:

1. The amount of material to be gathered, analyzed, and cast into narrative form is potentially overwhelming. Thus some form of cognitive economy in the form of problematic and thematic organization is necessary to render this material humanly intelligible, to both historians and their readers. The idea of revolution has been a fruitful organizing principle for history and philosophy of science as well as for social-political history—perhaps the more so as recent history of science has turned away from internalist to social history. There are historiographical and epistemological reasons for this shift, including the difficulty of treating large quantities of highly technical material. On the other hand, this very complexity can challenge the identification of discrete revolutions.
2. The idea of revolution is well suited to the narrative forms that historians prefer. Story is one of the primary ways in which complex meshes of human events are made both interesting and intelligible. Revolutions are event-like or process-like and thus fall naturally into Aristotelian narrative form as having a beginning, middle, and end. Moreover, with scientific revolutions the story can be made dramatic, even heroic. Possible story lines full of human interest easily suggest themselves: the fall of the mighty; the triumph of the underdog, ridiculed outsider, or creative genius; achievement of the impossible; the discovery of a vast new territory; and so on. So depicted, science is no longer a boring, grinding affair of piling up experimental facts in an emotionless manner.
3. As a discipline, history of science came to intellectual and professional maturity under the guidance of such leaders as Koyré, who turned the history of science into an agonistic field of big intellectual ideas (e.g., Platonic vs. Aristotelian), that is, a story of clashing intellectual positions, in which human beings are merely the carriers or the agents. As Burtt had before him, Koyré combated the standard, positivist-empiricist view that it was a turn to empiricism that produced the Scientific Revolution for the revolution involved little new empirical information (they said) and was primarily the reorganization of already available material around a new set of philosophical and metaphysical ideas. The Scientific Revolution was an intellectual revolution. To a large degree, Thomas Kuhn would extend this analysis to revolutions in general, as Butterfield also had suggested.
4. Kuhn's *The Copernican Revolution* (1957) and *The Structure of Scientific Revolutions* (1962) further challenged conservative conceptions of science, illustrated the organizing and narrative power of revolution by turning the history of science into a drama of profound and unexpected conceptual change that was fascinating to a wide audience, and had immense impact on several fields of thought and action. The new history of science together with the philosophies of science described in the next section created a battle of the "big systems" and ultimately, with the advent of the new sociology of science, the "science wars" (see Social Constructionism).
5. Mature history of science features sensitivity to cultural context, as against the old "whiggish" histories that arranged all of the history of science within the same, cumulatively growing conceptual and cultural framework. Professional historians taking a biographical approach place a premium on identifying cultural differences and on capturing the subjective problem contexts of individual scientists such as Galileo, Newton, Faraday, Darwin, and Einstein. It is as if the discipline of history, like some high-tech imaging technologies, contains built-in contrast enhancers that magnify sensitivity to conceptual change and thereby turn every subtle change into a major explanatory problem. Such sensitivity can heighten the appearance of revolutionary thinking.
6. The institutionalization of history and philosophy of science in graduate schools further magnified this sensitivity, for it provided a way in which young historians and graduate students could emphasize the importance and originality of their historical subjects. Again, carried to an extreme, the smallest variation can be made to look revolutionary.
7. Yet, revolutions can also be constructed in historical writing by going to the opposite extreme of telescoping previous developments so that they appear to be more episodic and less evolutionary than they in fact were. Such revolutions can be whiggish illusions insofar as they are construed as sudden breaks from tradition. Butterfield (1931) had already warned that general history inevitably produces whig fallacies.

8. Among the factors that incline historians and philosophers toward a revolutionary conception of scientific change are intellectual commitment to stage theories of psychological, social, or scientific development; rigid, rule-based accounts of human cognition and/or social organization (e.g., various forms of structuralism); logical empiricist assumptions about the logico-linguistic structure of science; and romantic accounts of creativity. All such approaches render evolutionary accounts of stage transitions implausible. In addition to some logical empiricist assumptions, Kuhn was influenced by the developmental psychology of Jean Piaget, a stage theorist, while others were influenced by Freud's account of development, another stage theory. Many historians have been influenced by French historico-philosophical thought, e.g., that of Gaston Bachelard and Michel Foucault, with their emphasis on long-term epistemological formations and their ruptures.

On the other hand, mature historians rightly reject "deus ex machina" explanations of profound change such as appeal to genius or divine inspiration. Instead, they seek to understand major developments as the culmination of many previous events in their cultural and institutional contexts. Thus, good historians no longer see Galileo's or Newton's or Einstein's work as dropping miraculously from the sky. They view it as sustained exercises in problem solving, based on the influence of many precursors whose work they set out to document. And they see the propagation and acceptance of that work as the result of a broad range of techniques of persuasion, including funding enticements. Their overall strategy is to break a major change (e.g., a revolutionary "discovery") into numerous small changes. The same cultural sensitivity mentioned above can also lend itself to a "continuity model" of scientific development, as opposed to the revolutionary or "discontinuity model," again by breaking down the conceptual changes into smaller and smaller units. Pierre Duhem, an early exemplar of this strategy, had notoriously contended that much of the Scientific Revolution actually occurred in the fourteenth century.

Victorious scientists rewriting the history of their discipline also face a tension. On the one hand, they want to emphasize the revolutionary character of the work with which they have been heroically associated. On the other hand, in order to maintain their scientific identity, they want to make that work appear to be the natural or rational culmination of the previous history of their discipline.

Cohen (1985, Ch. 2) describes four stages of revolution:

- "the intellectual revolution" or "the revolution-in-itself," in which a radical problem-solving success leads investigators to formulate a research program that will spell out the implications;
- commitment to the new research program;
- communication to colleagues or the "revolution on paper"; and
- conversion of the sometimes resistant field to the new program and its agenda.

Cohen (chap. 3) employs four kinds of historical tests for identifying revolutions. First, contemporary witnesses, both scientists and nonscientists, testify to the occurrence of a revolution. A revolution must be recognized as such at the time. (For instance, Funkenstein (1988) notes the radicalism of authors such as Descartes, Hobbes, and Newton in claiming that metaphysical reality is totally different from what anyone before had imagined.) Second, later documentary histories of the science claim that a revolution has occurred. Third, this attribution is confirmed by the judgment of competent historians of science and philosophy. And, fourth, it is confirmed by the general opinion of scientists working in the field today. Cohen, like Kuhn, calls attention to the important fact that practitioners facing a revolutionary choice give special weight to future promise, for the established approach almost always possesses greater scope and stronger confirmation than the young challenger.

Cohen's inclusion of today's perspective excludes announced revolutions that did not pan out. In history one is usually wiser after the event. And his emphasis on historically contemporary accounts avoids the danger of later scholars retrospectively inventing revolutions where there were none. Does Cohen give contemporary accounts too much weight? An ironic consequence of his view is that the Copernican revolution was not really a revolution, since it was largely ignored for decades. And, as noted above, the modern concept of scientific revolution was not yet available in earlier times.

What stretch of historical events should be included in a revolution, and who is to decide? Rarely, at the frontier of research, can the principal investigators formulate a radical research program that stands the test of time. It typically takes years, even decades, to distill out the core meaning of dramatic new developments, as the quantum revolution illustrates. Cohen's insistence that revolutions be highly innovative stands in tension with his claim that they are fully recognized by their makers and, to some degree, preprogrammed. If revolution is at all like biological speciation, then 'revolution' is a retrospective category. Cohen himself calls attention to the mechanisms by which scientific results and techniques can be gradually transformed, almost beyond historical recognition. Finally, how should one classify results that are not radical breaks from tradition, yet in the long run turn out to be remarkably productive? Cohen recommends a rather sharp distinction between revolutions themselves, as definite historical episodes, and their consequences.

Sociologists and cultural historians challenge standard historical and philosophical accounts of revolutions as signal intellectual achievements in much the same way that they challenge standard accounts of scientific discovery, given that revolutions are supposedly founded upon major discoveries. Like discoveries, revolutions are notoriously difficult to identify and delimit in their actual historical contexts. And like 'discovery,' 'revolution' is not simply a logical, or epistemological, or even a descriptive-historical term. Rather, it is an honorific label, retrospectively conferred by a scientific community (or its historiographers and philosophers) in order to celebrate and canonize designated authors and events as constituting the present state of the discipline and its authorized practices. As such, these social attributions serve important community and cultural functions, but they belong more to the genre of founder myths than to historiography or epistemology. No event is revolutionary by nature. That label is the product of a complex process of social negotiation (Schaffer 1994).

Revolution as a Unit for Philosophy of Science

There are also professional philosophical reasons why revolution is attractive as a unit of analysis for the philosophy of science. Paradoxically, these spring from both ahistorically and historically oriented philosophies. Kant was ahistorical. However, his introduction of the idea that the world of everyday and scientific experience is a product of a rational and specifically human conceptual scheme (the forms of intuition and the categories) immediately suggested to the historically oriented Hegel the possibility of alternative conceptual schemes, including the possibility that ancient and medieval peoples actually exemplified such alternative forms of life. In that case, human history organizes itself into distinct, stable epochs with less stable, perhaps revolutionary, transitions between them. Such an idea proved very suggestive to idealist philosophers, social theorists, and anthropologists, and it became one source of the nineteenth-century discovery of deep history and of culture. A century and a half later it would strongly influence Thomas Kuhn, who described himself as a "historicized Kantian" (Kuhn 2000). And it would become a leading heuristic of the new, anti-whiggish history of science to focus on cultural and intellectual differences rather than similarities. Revolutionary ruptures between intellectual and social formations already characterized a line of French thinkers, including Bachelard, Canguilem, and Foucault (Gutting 2003).

Prominent among ahistorical philosophers were the logical empiricists, who, between about 1930 and 1970, attempted to reconstruct scientific method in terms of fixed, formal logical systems, which permit no real conceptual growth. Thus scientific change that requires a new logical framework must appear to be revolutionary, absent any mechanism for how one logical system can evolve smoothly into another one quite different from it.

Most logical empiricists minimized the problem of change by assuming that science, at least as rationally reconstructed, develops within more or less the same conceptual framework and that theory change can be handled in terms of the logic of intertheoretic reduction (see below). Members of the operationalist movement held much the same position by virtue of the requirement that each theoretical term be fixed in advance by a permanent operational definition (see Bridgman, Percy). The physicist and operationalist Bridgman (1927) explicitly denied the need for revolutions in rightly conducted science:

> We should now make it our business to understand so thoroughly the character of our permanent mental relations to nature that another change in our attitude, such as that due to Einstein, shall be forever impossible. It was perhaps excusable that a revolution in mental attitude should occur once, because after all physics is a young science, and physicists have been very busy, but it would certainly be a reproach if such a revolution should ever prove necessary again. (2)

In the case of relativity, thought Bridgman, Newtonian physicists should have defined the concept of simultaneity more carefully in the first place and erected a corresponding physical theory that would have avoided the later need for an Einsteinean revolution—as if scientists could be so prescient of later developments and research programs could handle full complexity from the start.

By contrast, Kuhn (1962) claimed that scientific revolutions are necessary if science is to progress beyond its present conceptual and practical framework, and therefore inevitable, assuming that it will so progress. By its nature, said Kuhn, mature scientific practice eventually produces major conceptual and practical change, and so it must break out of the old framework of theory and practice. Whereas most writers, past and present, have considered the occurrence of political and scientific revolutions to be rare and highly contingent, Hegel, Marx, and, later, Kuhn regarded them (although not their precise content) as qualitatively predictable because inevitable. The views of these individuals were of course very different in other ways. Notably, Hegel's transitions and Marx's revolutions lead inevitably to a final goal, whereas Kuhnian revolutions are nonteleological and will continue without end as long as science continues to flourish.

As noted, the later logical empiricists attempted to handle the problem of major theoretical change by means of reduction of theories (e.g., Nagel 1961) (see Reductionism). Applied to a theory and its successor, reductionism is the conservative idea that the new theory is logically more general than its predecessor and hence explains it by incorporating it as a special case. Thus Maxwell showed that electromagnetic theory entails physical optics, which in turn entails ray optics as an idealization. Somewhat similarly, special relativity theory yields Newton's laws in the limit of low velocities. Thus the conceptual framework and the empirical content of science expand cumulatively rather than being undermined and replaced (cf. Sarkar 1992).

Kuhn (1962) and Paul Feyerabend (1962) opened what became the battle of the big systems by raising two objections against this account (see Feyerabend, Paul; Kuhn, Thomas). First, major theoretical change brings meaning change. Einsteinian masses are not the same as Newtonian masses, since the former but not the latter are velocity dependent, a major conceptual difference. Hence, to claim that Newton's laws are derivable, even approximately, from Einstein's, commits the fallacy of equivocation. Second, the succession of one major theory by another always involves "Kuhn loss," that is, phenomena that were explained by the old theory but cannot be handled by the new one. For example, once early-nineteenth-century scientists accepted Dalton's view that the atmosphere is a physical mixture rather than a chemical compound, they could no longer explain why the heavier gases did not settle out (Kuhn 1962, §X).

Other philosophers, notably Popper, responded to the twentieth-century revolutions in physics in a far more positive way than did Bridgman and the logical empiricists, while denying the status of genuine scientific revolutions to Freud's and Marx's work (see Popper, Karl Raimund). Inspired by Einstein, Newton, and a few other great men of science, Popper rejected the so-called Baconian method of induction from the facts. In its place he promoted a romantic picture of creative revolutionaries who put forward bold but empirically testable theoretical conjectures in order to solve deep intellectual problems. Popper became so enamored of his "critical approach" and its motivating idea ("we learn from our mistakes") that he and his followers urged "revolution in perpetuity." For Popper, revolutionary new ideas undermine and replace their predecessors, but they must incorporate everything that seems correct in the old viewpoint, a conservative requirement that tames his revolutionary impulse.

Feyerabend took Popper's revolutionary ideas to the limit. The early Feyerabend (1962) recognized that some empirical criticism requires a deep theoretical background and contended that scientists should develop multiple, competing theory systems in order to maximize criticism of each from the standpoint of the others. This was his "proliferation thesis." Yet Feyerabend also spoke of alternative theoretical viewpoints as being "incommensurable," a term introduced simultaneously by Kuhn (see Incommensurability). The later Feyerabend (1975) became still more critical than Popper of standard conceptions of scientific method and argued on both historical and philosophical grounds that there is no permanent logic of science or theory of rationality. Rather, whole new conceptions of method and rationality have been brought into existence through the rhetorical strategies and social manipulations of such men as Galileo. Feyerabend branded this revolutionary position "methodological anarchism."

Meanwhile, Feyerabend's friendly rival, Lakatos (1970), argued for a permanent "methodology of scientific research programs." Whereas Popper considered the formulation and testing of theories more or less in isolation, Lakatos's methodology explicitly recognized the importance of long-term traditions within the history of science in the form of competing research programs that fight battles of attrition. Each research program strives to produce more novel predictions and confirmations than its rivals. In this view there are no scientific revolutions, properly so called, wholly within a single research program. However, the splitting of an old program or the foundation of a major new program could be considered revolutionary (see Lakatos, Imre; Research Programs).

Thomas Kuhn on the Structure of Scientific Revolutions

Kuhn rejected Popper's idea of revolution in perpetuity as virtually self-contradictory, for to speak of revolution makes sense only as a reaction against a securely established, stable regime. Scientific revolution can be understood only by contrast with normal science. (Bachelard [(1934) 1984] had already allowed for "continued revolutions.") Kuhn also criticized Popper's critical approach on the ground that its thoroughgoing application would destroy science as currently practiced and turn it into something more like philosophy, complete with the latter's premium on critiquing everything, including the fundamental principles that define the field. Kuhn rejected Feyerabend's call for proliferation and, by implication, Lakatos's competing research programs (as ways of achieving a kind of perpetual revolutionary antagonism) on the economical ground that there are insufficient intellectual and financial resources to support such extravagance (Worrall 2003). After all, it is difficult enough to produce a single major framework (Newtonian, Einsteinian, Darwinian), with its associated community and sets of practices and institutions.

Besides, for Kuhn mature science distinguishes itself from other human endeavors precisely because it is monolithic: A single paradigm defines a normal scientific period, and work under a paradigm is conservative and tradition bound, convergent rather than divergent. Eventually work at this level of esoteric detail will produce anomalies that even the best practitioners fail to resolve. Long-term failures gradually weaken commitment to the paradigm to the point of crisis, allowing alternative approaches to spawn. If the relevant gatekeepers of science can be persuaded that one of these alternatives handles the major anomalies in a more promising way than the received paradigm, then a revolution will likely ensue.

Kuhn (1962) notoriously emphasized that the transition to the new paradigm cannot be wholly 'rational' in the received sense of the word, for, again, rationality and standards of good science in their domain-specific forms are relative to a paradigm, and the competing paradigms are incommensurable. The early Kuhn maintained that the victory of a new paradigm requires political argument, power plays, and rhetoric as well as something like religious conversion.

Though infrequent, Kuhn's revolutions are not merely isolated historical contingencies, for he regularized the idea of revolution as paradigm change. Many if not all mature sciences (or at least the physical and biological sciences) were born in revolt against the folk science that preceded them, and all inevitably generated crises, some of which required a paradigm change for their resolution. It is not clear whether revolutions away from folk science and revolutions within an already mature science are of the same type.

For Kuhn a large science such as physics, chemistry, or biology consists of a hierarchy of paradigms—an overarching paradigm for the field as a whole with smaller paradigms for specialties and subspecialties. Scientific revolutions can be relatively small and local, since a community of specialists may include only a couple dozen practitioners, scattered worldwide. A revolution within such a subspecialty may appear to be an incremental change to scientists working in other areas, and may be completely invisible to the general public (Hoyningen-Huene 1993, chap. 6). Thus Kuhn greatly proliferated scientific revolutions while relativizing them to specialist communities.

For Kuhn, normal scientific work does not seek profound innovation, yet unexpected major discoveries occasionally occur. These are revolutionary, since they fall outside the bounds of normal science; but they alone do not constitute full-fledged revolutions (Hoyningen-Huene 1993, chap. 6). Although Kuhn himself is sometimes guilty of telescoping historical developments in such a way that they appear more revolutionary than they were, it was Kuhn who emphasized that both discoveries and revolutions are structured events (Kuhn 1962 and 1978, Ch. 7).

Kuhn claimed that major scientific revolutions are radical changes in worldview and the associated technical practices, and that competing paradigms are incommensurable (Kuhn 1962, §X). Scientists working under competing paradigms literally live and practice their trades in different worlds. These strong claims, a product of Kuhn's "historicized Kantianism," drew a storm of protest from philosophers (e.g., Scheffler 1967; Shapere 1984, Chs. 3 and 4; Sankey 1994). Kuhn spent much of the rest of his career attempting to reformulate, clarify, and defend versions of these claims (Kuhn 2000). One of the most important critics was Stephen Toulmin, who, prior to Kuhn, had discussed revolutionary changes in framework principles that he termed "ideals of natural order" (Toulmin 1961). Toulmin (1972) rejected Kuhn's account of scientific development as too saltatory and countered with an evolutionary model of the growth of science. Kuhnian normal science, he complained, retains the logical empiricist conception of scientific rationality as logicality within a fixed system rather than as a more biological adaptation to changing circumstances. It was precisely this unjustified rigidity of normal science that required revolutionary breaks to a new, equally rigid framework. On the other hand, Kuhn himself frequently drew on evolutionary metaphors.

In a book left unfinished at his death, Kuhn abandoned history of science as a basis for his views, relying instead on philosophy of language and his understanding of human cognition. Kuhn now claimed that revolutionary incommensurability involves a major modification of the lexical structure of scientific knowledge, the meaning of key terms and their relationships. Revolutions alter the synthetic a priori judgments, the deep, quasi-Kantian constitutive principles of normal science (Kuhn 2000). These cognitive orientations are synthetic because they carry content about the world, but they are a priori because, during the period of normal science, they are not subject to empirical test or refutation. Revolutions involve major changes in the network of similarity relations that produce one Kantian "world" or another. Incommensurability entails the impossibility of accurate, technical translation but does not completely foreclose interpretation, mutual intelligibility, and rational comparability.

In addition to the aforementioned French thinkers, revolutionary discontinuities in the history of science were introduced by Fleck ([1935] 1979), who spoke of community thought-styles and thought-collectives. Fleck's work apparently influenced Kuhn, as did Polanyi's (1958) treatment of scientific practices and the associated "tacit knowledge." In turn, many analysts have been influenced by Kuhn's treatment of scientific revolutions. Margolis (1993) reduces all cognition to pattern matching and the associated "habits of mind." Even logical reasoning is pattern matching. There is no special faculty of reason or logic. Margolis's principal thesis is that sometimes what separates two incommensurable Kuhnian paradigms and produces mutual incomprehension is not a seemingly unbridgeable logical or conceptual gap but rather a cognitive barrier, a deeply ingrained habit of mind that prevents an available argument or conceptual transformation from becoming salient. He thus attempts to explain whiggish perplexities of the type: Why did it take person or community *X* so long to recognize that *Y*? For example, why did it take so long to make the transition from the Ptolemaic to the Copernican paradigm, given that all the key steps of the argument had been available for centuries? Margolis contends that the blocking habit or intuition may be peripheral to the paradigm in question. In Margolis's account it was the availability of new maps of the world that broke the nested spheres model of the universe and finally enabled Copernicus to appreciate the power of the transforming argument. It then took the few followers of Copernicus another half century to convince others of the argument's merits. Not even known logical arguments are automatically persuasive. For Margolis, a revolutionary development need not be subversive; it may simply open up vast new territory. His chief example is the probabilistic revolution of the nineteenth century, a topic thoroughly explored by Hacking (1983 and later works).

Thagard (1992) distinguishes several types and degrees of conceptual change in science, relative to the organization of scientists' cognitive representations of the taxonomy of their domain, including part/whole and hierarchical relations. The two most significant kinds of changes are branch jumping and tree switching, both of which occur in scientific revolutions. Branch jumping moves a concept from one branch of a hierarchical tree of conceptual categories to another, as when Brownian motion was reclassified from a biological to a physical phenomenon. More revolutionary examples are Copernicus's reclassifying the sun as a star and the Earth as a planet and Darwin's reclassifying human beings as animals. Tree switching involves changing the organizing principle of the taxonomic tree, a more radical change, and is illustrated by Lavoisier's replacement of a phlogiston account of combustion by his new account in terms of oxygen. (For an account of such conceptual change based on Kuhn's later work, see Barker, Chen, and Andersen 2003.)

Business analyst Christensen (1997) does not mention Kuhn, but his distinction of two types of technologies and of technological innovation in the business world indirectly challenges Kuhn's claim that the mature sciences exhibit only one type of revolutionary scenario, for the distinction may apply to science as well. Christensen notes that well-established industry leaders who employ good business practices may nonetheless be subverted by smaller companies, an observation that is suggestive for both the history of science and the history of technology. He distinguishes sustaining technologies from disruptive technologies. Leading companies normally pursue sustaining technologies, focusing their research and development on the improvement of established product lines. Sustaining technologies should not be considered the technological counterpart of Kuhnian normal science, however, for good companies deliberately seek innovation and the innovations can be revolutionary (consider the history of large aviation companies, for example). On the other hand, disruptive technologies need not involve breakthrough technology, only the clever assembly of "off the shelf" technology (e.g., the Sony Walkman and the personal computer). They are promoted by small, fringe companies and their scattered group of followers. Such technologies do not initially appear to threaten the large companies and their product lines. Applying these ideas to the history of science discloses many examples of once-peripheral practices that eventually overtake dominant ones without there necessarily being a direct clash, certainly not a logical one.

The usual association of scientific revolutions with major changes in the world picture is theory centered. While Kuhn (1962 §X) mentions this conception of revolutions, he also stressed the practices of the various scientific communities (Rouse 2003). Much of what scientists and other experts know is embodied in their subarticulate practices and in the "acquired similarity relations" that have shaped their cognitive faculties. In this view, a revolution alters the basic form of scientific community life and, accordingly, the basic form of human cognition. In the last quarter-century, there has been a turn away from theory-centered accounts to those emphasizing social practices, especially experimental ones (see Experiment). Here the emergence of a new technique, perhaps a practice built around a newly invented piece of equipment, can be called revolutionary in opening up new areas of inquiry, whether or not it refutes an established practice. (Note that Kuhnian paradigm replacements are not straightforward refutations either.) It is useful to speak of some developments as *revolutionary* without reifying them as *revolutions*.

Types and Characteristics of Revolutions

The following is a list of characteristics often attributed to scientific revolutions. The items overlap in various ways, and each is subject to qualification or outright challenge:

1. Revolutions achieve scientific change on a large scale.
2. Revolutions originate as major discoveries.
3. Revolutions are episodic and event-like, rather than states, standing conditions, or long-term tendencies.
4. Revolutions are relatively rapid—as in 'the Scientific Revolution' being used to designate an entire historical period and a corresponding field of historical study rather than a specific series of episodes (Hall 1954). Single revolutions do not last that long. Should not a revolution in the event sense occur on a human scale, within a human lifetime or less? Scientific and technological revolutions tend to take longer than political revolutions to achieve finality or even recognition, and it is often more difficult to decide when they are 'over.'
5. "Revolution" is normally a retrospective, success/achievement term, although 'revolutionary' is not, especially when applied to persons rather than events. There cannot be a failed revolution, but there can be failed revolutionaries and failed revolts. Sociologists and social historians conclude that being a revolution is not intrinsic to a change or process: It must be *socially* recognized as such. Social attribution is crucial (Schaffer 1994).
6. Each revolution is unique. Oliver Wendell Holmes, Jr. ([1861] 2001) once remarked, "Revolutions do not follow precedents nor furnish them," in which case historical and philosophical understanding of revolutions becomes still more difficult. Barker et al. (2003) contend that revolutions can occur slowly and can have important characteristics in common with each other.
7. Revolutions are highly innovative. They are not direct applications or simple adaptations of precedents.
8. Revolutions create fundamentally new worldviews or practical forms of life. Recent science writers have sometimes used the metaphor of creation, e.g., Crease and Mann (1996) and Judson (1979).
9. Revolutions are convulsive and subversive and, accordingly, meet strong resistance. They involve revolts against an established orthodoxy, violations of received dogmas and procedures. This view rejects the "free expansion" model of revolution as a rapid and unhindered advance into new territory. 'Revolution' is increasingly used in this latter sense today.

10. Revolutions are divergent and radical, while evolution is convergent and conservative. Revolutions constitute breaks and discontinuities, and are hence incompatible with purely evolutionary accounts of scientific development. (Yet what process is more creative in the long run than biological evolution?)
11. Revolutions are breakouts from an old conceptual framework. This is an overly intellectual, theory-centered view that gives insufficient emphasis to revolutionary practices and technologies, including organizational technologies.
12. Revolutions involve replacements rather than reductions of the previous scheme. This was the claim of Feyerabend and Kuhn against the logical empiricists and Popperians. Moreover, not all replacements are refutations.
13. Revolutions are paradigm changes or "paradigm shifts," in Kuhn's sense.
14. Revolutions require incommensurability between the new and old programs. They create considerable cognitive dissonance, mutual incomprehension, and discomfort. This is the feature that Kuhn and Feyerabend always emphasized.
15. Revolutions involve reconceptualizing already familiar materials and practices (especially the constitutive principles and practices that define the paradigm) rather than introducing substantially new empirical content. Koyré and Kuhn tended to favor this characterization.
16. Revolutions are intrinsically social, not individual.
17. Revolutions involve a reorganization of scientists' cognitive representations, habits of mind, or similarity metrics, e.g., through the adoption of transformed taxonomies (Kuhn 1962 and 2000).
18. Revolutions are not mere changes in the content claims of a scientific field, however radical, for revolutions also transform goals, standards, and methods.
19. Revolutions alter the administrative and economic landscape of the field—for instance, by reorganizing the governing institutions and organs of science. Hacking (1983) contends that major scientific revolutions couple major changes in content with institutional changes. The Scientific Revolution was accompanied by the emergence of many new scientific societies and their publications. The nineteenth-century probabilistic revolution was accompanied by the appearance of government social statistics bureaus.
20. Revolutions involve the emergence of new fields or specialty areas or the splitting of a field into separate disciplines.
21. Fully genuine scientific revolutions produce a proliferation of new results across several scientific fields.
22. Fully genuine scientific revolutions change one's understanding of the universe and one's place in it and thereby challenge wider cultural values. Kuhn's technical concept of revolution denies that this feature is necessary.
23. Revolutions are nonlinearities in scientific development, including not only discontinuities but also exponential spurts, either temporal or logical (as when a relatively compact development has immensely fertile consequences). In this respect, revolutions are "chaotic" and unpredictable. People do not know in advance where a revolution will take them. This is a "consequentialist" interpretation of revolution and conflicts with the more "generative" views espoused by Cohen. The nonlinearity of revolutions is highlighted by Margolis's (1993) account of barriers, for what removes a cognitive barrier to revolution may be a development far-removed from the center of the scientific action.
24. Revolutions are progressive. (Kuhn and Feyerabend denied this in the sense of progress toward a predetermined truth.)

Given the multiplicity of criteria that may come into play, there need be no opposition between revolution and rapid evolution. The closer one looks at revolutionary episodes and the more carefully one explores the contextual sources of the "revolutionary" work (especially noting those that lie outside the field as it was previously constituted), the more numerous and relatively smaller the individual transitional steps become, however rapidly they may have progressed. Evolution is capable of producing revolutionary changes when plotted against a more course-grained timescale. One is tempted to label "revolutionary" any development that yields extremely fertile practices, whether or not these consequences were immediately apparent and whether or not they were subversive of an entrenched orthodoxy. Yet, less fertile but rapid developments also appear revolutionary. Again, it is partly a question of scale, of getting a large output from a small input.

The Wider Importance of the Debate about Scientific Revolution

The discussion of scientific revolutions and the attendant concepts of paradigm, conceptual scheme, and incommensurability have enjoyed wide influence within general philosophy and popular culture. Kuhn's distinction of normal from revolutionary science and his rejection of traditional conceptions of realism, rationality, and objectivity have been one resource for recent feminist movements and other postmodern cultural developments. Controversies over scientific revolutions also fuel the large debate over realism and social construction within philosophy and sociology of science—the so-called "science wars" (see Social Constructionism). Kuhn's historical Kantianism raises once again central issues concerning the nature of human cognition and one's epistemic relations to reality. For example, Rorty (1979 and 1991) makes substantial use of Kuhnian distinctions. MacIntyre (1980) brings them into the moral sphere. Scientific revolutions inform the work of Quine (1953) and Davidson (1984) on radical translation and radical interpretation. Davidson, for example, rejects the dualism of scheme and content as a "third dogma of empiricism."

There is no coherent, agreed-upon concept of scientific revolution. Accordingly, many science studies practitioners, including some philosophers of science, avoid the term as problematic; but others find it essential for accounts of the development of modern science.

THOMAS NICKLES

References

Bachelard, Gaston ([1934] 1984), *The New Scientific Spirit*. Boston: Beacon.

Barker, Peter, Xiang Chen, and Hanne Andersen (2003), "Kuhn on Concepts and Categorization," in T. Nickles (ed.), *Thomas Kuhn*. New York: Cambridge University Press, 212–245.

Bridgman, P. W. (1927), *The Logic of Modern Physics*. New York: Macmillan.

Burtt, E. A. (1932), *The Metaphysical Foundations of Modern Physical Science*. New York: Harcourt Brace. Revised edition, 1932.

——— (1931), *The Whig Interpretation of History*. London: Macmillan.

Butterfield, Herbert (1949), *The Origins of Modern Science, 1300–1800*. London: G. Bell.

Christensen, Clayton (1997), *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Cambridge, MA: Harvard Business School Press, 1997.

Cohen, H. F. (1994), *The Scientific Revolution: A Historiographical Inquiry*. Chicago: University of Chicago Press.

Cohen, I. B. (1985), *Revolution in Science*. Cambridge, MA: Harvard University Press.

Crease, Robert, and Charles Mann (1996), *The Second Creation: Makers of the Revolution in Twentieth-Century Physics*. New Brunswick, NJ: Rutgers University Press.

Davidson, Donald (1984),. "On the Very Idea of a Conceptual Scheme," in *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 183–198.

Feyerabend, Paul (1975), *Against Method*. London: New Left Books.

——— (1962), "Explanation, Reduction, and Empiricism," in Herbert Feigl and Grover Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press, 28–97.

Fleck, Ludwik ([1935] 1979), *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.

Funkenstein, Amos (1988), "Revolutionaries on Themselves," in William Shea (ed.), *Revolutions in Science: Their Meaning and Relevance*. Canton, MA: Science History Publications, 157–163.

Gutting, Gary (2003), "Thomas Kuhn and French Philosophy of Science," in T. Nickles (ed.), *Thomas Kuhn*. New York: Cambridge University Press, 45–64.

Hacking, Ian (1983), "Was There a Probabilistic Revolution, 1800–1930?" in Michael Heidelberger, Lorenz Krüger, and Rosemarie Rheinwald (eds.), *Probability Since 1800: Interdisciplinary Studies of Scientific Development*. Bielefeld, Germany: B. Kleine Verlag, 487–506.

Hacking, Ian (ed.) (1981), *Scientific Revolutions*. Oxford: Oxford University Press.

Hall, A. R. (1954) *The Scientific Revolution, 1500–1800*. London: Longmans Green.

Hatto, Arthur (1949), "'Revolution': An Enquiry into the Usefulness of an Historical Term," *Mind* 58: 495–517.

Holmes, Jr., Oliver Wendell ([1861] 2001), Letter to Felton, 24 July. Quoted by Louis Menand, *The Metaphysical Club*. New York: Farrar, Straus and Giroux, 33.

Hoyningen-Huene, Paul (1993), *Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science*. Chicago: University of Chicago Press.

Judson, Horace (1979), *The Eighth Day of Creation: The Makers of the Revolution in Biology*. New York: Simon and Schuster.

Koyré, Alexandre ([1939] 1978), *Galilean Studies*. Translated by John Mepham. Atlantic Highlands, NJ: Humanities Press. Originally published as *Etudes Galiléennes*. Paris: Hermann.

Kuhn, Thomas (1957), *The Copernican Revolution*. Cambridge, MA: Harvard University Press.

——— (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press. Second edition with postscript, 1970.

——— (1978), *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.

——— (2000), *The Road Since Structure: Philosophical Essays, 1970–1993*. Edited by James Conant and John Haugeland. Chicago: University of Chicago Press.

Lakatos, Imre (1970), "Falsification and the Methodology of Scientific Research Programmes," in Lakatos and Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press, 91–195.

Lindberg, David, and Robert Westman (eds.) (1990), *Reappraisals of the Scientific Revolution*. Cambridge: Cambridge University Press.

MacIntyre, Alasdair (1980), "Epistemological Crises, Dramatic Narrative, and the Philosophy of Science," in

G. Gutting (ed.), *Paradigms and Revolutions*. Notre Dame, IN: University of Notre Dame Press, 54–74.

Margolis, Howard (1993), *Paradigms and Barriers: How Habits of Mind Govern Scientific Beliefs*. Chicago: University of Chicago Press.

Nagel, Ernest (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace and World.

Polanyi, Michael (1958), *Personal Knowledge*. Chicago: University of Chicago Press.

Quine, Willard Van (1953), "Two Dogmas of Empiricism," in *From a Logical Point of View*. Cambridge: MA: Harvard University Press, 20–46.

Rorty, Richard (1979), *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.

——— (1991), *Objectivity, Relativism, and Truth: Philosophical Papers*, vol. 1. Cambridge: Cambridge University Press.

Rouse, Joseph (2003), "Kuhn's Philosophy of Scientific Practice," in T. Nickles (ed.), *Thomas Kuhn*. New York: Cambridge University Press, 101–121.

Sankey, Howard (1994), *The Incommensurability Thesis*. Aldershot, UK: Avebury.

Sarkar, Sahotra (1992), "Models of Reduction and Categories of Reductionism," *Synthese* 91: 167–194.

Schaffer, Simon (1994), "Making Up Discovery" in Margaret Boden (ed.), *Dimensions of Creativity*. Cambridge, MA: MIT Press, 13–51.

Scheffler, Israel (1967), *Science and Subjectivity*. Indianapolis: Bobbs-Merrill.

Shapere, Dudley (1984), *Reason and the Search for Knowledge*. Dordrecht, Holland: Reidel.

Thagard, Paul (1992), *Conceptual Revolutions*. Princeton, NJ: Princeton University Press.

Toulmin, Stephen (1961), *Foresight and Understanding*. New York: Harper & Row.

——— (1972), *Human Understanding*. Princeton, NJ: Princeton University Press.

Worrall, John (2003), "Normal Science and Dogmatism, Paradigms and Progress: Kuhn 'versus' Popper and Lakatos," in T. Nickles (ed.), *Thomas Kuhn*. New York: Cambridge University Press. 65–100.

See also **Bridgman, Percy; Feyerabend, Paul; Incommensurability; Induction, Problem of; Kuhn, Thomas; Lakatos, Imre; Logical Empiricism; Popper, Karl Raimund; Reductionism; Scientific Change; Social Constructionism**

SCIENTIFIC STYLE

The term 'scientific style' has been used in a number of different ways, but with a characteristic strand of interconnected concepts and connotations: historicity, worldview, world picture, *Zeitgeist*, artistic style, thought collective, thought style, episteme, conceptual schemes, paradigms, themata, pluralism, relativism, and unity and disunity of science.

In the footsteps of Kant's idealist philosophy, German culture made central the notions of worldview or world conception (*Weltanschauung*) and world picture (*Weltbild*), which emphasized an overall unified conception of meaningful life involving cognition, emotion, and volition. It had individual as well as cultural and social connotations. In particular, it stressed culture, history, and the expression of values. For instance, Windelband's (1891) *History of Philosophy* was devoted to the study of individual worldviews. Galileo is mentioned as the spearhead of the "scientific world view" based on the mathematization of nature. The priority of meaning and values in historical method raised the question of the objectivity of the worldviews and of the social sciences in general, and in *Theory of World Views*, Dilthey ([1911] 1931) extended the use of the notion of worldview beyond philosophy, to religion and art. Since the early 1880s his approach to the historical sciences included the study of cultural phenomena in terms of typology, development, and environment.

The pursuit of a unified scientific worldview or world picture became central to German-speaking scientists from Mach and Planck (protagonists of a heated debate about the true physical world picture at the turn of the twentieth century) to Einstein. In more philosophical quarters, however, the emphasis was on plurality and history. In the 1920s Mannheim ([1929] 1936) introduced the term 'thought style' (*Denkstil*). He emphasized the historical-social nature of worldviews, associated with particular groups, and how this nature determined their subject mater, ways of setting problems, and conceptual apparatus. Thought styles, like worldviews, are radically perspectival and historical. According to Mannheim, their irreducible value-ladenness forces on the social sciences a new kind of objectivity based not on their elimination but on their critical awareness and control.

In the mid-1930s, in response to a sense of European intellectual crisis, Husserl, Heidegger, and Cassirer wrote about Galileo, revisiting what they considered the birth of modern civilization, linked to the rise of science. For Heidegger, Galileo inaugurated the exploration of the world predetermined by the mathematical. Husserl had read Windelband and referred to Galileo's scientific worldview already in the 1910s. For Husserl ([1936] 1970) the scientific worldview, the mathematization of nature, revealed the ideal essences of perceived things, but it was not the whole truth of the life-world and had be rooted in the subjective. For Cassirer (1937), Galileo's "universal style" constituted a system of symbolic forms—like those of art and religion—and reflected the ideal that the human mind was adequate to all of nature.

Also in the mid-1930s, Fleck ([1935] 1979) published the first monograph devoted to the examination of scientific *styles*, in this case applied to medicine. For Fleck it was a means of historicizing science itself, of discussing scientific change, and of doing so from a social perspective, namely, that of the "thought collective" (*Denkkollecktiv*). This approach stood in sharp contrast to Popper's (1935) unifying and demarcating emphasis on the autonomous logic of scientific method expressed in his *Logic of Scientific Discovery*. Fleck explicitly contrasted his ideas to Carnap's early notion of unified science based on ultimate elements of direct experience. A style is characterized by common features in the problems of interest to a thought collective, by the judgments that the thought collective considers evident, and by the methods that it applies as a means of cognition. It constrains individuals by determining what cannot be thought in any other way. (Fleck drew an analogy to perceptual *Gestalts*.) In this way facts can be both socially determined and factual or objective.

Toulmin (1961) argues that the style of theory and interpretation determines the terms in which one formulates the questions that are asked with provisional theories and in which one is answered with experiments. Holton (1973) speaks, by analogy with art, of scientific styles and links the stylistic commitments of a scientist with themata, preconceptions that drive theoretical and experimental work. Like many stylistic categories in art history, many themata come in opposite pairs, and, like artistic commitments, they may have characteristically aesthetic dimensions such as symmetry or simplicity. One such themata is a unified world picture or worldview, which Holton traces in Einstein's work.

Wisan (1981) has seen in Galileo's works the emergence of a new scientific style. She has developed the point by introducing an internalist analysis of the concept of scientific style by explicit analogy with the role and understanding of style in art history, especially in Wölfflin's ([1915] 1950) formalist approach. This allows her to make sense of scientific change by sorting out both innovations and continuities. Her notion of scientific style is characterized in terms of structure, content, techniques, and expressive quality. Thus, for instance, she distinguishes between Aristotle's and Galileo's treatments of motion, respectively: discursive and classificatory versus geometric and axiomatic; definitions of fundamental concepts versus unified treatment of local motions; logical analysis and explication of concepts versus geometrical derivations of mathematical consequences; substance and essence versus mathematico-physical magnitudes.

The most fully developed treatment and application of the notion of scientific style came subsequently in Crombie's (1993) *Styles of Scientific Thinking in the European Tradition*. Crombie introduces an externalistic notion of scientific style that provides a comparative intellectual anthropology of science and requires an interdisciplinary approach. This view notably contrasts with an emphasis on the logic of a unified scientific method. Crombie's notion of style is characterized by the following kinds of commitments: conceptions of nature within the general scheme of existence and its knowability by humans; conceptions of science and of the organization of scientific inquiry, argument, and explanation; a vision of what is desirable and possible in view of evaluations of nature, purpose, and circumstances of human life; and a commitment to the physical and biological environment in which humans find themselves (even if it is to change it). Crombie suggests six main styles:

(i) the Greeks' axiomatic method—the search for postulates and principles from which phenomena could be demonstrated (e.g., Euclid);
(ii) the medieval logic of experiment—the search for control and explanation of regularities (e.g., Bacon);
(iii) hypothetical argument and models—the search for analogies (e.g., Harvey);
(iv) taxonomical ordering of variety by comparison—classifications might also reveal causes (e.g., Linnaeus);
(v) probabilistic and statistical analysis of events and populations—the search for stability of chance and uncertainty in numerical regularities (e.g., Maxwell); and

(vi) historical derivation of genetic development (e.g., Darwin).

Hacking's (1985) influential concept of style of scientific reasoning originates from Crombie's and Hacking's own work on the preconditions for the emergence of probabilistic thinking, influenced by Foucault's archaeology of knowledge—a historicized Kantian epistemology of worldview-like epistemes. These preconditions included a transformation of the medieval notion of opinion into a concept of internal evidence. The new concept opened a space for theories of probability that enabled the possibility of thinking about inductive logic, statistical inference, and some interpretations of quantum mechanics. For Hacking (1985), a style of reasoning constitutes a space of intelligibility and objectivity, the precondition for truth and falsehood, prior to the truth of propositions. For instance, the proposition "Mercury salve is good for syphilis because mercury is signed by the planet Mercury, which signs the marketplace, where syphilis is contracted" is a candidate for truth-or-falsehood determination only within a style of reasoning central to the Renaissance based on the concepts of resemblance and similitude. Hacking contrasts his notion of style with Quine's (Quine and Ullian 1970) web of beliefs and Davidson's (1974) conceptual schemes, which deal only with truth; with Kuhn's (1962) paradigms, which are not cumulative and depend on specific exemplary items of knowledge; and also with logic, which deals only with truth preservation. For Hacking (1985), the historicity, pluralism, and constructive aspects of styles of reasoning do not undermine the notions of objectivity and rationality, but they do provide a model of the disunity of science (Galison and Stump 1996).

Jordi Cat

References

Cassirer, E. (1937), "Wahrheitsbegriff und Wahrheitsproblem bei Galilei," *Scientia* 67: 120–130 and 185–193.

Crombie, A. A. (1993), *Styles of Scientific Thinking in the European Tradition*, 3 vols. London: Duckworth.

Davidson, D. (1974), "On the Very Idea of a Conceptual Scheme," *Proceedings and Addresses of the American Philosophical Association* 47: 5–20.

Dilthey, W. ([1911] 1931), *Weltanschauungslehre, Abhandlungen zur Philosophie der Philosophie*. Leipzig: Teubner.

Fleck, L. ([1935] 1979), *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.

Galison, P., and D. Stump (eds.) (1996), *The Disunity of Science*. Stanford, CA: Stanford University Press.

Hacking, I. (1985), "Styles of Scientific Reasoning," in J. Rajchman and C. West (eds.), *Post-Analytic Philosophy*. New York: Columbia University Press.

Holton, G. (1973), *Thematics Origins of Scientific Thought*. Cambridge, MA: Harvard University Press.

Husserl, E. ([1936] 1970), *The Crisis of the European Sciences and Transcendental Phenomenology*. Evanston, IL: Northwestern University Press.

Jones, C. A., and P. Galison (1998), *Picturing Science, Producing Art*. New York: Routledge.

Kuhn, T. S. (1962), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Mannheim, K. ([1929] 1936) *Ideology and Utopia*. New York: Harcourt.

Popper, K. R. (1935), *Logik der Forschung*. Vienna: Julius Springer Verlag.

Quine, Willard Van, and J. S. Ullian (1970), *The Web of Belief*. New York: Random House.

Toulmin, S. (1961), *Foresight and Understanding*. New York: Harcourt.

Windelband, W. (1891), *A History of Philosophy*. London: Macmillan.

Wisan, W. L. (1981), "Galileo and the Emergence of a New Scientific Style," in Hintikka, J., Gruender, D., and Agazzi, E. (eds.), *Theory Change, Ancient Axiomatics, and Galileo's Methodology*. Dordrecht: Reidel, 311–339.

Wölfflin, H. ([1915] 1950), *Principles of Art History: The Problem of Development of Style in Later Art*. New York: Dover.

JOHN SEARLE

(31 July 1932–)

John Searle was educated at the University of Wisconsin and at Oxford University as a Rhodes Scholar. His Ph.D. was directed by Peter Geach. For almost all of his professional career, he has taught at the University of California at Berkeley, where he is Mills Professor of Philosophy.

Searle has developed sophisticated and interlocking theories of language, mind, and social reality. These theories are held together by a realist metaphysics and a correspondence theory of truth. Primarily, he has contributed to the philosophy of science (i) a criticism of the dominant project of cognitive science, (ii) a theory of the structure of consciousness, which he thinks should guide or at least constrain cognitive science, and (iii) a theory of the nature of intentionality. Although this article will focus on these three issues, it is necessary to begin by saying something about his philosophy of language, because, like many twentieth-century philosophers, he has to a large extent read off the structure of the mind from the structure of language. Once his philosophy of language is explained, his other views relevant to the philosophy of science will be explained in reverse order.

The Structure of Speech Acts

Following J. L. Austin, Searle claims that the basic unit of language is the speech act. The most explicit kind of speech acts are those that Austin referred to as explicit performatives or illocutionary acts, such as those that typically would be performed by saying, for example, "I state that Ava will be at the party," "I promise that Ava will be at the party," "I ask, you, Ava, to be at the party." Searle points out that as indicated by the grammatical structures of these sentences, the speech acts are helpfully divided into two parts. First, each has the same content, in expressing the proposition that Ava will be at the party. Second, each of the relevant propositions performed by these sentences differs with respect to the force with which each proposition is expressed: stating, promising, and asking, respectively.

Most of the utterances used to perform speech acts are not as explicit as the ones in the examples above. In the right circumstances a person may utter an expression that indicates just part of the proposition or the force and still succeed in communicating the entire thought by relying on the hearer's knowledge of the context, for example "[I request that you close] the door," "[I order you to] leave!," "I insist [that you leave now]," respectively.

As regards the analysis of specific explicit, nondefective speech acts, several fairly sharp types of conditions are detectable. For the following (simplified) analysis of nondefective promising, the various kinds of conditions are italicized:

> A speaker *S*, in uttering a sentence *T,* promises an addressee *H* that *S* will perform an action *A* only if
> *Preparatory:* it is not obvious in the normal course of events that *S* will do *A*; and *H* wants *S* to do *A*.
> *Sincerity: S* intends that *S* will do *A*.
> *Propositional content: A* is a future action.
> *Essential: S* and *H* recognize that in uttering *T*, *S* incurs an obligation to do *A*, and *S* and *H* recognize that the other recognizes this.

Speech Acts and Intentionality

Searle's theory of speech acts has obvious connections with mental phenomena. The sincerity condition is always some mental state—for example, intending for promising, believing for stating, and desiring for commanding or requesting. Mental states, like speech acts, usually have propositional contents (e.g., intending that one will paint the house). Some essential conditions are connected with mental phenomena too; for example, an apology is an expression of sorrow, and a congratulation is an expression of joy.

Some general features of speech acts that also indicate a relation to mental states can be codified in four dimensions: (1) illocutionary point (purpose), (2) the direction of fit between the words and the world, (3) the type of sincerity, and (4) the propositional content; and some of these have correlates in mental phenomena (see Table 1).

Table 1 Major features of speech acts

	Illocutionary point	Direction of fit	Sincerity	Propositional content
Assertives or representatives	Commits *S* to truth of proposition	Word to world	Belief	Unrestricted
Directives	Commits *H* to do *A*	World to word	Wanting	*H* does *A*
Commissives	Commits *S* to do *A*	World to word	Intention	*S* does *A*
Expressives	Expression of attitude	None	Some psychological state	*S* or *H* (expand) + property
Declarations	"Declarational attitude"	Word to world	None	Some state of affairs

Assertives and beliefs have word-to-world direction of fit; both are supposed to match the way the world is. In contrast, directive and desires (or wants) have world-to-word direction of fit, as do commissives and intentions. Concerning the sincerity condition, only declarations do not have some requisite mental state. For assertives, it is belief; for directives, it is wanting or desiring; for commissives, it is intending; and for expressives, it is some pro or negative state, such as joy, anger, or resentment. Further, the four types of speech acts just mentioned and their correlative psychological states have the same propositional contents. To state that *p* is to believe that *p*, and to command that a hearer do an action *A* is to desire that the hearer do *A*.

Given these and other correspondences between language and mental phenomena, one might ask which, if either, is more basic. Some philosophers hold that language is more basic than the psychological states it expresses, because the states are made possible by the language used to describe them. Roughly, people posit the existence of psychological states in order to explain the behavior of people, and by extension some other creatures, mostly vertebrates. Psychological states and events are entities of a primitive scientific hypothesis. (It is this view that inclines some philosophers to hold that beliefs do not really exist; that is, beliefs are parts of primitive hypotheses that neurological science may show are as much a fiction as bodily humors.) In contrast, Searle thinks the opposite is the case. In his terminology, psychological phenomena have *intrinsic intentionality* (see Intentionality), by which Searle means "that feature of the mind by which mental states are directed at, or are about or of, or refer to, or aim at, states of affairs in the world" (Searle 1998, 64–65). Like force, mass, and gravitational attraction, things that are intrinsically intentional are such independently of any community of observers. It is internal to an intentional experience that it seems to be about something. Some things, notably language, have "derived intentionality." Individual words and sentences are directed at or are about things in the world because people intend them to be so. Sentences mean things because people have psychological states that are expressed by the sentences. Searle's criticism of one form of artificial intelligence, to be discussed below, turns on the claim that computers have only derived intentionality. Computer states and their output mean things only because people take those states to mean those things.

The priority of psychological phenomena over linguistic phenomena is intimately related to another property shared by speech acts and mental phenomena. Just as statements can be true, so can beliefs. Just as promises can be kept, intentions can be fulfilled. Just as directives can be obeyed, desires can be satisfied. In general, illocutionary acts and psychological states alike have conditions of satisfaction: "An intentional state is satisfied if the world is the way it is represented by the intentional state as being" (Searle 1998, 103). A belief is false, an intention is not fulfilled, a desire is not satisfied, and so on when their conditions of satisfaction do not obtain.

Although there is a connection between basic kinds of illocutionary acts with belief and desire, they are not the primary forms of intentionality. Perception and action are. At the heart of these two concepts is that of having intentional causation in their conditions of satisfaction. When a person sees something, the object causes the seeing, and the seeing represents (or, perhaps, better, presents) to the person what is seen. When a person raises an arm (in Gödel-Lob [GL] logic), the person's intention to raise it causes the arm to go up, and that intention contains a representation of that action. In short, perception and action involve both a causal and an intentional aspect, and they need to work together. If they did not work together, human survival would be impossible. Intentional causation is quite different from Humean causation, which consists of regular, contiguous conjunctions of events. It is essential to intentional causation for the cause to be a representation of the effect or the effect to be a representation of the cause. If a person drinks water as a way of satisfying (GL) thirst, then that person's (GL) mental state, that is, the desire the person drink water, causes it to be the case that she drinks water. The desire both causes and represents its condition of satisfaction. As regards intended actions, in contrast with accidental or mistaken actions, intentions form part of the conditions of satisfaction. A person who raises an arm must intend that the arm go up and that the arm's going up be caused by the intention that it go up. Actions are causally self-referential.

This causal self-referentiality also applies to perception. For someone to see something—say, a tree—that very tree must cause the visual experience of seeing it. The same holds for memory. If someone remembers playing tennis at Wimbledon, then the remembered event of playing tennis at Wimbledon must cause the memory. The causally self-referential components of seeing and remembering

Table 2 Components of international states and actions

	Seeing	Believing	Acting	Remembering
Intentional component	Visual experience	Belief	Experience of acting	Memory
Presentation or representation	Presentation	Representation	Presentation	Representation
Causally self-referential	Yes	No	Yes	Yes
Direction of fit	Mind to world	Mind to world	World to mind	Mind to world
Direction of causation	World to mind	World to mind	Mind to world	World to mind

form part of the conditions of representation of these phenomena.

As this discussion of acting, perceiving, and remembering suggests, the concept of direction of fit applies to them just as it applies to various categories of illocutionary acts. Since one's visual experience needs to fit the world, seeing has mind-to-world direction of fit; and since one tries to make the world satisfy one's desire, action has the world-to-mind direction of fit.

The various components of seeing (an instance of perceiving), believing, acting, and remembering are summarized in Table 2.

As explained thus far, it may appear that Searle has an atomistic view of mental phenomena. But that would be a misimpression. All mental phenomena take place against a backdrop of pre-intentional "capacities, abilities, tendencies, habits, dispositions, … and 'know-how'" (Searle 1998, 107–108) and within a network of beliefs, desires, memories, and so on. Among other things the network explains what makes a particular perception a perception of some object well known to the perceiver. One of the conditions of satisfaction for seeing a familiar object is that the object causing the visual experience be a particular object that occurs as part of the conditions of satisfaction of memories in the network.

The Structure of Consciousness

Acting, perceiving, believing, and desiring are so pervasive and salient that one might ignore the fact that each requires consciousness, at least in the basic cases (see Consciousness). Paradigmatic cases of acting, perceiving, believing, and desiring are conscious. When consciousness is not explicitly present, it must be theoretically possible for the mental state to be able to become conscious. A large literature has recently sprouted about zombies, beings that are functionally like humans, but nonconscious. Searle thinks such views are absurd or dangerously misleading. As a matter of biological fact, one cannot have seeing or acting without consciousness: "In real life you cannot subtract the consciousness and keep the behavior….[W]e only understand intentionality in terms of consciousness" (Searle 1998, 63–65).

Searle uses the point that that every mental state must be theoretically or potentially conscious to criticize Noam Chomsky's claim that humans follow rules of universal grammar even though these rules are not conscious (see Chomsky, Noam). According to Searle, Chomsky's view is incoherent because he does not say what feature of these rules makes them mental rather than purely physical brain states. An even more important application of this point concerns Searle's criticism of a dominant understanding of cognitive science, as seen below.

Consider now the structural features of consciousness itself. Although Searle identifies ten features of consciousness, they may be discussed under five categories, or features.

The first is that consciousness is *ontologically subjective.* It exists only "as experienced by an agent" (Searle 1998, 73). Some contemporary philosophers want to deny this because it does not seem to fit into their world. Being physicalists, they must also be materialists. They want to avoid anything like Cartesian dualism. Searle fights against this conflation of the physical and the mental, and maintains that commitment to both the mental and the nonmental does not necessitate a commitment to dualism. For him, consciousness is just as real as material objects; but it has a first-person (subjective) ontology, in contrast with the third-person (objective) ontology of the things usually studied by natural science.

The second feature of consciousness is its unity. The various instances of awareness that a person has at one time of sight, sound, smell, thought, and so on are part of one experience. This might be called "vertical unity," and it exists at any instant. Consciousness is unified in another way: across time. Consciousness would not be coherent if there were no short-term memory. Searle calls this "horizontal unity." It is tempting to think that the unity of consciousness is achieved by hooking together discrete mental states, that unity is the result of binding distinct elements. But Searle thinks that consciousness by its nature has a unity, and the various "contents" of consciousness are better understood as modifications of consciousness. Searle uses the metaphor of a field: "[We should] think of consciousness ... as a vast field, and think of the particular percepts, thoughts, experiences, and so on, as variations and modifications in the structure of the field" (Searle 1998, 82). Related to the unity of consciousness is the fact that experiences are normally structured. The mechanisms underlying this cause one to see a few lines as a human face or a sequence of stationary flashing lights as moving.

The third feature of consciousness connects a person to the external world (GL) in two ways: cognitive and volitive. People represent the way the world is and how they would like the world to be. Their orientation to the world is always colored in some way. People are always in some dominant mood or other that affects how things get experienced. What might cause anger in one mood might cause bemusement in another. People recognize that they are sometimes in good moods or bad moods; indeed all conscious states are "pleasurable or unpleasurable to some degree" (Searle 1998, 80). For the most part, it is difficult or impossible to describe precisely quotidian moods.

A fourth feature is that there are degrees of conscious attention, which again is in two dimensions. A person may be more or less attentive to the main focus of consciousness, say, a lecture; and a person is more attentive to that main focus than to what is peripheral. Thus, one might be paying much attention to a lecture and also be paying a little attention to the squirming child to the left.

The fifth feature seems to be closely related to Searle's idea of the network: the connectedness of all experience. People typically have a sense of their own situatedness: where they are roughly in space and time, what season of the year it is, what city and country one is in. It would be hard, if not impossible, to understand anything if one did not have some idea of where one was in space and time. Also, experiences come in varying degrees of familiarity. Entering a building, one has a sense for the kind of building it is, the kinds of facilities it will have, and where some of them are likely to be. Even otherwise strange places have their own familiarity. One would not expect to come upon a ski lodge in a jungle. A person who currently sees only the cover of a book expects to have certain tactile sensations if the person touches it; and if one opens it, one expects it to open to pages with print. Some gag gifts are "books" with provocative titles consisting of one hundred or more blank pages.

The Critique of Strong Artificial Intelligence

One of the merits of Searle's view is the fact that it accommodates both common sense and the scientific view of things. Mind is not the same as body; but mental phenomena cause, are caused by, and are realized in brain states. Consciousness is to the brain as solidity or liquidity is to molecules. One thing that guarantees the irreducibility of the mental is its place within a level of description that does not normally include talk about brain cells. It is also guaranteed, according to Searle, by the fact that consciousness, the qualitative experience of sights, sounds, and smells, is real. It is ontologically subjective but nonetheless a part of the physical world and must be explained. Comparing consciousness to the working of a four-cycle internal combustion cylinder, he says that, at the macro level of discourses, it is appropriate to talk about the spark plug firing and the explosion of the gas in the cylinder. At the micro level, the concepts of firing and exploding have no place, but "oxidation of individual hydrocarbon molecules" does. Both are causal accounts. Similarly, at the macro level it is appropriate to talk of consciousness and thinking, even though it is not appropriate at the micro level. Searle did not appreciate how central the ineliminability of consciousness was to his critique, and he rarely mentions consciousness (Searle 1980). This changed in his later publications: "The common sense objection to strong AI [artificial intelligence] was simply that the computational model of the mind left out the crucial things about the mind such as consciousness" (Searle 1992b, 45; 1998, 98).

Searle distinguishes between strong AI and weak AI. Strong AI is the view that "brain processes (and mental processes) can be simulated computationally" (Searle 1992b, 201–202). The hardware of the mind is unimportant. It could be carbon-based brain cells, as in human beings; but it could just as easily be silicon chips, vacuum tubes, water tubes, or even appropriately trained cats and dogs. What is important to being a mind, according to strong AI, is that the machinery instantiate a computer program. To do this is to think or to understand or to have desires and intentions. An appropriately programmed computer is a mind. Consequently, to understand what computer programs are is to understand what mental phenomena are.

Searle thinks that strong AI is false. In order to refute it, he constructs the following (simplified) thought experiment. Suppose that a person *C* who does not know Chinese is locked in a room and given a book, written in *C*'s native language, that correlates Chinese characters in one column with

Chinese characters in another. The book includes the instructions that when a slip of paper is passed through a slot in the room, *C* looks up the characters written on it in the book, finds their correlates, chooses another slip of paper that has the correlates written on it, and passes it back through the slot. Further suppose that *C* is so good at this operation that people outside the room who do not know what goes on inside the room would have good evidence for thinking that someone or something inside the room knows Chinese. From the external point of view, the paper output is functionally no different from what one would expect of someone or something that knows Chinese.

On this, Searle and Chomsky are on the same side. Searle now makes two observations: What *C* is doing in the room is essentially what a computer does; and *C* obviously does not understand or know Chinese. It follows that computers do not understand or know languages. QED.

One difference between Searle's original thought experiment and the rendition of it just given is that instead of talking about *C,* Searle describes the scenario in the first person: "Suppose that I'm locked in a room," etc. Searle wants readers to put themselves in the position of the person in the room. It is the reader who is looking up the characters in the instruction book and then choosing the appropriate output. The reader has a privileged perspective on what the reader knows. Further, since there is no other person or thing connected with the room that plausibly knows Chinese, the reader seems justified in concluding that instantiating a computer program is not sufficient for knowing a language, or anything else. The basis for the readers' belief that they do not know Chinese is their recognition that they do not know it.

Searle maintains that various mistakes in cognitive science, not to mention the philosophy of language, result from taking the third-person point of view, either because one thinks that that is the only scientific way to study anything or because there is no genuine first-person point of view, the latter being either epiphenomenal or an illusion. In contrast, Searle believes that the first-person perspective is not just a legitimate one but is a privileged perspective for judging what a person knows. As he says: "Remember, in these discussions, always insist on the first person point of view" (Searle 1980, 451).

In the original debate, neither Searle nor his critics called attention to the role of consciousness in this thought experiment. For all of them, the main issue was put in terms of what a computer could know and whether one could think. It is plausible that an important source of the impasse between them was a disagreement about the criteria for thinking and knowing. Searle implicitly thought that consciousness was a necessary condition for thinking. The proponents of artificial intelligence maintained that satisfying a functional or behavioral criterion was sufficient for thinking. This was why they appealed to the Turing test.

Someone trying to mediate the dispute between Searle and his opponents might suggest that the proponents of strong AI concede to Searle that computers today and for the foreseeable future will not be conscious, and that Searle concede to the proponents of strong AI the behavioral criteria they want for words like 'think' and 'know' as applied to computers. In this proposal, the different criteria applied to humans and computers would be analogous to the varying criteria that are applied to the word 'strong' in relation to humans and elephants.

At least some proponents of strong AI would probably reject the proposal, either because they think that consciousness is an illusion, so to speak, or because they think that consciousness emerges from the complex computations that would be performed by *C.* The consciousness might not attach to *C* exactly, but to the room and contents as a whole. This is the "systems" objection to the Chinese-room scenario.

However, Searle would probably reject the proposal because he thinks that the behavioral criterion suggested for thinking is so far from capturing what is essential to thinking and knowing that it changes the meaning of 'think' and 'know.' In addition to the point that thinking depends on consciousness, Searle says that thinking is intrinsically intentional, while computer states are not. The physical states or outputs of computers have derivative intentionality, as language does. What goes on within a computer are purely formal or purely syntactic operations. Therefore, the computer states are not intrinsically semantic; they are not intrinsically about anything in the world. Human beings, however, can interpret or assign a meaning to the various states or outputs of the computer, just as they do for linguistic signs. In other words, "the characterization of a [computer] process as computational ... is essentially an observer-relative characterization. ... It requires the assignment of a computational interpretation by some agent" (Searle 1992b, 210–211). A computer has nothing inside it that has direction of fit, propositional content, or conditions of satisfaction.

Searle's thought experiment was opposed by most cognitive scientists. Among the many objections advanced, two are important. According to the systems objection, it is not *C* who is supposed to understand; *C* is just the central processor; rather, what understands Chinese is the entire system, which includes *C*'s ledger with the rules of correlation and *C*'s calculations. Consequently, the fact that *C* recognizes that one does not know Chinese is not decisive. Searle's reply aims to show that even when all the aspects of the system are considered together, it is obvious that nothing in or about the room knows Chinese. Suppose that everything in and about the room is put into the head of *C*, including the information in the ledger. *C* would still know that she does not understand Chinese, says Searle. Hence, no system relevant to the room understands Chinese.

The other main objection to Searle's thought experiment is the "robot" objection. The reason *C* does not understand Chinese is that one lacks substantive causal interaction with the world outside the room. *C* would know Chinese if one interacted with one's environment. According to the robot objection, if *C* can observe what is outside via a television camera and if one can move the room from place to place, say, on wheels, if *C* can manipulate objects in the environment, with armlike devices, and if *C* can do all of this in virtue of the Chinese instructions that enter the room and then get decoded via the rules in the ledger, then *C* knows Chinese. In fact, *C* is no longer needed at all. Let *C* be a computer that causes things to happen in virtue of the machinery just described. In such a case, the computer in the room would understand Chinese and have other mental states. The proponent of strong AI may then observe that a human being is just like the robot, except that a brain stands in for a computer, eyes for the television camera, and so on (Searle 1980, 420.)

Searle's response to the robot objection in effect consists of two parts. The first concerns the case in which *C* is hooked up with cameras, wheels, and levers. Searle observes that in this case *C* is starting to learn Chinese. When *C* learns from the (GL) rule book that *C* is to grab a salt shaker and pass it to someone viewed on the television screen, *C* is giving a semantic content to the previous purely formal or syntactic elements, and hence taking a step toward language comprehension. But this case is very different from the original one in which *C* was manipulating purely formal elements. The second part of Searle's reply concerns the allegation that the human (in a human body) is like a computer (in a robot). Searle does not deny that machines think. Indeed, he claims that only machines think. (Humans are machines.) Rather, he claims that it is a matter of physical fact about this world that the carbon-based structures of the brain are such that their specific composition gives rise to consciousness, and the silicon-based structures of a computer do not. Searle is not saying that no other kind of material in this world could also cause consciousness. In some solar system of some galaxy in the universe there may be conscious creatures who have a very different physical composition. Similarly, he is not saying that no silicon-based machine of a different design and complexity could not give rise to consciousness. But in this world, given their current composition and mechanisms, computers do not think.

Searle's critique of strong AI has the consequence that cognitive science should abandon the idea that computer programs are accurately representing either mental or neurological processes. His views about the structures of consciousness, identified by taking the first-person point of view, constrain any adequate study of consciousness and intelligence.

A. P. MARTINICH

References

Fotion, Nick (2000), *John Searle*. Princeton: Princeton University Press.

Garnett, William (1987), *The Springs of Consciousness*. Cornwall: Tabb House.

Lepore, Ernest and R. van Gulick (1991), eds. *John Searle and His Critics*. Cambridge, MA: Blackwell Publishers.

McGinn, Colin (1991), *The Problem of Consciousness*. Oxford: Blackwell Publishers.

Martinich, A. P. (2001), "John Searle." In *A Companion to Analytic Philosophy*, eds. A. P. Martinich and David Sosa. Malden, MA: Blackwell Publishers, 434–50.

——— (1984). *Communication and Reference*. New York: Walter de Gruyter.

Searle, John (1969), *Speech Acts*. Cambridge: Cambridge University Press.

——— (1979), "What Is an Intentional State?" *Mind* 88: 74–92.

——— (1979), *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

——— (1980), "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3: 417–24.

———(1982), "The Myth of the Computer." *New York Review of Books* 29, no. 7: 3–6.

———(1983), *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

———(1984), *Minds, Brains, and Science*. Cambridge, MA: Harvard University Press.

——— (1989), "Consciousness, Unconsciousness and Intentionality." *Philosophical Topics* 12: 193–209.

——— (1992a), "Collective Intentions and Actions." In *Intentions in Communication*, ed. P. R. Cohen, J. Morgan, and M. E. Pollock. Cambridge, MA: MIT Press.

——— (1992b), *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

——— (1995a), *Construction of Social Reality*. New York: The Free Press.
——— (1995b), "Consciousness, the Brain, and the Connection Principle." *Philosophy and Phenomenological Research* 45: 217–32.
——— (1997), *The Mystery of Consciousness*. New York: *New York Review of Books*.
——— (1998), *Mind Language and Society: Philosophy in the Real World*. New York: Basic Books.

UNITS AND LEVELS OF SELECTION

See **Biology, Philosophy of; Evolution; Natural Selection**

SIMPLICITY

See **Parsimony; Unity and Disunity of Science**

SOCIAL CONSTRUCTIONISM

The first use of the words 'social construction' to signal a theoretical orientation was by Peter Berger and Thomas Luckmann (1966), in their book *The Social Construction of Reality*. Berger and Luckmann argued that "constructed" social reality is an empirically tractable domain; an objective world of a different order than the world of natural science. They did not question the objectivity of either natural or social reality, but instead attempted to develop a distinctive theoretical approach to the latter. They drew upon phenomenology (particularly Alfred Schutz's [1972] phenomenology of the social world) but adapted it to a more conventional form of social explanation. Unlike Schutz, Berger and Luckmann did not attempt to explicate the life-world from the standpoint of a typified member. They used the word 'construction' as an Americanized and sociologized variant of the phenomenological notion of "constitution" (an acausal coordination of intentional acts and intentional objects), and they reverted to a more familiar form of structural theory, using a historical overview from which to develop a general explanation of how subjective actions give rise to stable, externally constraining, and (for all practical purposes) objective institutions. Their account of institutions combined a critique of positivism—arguing against one-sided views of social institutions as being "out there" in a world independent of human agency—with an empirical outlook that invited investigation of how such institutions become objectified.

Berger and Luckmann assumed that their genealogy of social institutions did not apply to natural regularities. Instead, their concept of social construction described the emergence of institutions from rituals, customs, and beliefs in particular historical circumstances. For Berger and Luckmann, the compulsive and constraining force of institutional reality is no less binding than natural laws, but this force inhabits a second-order reality of norms and evaluations—a reality that becomes second nature, entrenched in the actions of properly socialized members. Berger and Luckmann deemphasized ontology and made the point that members of particular cultures who lack cosmopolitan perspicacity are unlikely to recognize the difference between socially constructed institutions and natural realities, since both make up the unquestioned background of daily practice. Decades later, John Searle (1995), in *The Construction of Social Reality*—a title with an uncanny, although apparently unintentional, resemblance to its unacknowledged predecessor—took a far more explicit interest in drawing ontological distinctions. Searle articulated two distinct lessons about socially constructed realities: first, that "institutional facts" like the value of currency or the rules of chess are both *real* and essentially dependent upon the existence and maintenance of collective ideas and actions; and second, that such socially constructed realities differ essentially from "brute facts" like the height of Mount Everest or the double-helical structure of DNA. Contemporary social constructionism in the social sciences is distinguished by an irreverent stance toward that very distinction. Like Berger and Luckmann, Searle aimed to establish that constructed reality *is* real, albeit not in the same way as heights of mountains and structures of molecules. Contemporary constructionists take for granted that social constructions are real, and are more inclined to extend reality construction to cover geological measures and molecular structures.

The Diffusion of Constructionism

In the three decades between the publication of Berger and Luckmann's and Searle's essays on social construction, the theme of 'construction' became widespread in the humanities and social sciences. The term 'construction' assumed a life of its own independent of Berger and Luckmann's theory, which one could say describes the academic institutionalization of constructionism. Especially in the 1980s and 1990s, social and cultural construction became a central theme for a movement in the humanities and social sciences. An indication of how far the movement spread is a lengthy list compiled by Ian Hacking (1999, 1) of books with 'construction' in their titles. This list includes works on topics as diverse as scientific facts, illness, women refugees, and Zulu nationalism. The various constructions include culturally prevalent ideals and ideas, as well as more substantive institutional patterns and programs.

As Hacking points out, contemporary social constructionism has little affinity with cognate developments in earlier eras. The nineteenth-century mathematicians and early-twentieth-century artists sometimes called 'constructivists' tended to favor formalistic schemes and procedures. Today's social and cultural constructionists tend to take a skeptical or critical view of rules, algorithms, and other formal structures, emphasizing the interpretive flexibility of formalisms and stressing the role of locally situated actions. Hacking recommends that the term 'constructionism' be reserved for the recent movement, in order to distinguish it from unrelated forms of 'constructivism,' but both terms continue to be used interchangeably. He argues that social constructionists typically challenge conventional sensibilities by demonstrating that a phenomenon that *appears* to be natural, determinate, or inevitable (e.g., the natural basis of sexuality/gender, the medical status of one or another mental illness, the inevitable march of technological progress) is actually more flexible and less determined than usually imagined. Often combined with such arguments is a denunciation of the state of affairs held responsible for the construction. Such denunciations encourage efforts to reform the existing state of affairs by counteracting assumptions about their natural foundations and inevitability.

Hacking's inventory demonstrates the extent to which constructionist arguments depend upon *what* they take up for analysis. To successfully provoke interest and stir controversy, a constructionist study typically starts with a subject that is taken for granted or assumed to be inevitable (Hacking 1999, 12). The challenge for the study is to demonstrate how a condition that seems inevitable might have turned out differently, and might yet be changed through concerted social action. Some subjects (e.g., technological innovations) might seem unpromising for such an argument, because everyone already knows that they are constructed. However, successful constructionist programs can frame the subject in an appropriate way—for example, by taking aim not at technology per se but at the *idea* that technological progress is governed by rational imperatives (Bijker, Hughes, and Pinch 1987). Now that the constructionist idiom has been adopted in one field of study after another, it has *diffused* in more ways than one. While constructionist programs can be found throughout the social sciences and humanities, and also in schools of business management, law, nursing, social work, and education, and even in corporate technology research centers, just what hangs on the idea of "construction" has become quite diffuse. Depending upon which corners of the constructionist universe one explores, it is possible to find accents of idealism, realism, historicism, pragmatism, functionalism, empiricism, and instrumentalism.

The Construction of Science

Perhaps the most provocative examples of constructionism are found in the field of science studies (a transdisciplinary field including historians, sociologists, anthropologists, and a few philosophers of science). Starting in the early 1970s and inspired by Thomas Kuhn's *Structure of Scientific Revolutions* ([1962] 1970), proponents of the strong program in the sociology of scientific knowledge (SSK) argued that even the most established scientific and mathematical truths were subject to "social" explanation. David Bloor (1976) and other proponents of the strong program did not call themselves constructionists, but their writings are frequently cited in connection with the genre. Two of the most influential and controversial proposals from the strong program were that a sociologist of knowledge should attempt to be *impartial* toward the beliefs studied and should give *symmetrical* explanations of why they are believed, regardless of their (alleged) truth or falsity, rationality or irrationality, or success or failure. A practical rationale for impartiality and symmetry is that a sociologist is unlikely to know which side in a scientific controversy is on the side of truth and rationality. In scientific controversies it often is the case that proponents on both sides explain their own views as consequences of correct procedure while charging their opponents with ignorance, irrationality, and vested interests. A sociologist is in a poor position to guarantee that one or another currently accepted scientific fact or law will continue to be held true for all time. In addition, consistent with a more general policy of methodological charity, a sociologist or anthropologist may want to suspend initial preconceptions about the rationality or irrationality of a belief system when seeking to explain the origins of that belief or understand why it is accepted. Symmetry and impartiality are methodological policies, but they are often confused with metaphysical positions that deny the possibility of having true knowledge, and the strong program is sometimes accused of treating all systems of belief indiscriminately, as equally valid.

The policies of symmetry and impartiality, and the attempt to explain scientific knowledge as socially caused belief—the hallmarks of the strong program—raise many conceptual and methodological problems, such as the apparent incongruity between relativism about nature and realism about society, the puzzling meaning and implications of the claim that "true" and "false" knowledge or belief can be explained in the same general way, and the confusion that results from adopting a general explanatory term ('construction') that is commonly used to dismiss empirical claims (Laudan 1981; Coulter 1989; Lynch 1993). However, Bloor and his colleagues do not deny truth, nor do they suggest that specific scientific results are untrue, irrational, or unsuccessful. Indeed, Barnes, Bloor, and Henry (1996) identify their position with realism.

The constructionist idiom became explicit with the publication of Latour and Woolgar's (1986) study of a biochemistry laboratory at Salk Institute in San Diego. The authors framed their study with the metaphor of an anthropologist studying an exotic tribe in which the natives share esoteric knowledge, speak (technical) dialects unintelligible to an outsider, and perform arcane rituals with strange apparatuses and sacrificial animals. Latour and Woolgar spoke of facts as anthropological constructions. They defined a fact as a statement of the form '*X* is *Y*' from which qualifications, references to circumstances, and grammatical markers of uncertainty have been deleted. The main case they examined was a successful effort to identify the molecular structure of a particular growth hormone (eventually awarded the Nobel Prize). Latour and Woolgar insisted that the scientists did not begin with the fact they eventually helped establish; instead, the scientists put forward a series of tentative, and often contested, "representations," which were defended against rival efforts to "deconstruct" them, and only after the controversy was settled was the fact endowed with transcendental status and stripped of any reference to the uncertainties and contingencies that were conspicuous at earlier times. The genealogy of the "fact" placed it within a long chain of representations (in language, in literary inscriptions) at the end of which it became settled and took its place in nature.

Latour and Woolgar attempted to avoid the suggestion that the scientific fact they studied was *merely* a product of ideology, bias, or some other discrete "social" factor. To ward off the idea that they were making a straightforward sociological explanation, they went so far as to drop 'social' from the title of their book when they published a revised edition in 1986, so that the subtitle read *The Construction of Scientific Facts*.

Nevertheless, Latour and Woolgar (1986, 237) courted confusion by making provocative claims such as, "reality is the consequence, rather than the cause," of scientists' constructive activities. Alan Sokal (2001, 16), for example, interpreted the assertion as an ontological claim implying that natural reality does not preexist its discovery. However, the assertion becomes less provocative, and less absurd, when one understands Latour and Woolgar's genealogy of fact, *not* as a causal explanation, but as a semiotic analysis of the way scientists reformulate their factual accounts over time. It makes some sense to say that "reality" comes after (is a "consequence" of) a series of referential constructions if one maintains an agnostic attitude toward the ontological status of the "fact" in question and draws no distinction between the idioms of realistic reference and their ultimately "real" referents. The source of confusion is not necessarily on the side of naïve realist readers, however, since Latour and Woolgar's grammatical formulations often seem designed to provoke shock and confusion.

Other influential studies, such as Andrew Pickering's (1984) *Constructing Quarks*, adopted the constructionist idiom. Pickering employed a Kuhnian perspective (one that Kuhn did not personally endorse) to describe a revolution in late-twentieth-century particle physics. Although sometimes cast by critics as a simple causal story of external social and political influence on a scientific community, Pickering's narrative and those of other constructionists deploy more subtle arguments involving two distinct steps (see Collins 1985 for related arguments and case studies):

1. The constructionist invokes skeptical philosophical arguments and interprets them in a sociological manner, to say that the victorious theory was underdetermined by experimental evidence (see Underdetermination of Theories), that the experimental evidence was theory-laden (see Observation), and that no crucial test was responsible for the eventual victory of one of the rival research programs.
2. Having noted the lack of unequivocal evidence supporting the victorious theory, the constructionist examines the historical record for persistent disputes about particular experimental designs and interpretations. Then, when such disputes can be found (as they frequently are), the constructionist adduces evidence of distinct theoretical commitments and experimental styles, national divisions between research groups, vested interests, and other social and cultural orientations that may account for the alignment of key parties in the controversy.

The extension of constructionism to cover the "products" of the natural sciences and mathematics proved to be highly provocative and also rather confusing. In part, the confusion stemmed from the fact that the term 'construction' already had currency in the vernacular of science. It is commonplace, and noncontroversial, to speak of the construction of models, theories, and proofs. However, 'construction' and related expressions, such as 'artifact' and 'manufacture,' are fighting words in disputes about experimental results. When used in sociological explanations of particular episodes and general developments in science, such terms can seem to imply that specific results were concocted, or even that all scientific results are dubious. Whether such confusion is deliberately fostered by attention-seeking authors or results from naive misreadings of their arguments and descriptions remains an open question.

Constructionism became very influential in science-studies circles through the 1980s and 1990s and was an important part of a broader set of developments that included new professional societies, journals, and university science-studies programs. It would be grossly inaccurate to identify such developments with a single "school" or style of explanation, though constructionism was a major theme and topic of debate at the time. As an increasing number of social-historical and ethnographic case studies of science, technology, and medicine accumulated, it became common to cite the literature as evidence that traditional *idealized* versions of scientific method have been superseded by a developing empirical knowledge of a messy, uncertain, and contingent array of scientific practices. Critics occasionally pointed out that such empirical claims about *actual* science sat uneasily with a relativistic stance toward natural reality. In addition to being subject to variants of the *tu quoque* argument against relativism, social constructionists were accused of political quietism and dismissed for their lack of "normativity." For many feminists, constructionism was a valuable tool for opposing objectivism and male privilege, but it did not go far enough in the directions they wished to take it (Harding 1996; Haraway 1991). The question of whether a "symmetrical" stance toward science is compatible with normative programs of feminist epistemology and other efforts to democratize scientific knowledge continues to be debated.

The Science Wars and After

Social constructionism became a focus during the "science wars" in the mid-1990s. Unlike the many criticisms of constructionist science studies that had occurred over the previous twenty years, those associated with the science wars were put forward by natural scientists. The indignant missives written by a few of these scientists reached a higher level of explicit hostility and attracted much wider attention than had previous criticisms in academic journals and interdisciplinary symposia. Not only did the critics charge social constructionists with philosophical relativism and take issue with particular historical interpretations, they lumped social and cultural studies of science together with "postmodernist" attacks on "truth" and with diverse anti-science and anti-modern movements (Gross and Levitt 1994). Though often billed as a conflict between science and anti-science, the science wars involved social and natural scientists on both sides.

The vociferous critics of constructionism included a fair number of social scientists, philosophers, and other scholars who took the opportunity to pursue disputes that had been going on for many years. In addition, a number of scientists and mathematicians (as well as many historians and sociologists of science with backgrounds in science and engineering) defended social and cultural studies from some of the attacks. Physicist Alan Sokal's (1996) "hoax" article, which appeared in the cultural studies journal *Social Text*, touched off a wave of publicity about the science wars. Sokal's article claimed amazing parallels between quantum gravity theory and postmodern literary theory. When Sokal acknowledged just after his article was published that it was complete nonsense, he purported to expose the ignorance of science and lack of standards in the cultural studies field. However, his target differed from the anti-science that Gross and Levitt denounced. Sokal took delight (as well as offense) in revealing an odd variant of scientism in the arcane critical writings of French intellectuals and their American exponents. It seemed that, like more common varieties of science popularizers and science emulators, famous French intellectuals also had a tendency to couch their ideas in garbled versions of quantum theory, Einstein's theory of relativity, and chaos theory. Sokal was accused of being unimaginative and uncharitable in his readings of the authors he criticized (Stolzenberg, 2001), but he did not charge science studies with general hostility toward science. If anything, he charged particular writers and their followings with idolatry——with taking the name of Einstein in vain.

If scientism is an illegitimate extension of science into metaphysics, then the charge of scientism cuts both ways. A few of the self-proclaimed defenders of Science and Reason (who used these categories in an abstract, quasi-religious way) who joined the attack against social and cultural studies of science were authors of controversial sociobiological explanations of human capacities and sexual preferences. Others, who worked in more credible fields of physics and biology, wrote articles and books in which they disavowed philosophy while at the same time advancing their own intuitive philosophies of science. Both critics of Western science and defenders of Science and Reason could be accused of confusing science *as practiced* in specific organizational and historical circumstances with Science as an idealized worldview. In any case, the specificity of constructionism tended to get lost in the fray. Constructionist studies characteristically focus on specific historical sequences and organizational circumstances, but the science wars invited much broader claims and sweeping arguments than could ever be documented with detailed historical or sociological studies. As a philosophical debate joined by very few philosophers—in which the scientists involved rarely referred to their own specialized research, and the social scientists rarely went into detail about their historical and ethnographic research—it was not surprising that the science wars ran out of steam after a few years. The dispute had started to wane by the late 1990s, as a few participants continued a less fractious dialogue about the claims and implications of social studies of science (Labinger and Collins 2001).

By now, largely as a result of the widespread adoption and abuse of constructionism, the ideas associated with the genre have become so diffuse that the word construction and the tropes associated with it have lost their meaning, as well as much of their former cachet. In science-studies circles, strident polemical efforts to rewrite "nature" as construction, or "discovery" as "invention" appear less frequently and provoke less interest than they once did. Latour and Pickering, who helped establish the constructionist idiom in social studies of science, now pursue postconstructionist attempts to re-introduce nonhuman sources of agency into their theoretical schemes (Pickering 1995; Latour 1999). More interesting, perhaps, are continuing efforts to trace how variants of the nature/society and science/nonscience distinctions are deployed in specific institutional settings with tangible practical, political, and legal consequences (Gieryn 2000).

It is tempting to think of constructionism as a coherent epistemology, roughly akin to classic idealism but with more emphasis on the constitutive roles of embodied practices, social alignments, and technological instruments. However, in its current incarnation, constructionism has become too diffuse to support a single philosophical perspective. Moreover, the orientation of many social-historical and ethnographic studies to particular settings, and the emphasis on the specificity of research in different fields, is not easily subsumed under a single theory of knowledge. Instead of being a coherent epistemology, constructionism might better be considered an approach to "epistemography" (Dear 2001): an empirical study of particular historical and institutional settings in which participants organize and deploy what counts, for them, as observation, experimental evidence, truth, and knowledge.

MICHAEL LYNCH

References

Barnes, Barry, David Bloor, and John Henry (1996), *Scientific Knowledge: A Sociological Analysis*. Chicago: University of Chicago Press.

Berger, Peter, and Thomas Luckmann (1966), *The Social Construction of Reality*. New York: Doubleday.

Bijker, Wiebe, Thomas Hughes, and Trevor Pinch (eds.) (1987), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press.

Bloor, David (1976), *Knowledge and Social Imagery*. London: Routledge and Kegan Paul.

Collins, H. M. (1985), *Changing Order: Replication and Induction in Scientific Practice*. London and Beverly Hills: Sage.

Coulter, Jeff (1989), *Mind in Action*. Oxford: Polity Press.

Dear, Peter (2001), "Science Studies as Epistemography," in Jay Labinger and H. M. Collins (eds.), *The One Culture? A Conversation about Science*. Chicago: University of Chicago Press, 128–141.

Gieryn, Thomas (2000), *Cultural Boundaries of Science: Credibility on the Line*. Chicago: University of Chicago Press.

Gross, Paul, and Norman Levitt (1994), *Higher Superstition*. Baltimore, MD: Johns Hopkins University Press.

Hacking, Ian (1999), *The Social Construction of What?* Cambridge, MA: Harvard University Press.

Haraway, Donna (1991), *Simians, Cyborgs and Women: The Reinvention of Nature*. New York: Routledge.

Harding, Sandra (1996), "Standpoint Epistemology (a Feminist Version): How Social Disadvantage Creates Epistemic Advantage," in Stephen Turner (ed.), *Social Theory and Sociology: The Classics and Beyond*. Oxford: Blackwell.

Kuhn, Thomas ([1962] 1970), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Labinger, Jay, and H. M. Collins (eds.) (2001), *The One Culture? A Conversation about Science*. Chicago: University of Chicago Press.

Latour, Bruno (1999), *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.

Latour, Bruno, and Steve Woolgar (1986), *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Laudan, Larry (1981), "The Pseudo-Science of Science?" *Philosophy of the Social Sciences* 11: 173–198.

Lynch, Michael (1993), *Scientific Practice and Ordinary Action: Ethnomethodology and Social Studies of Science*. New York: Cambridge University Press.

Pickering, Andrew (1984), *Constructing Quarks: A Sociological History of Particle Physics*. Chicago: University of Chicago Press.

——— (1995), *The Mangle of Practice: Time, Agency, and Science*. Chicago: University of Chicago Press.

Schutz, Alfred (1972), *The Phenomenology of the Social World*. London: Heinemann.

Searle, John (1995), *The Construction of Social Reality*. London: Penguin.

Sokal, Alan (1996), "Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity," *Social Text* 14: 217–252.

——— (2001), "What the *Social Text* Affair Does and Does Not Prove: A Critical Look at 'Science Studies,'" in Keith Ashman and Philip Baringer (eds.), *After the Science Wars*. London and New York: Routledge, 14–29.

Stolzenberg, Gabriel (2001), "Reading and Relativism: An Introduction to the Science Wars," in Keith Ashman and Philip Baringer (eds.), *After the Science Wars*. London and New York: Routledge, 33–66.

PHILOSOPHY OF THE SOCIAL SCIENCES

Social scientists study the group behavior of human beings. Philosophers of social science study the methods of inquiry and patterns of explanation appropriate for the social sciences. A perennial topic of interest to philosophers of social science is the question of how different the methodology of the social sciences should be from that of the natural sciences. Aristotle believed that final causes played a crucial role in the explanation of human behavior, whereas natural scientists need look for only efficient and material causes. At the beginning of the twentieth century there was a concerted attempt to naturalize the social sciences (see Unity of Science Movement). The debate about whether social science requires a special form of understanding still resonates in current philosophical discussions.

What are some of the reasons for believing that social inquiry poses unique problems? In *The Poverty of Historicism* Karl Popper (1957) argues that attempts to predict human behavior are undercut by a radical openness of the system that has no parallel in the physical sciences (see Popper, Karl Raimund). What people do is influenced by the ideas they have about what is possible. It is also known that a person in a practical problem situation may come up with a completely novel solution to that problem and act on it. But by definition, it is impossible for anyone today to predict what will only be discovered or first thought of tomorrow! The ability of human beings to create new ideas and the responsiveness of human systems to those ideas seems to be different in kind from the openness of the solar system to asteroids.

Philosophers have also noted the phenomenon of "reflexive predictions," in which the very act of making a prediction can change the behavior it was trying to foretell. Such considerations have convinced many philosophers that social scientists should devote more attention to explanation than to prediction; but as will be discussed, there is little agreement about exactly which models of explanation are appropriate.

Another distinctive feature of social science is the fact that many of the basic phenomena to be studied come into existence only because people in a given society agree to their existence. What counts as money, marriage, or murder depends on a complex system of mutual understandings and trust that floats on top of and is largely autonomous from the physical and biological world. These important constituents of social life are created, maintained, and destroyed according to the shared intentions of the people who are being studied. For J. L. Austin a key feature of the social world is the ability to create obligations by means of "performative utterances." John Searle (1995) has proposed an analysis of institutions that is rooted in the collective intentions of members of a society (see Searle, John). The fact that so much of social reality is socially constructed strongly suggests that the methods and structure of social science will exhibit distinctive features.

So there are certainly prima facie reasons for expecting that the typical problems faced by social scientists may turn out to be quite different from the challenges found in the physical sciences that have formed the basis for most philosophizing about the nature of science. But various traditional philosophical misgivings about the very possibility of a science of human societies are undercut by the current success of social scientists who have been getting on with their job. For this reason much current philosophy of social science deals with issues that arise in the daily practice of social inquiry.

Varieties of Social Explanation

Philosophers of social science have worked with theorists in clarifying basic features of various research programs and exhibiting their strengths and weaknesses. The following approaches are discussed by Little (1991). Martin and McIntyre (1994), Salmon (1992), and Gasper (1991) also offer good summaries of various kinds of explanatory approaches.

Rational Choice Program

On a daily basis people interpret the activities of others around then by attributing to them beliefs and desires such that their behavior becomes an appropriate choice, given their situation as they perceive it. Economists have developed the basic tenets of "folk psychology," as this form of common sense reasoning is called, into a theory of Rational Choice. In its mathematical form the theory proposes normative solutions to problems about combining preferences and group decision-making and tackles long-standing paradoxes, such as the Prisoner's Dilemma (see Game Theory). Social scientists use rational choice theory (RCT) to provide descriptive models of dating behavior, the deterrence of criminal behavior, and the economics of everyday life.

However, philosophers and social theorists have raised many critical queries about the status of rational choice explanations. To take one much-discussed example: If one applies a straightforward RCT analysis to the decision of an individual to vote in a nationwide election, it seems that it is almost always irrational for any individual to vote. The probability that one vote will make a difference is so small that any minor inconvenience accompanying the process of voting quickly outweighs the expected utility of taking the trouble to vote. Yet people do cast votes. Does this mean that RCT fails to explain their action? Rational choice theorists have a variety of responses. The most popular one is to say that individuals living in a democracy feel they have a duty to vote and so this factor enters into their decision. But duties do not fit smoothly into the mathematical machinery of RCT. And it seems ad hoc to introduce such factors only to save the model. If one admits that duties can simply override the factors more traditionally invoked by economists, the explanatory program of RCT is seriously weakened.

The approach has been faulted on other methodological grounds. On the one hand, RCT accounts are very narrow in scope. Since one must feed into the model of the situation the beliefs and desires or subjective probabilites and utilities of the actors, RCT leaves unanalyzed and unexplained some of the most important aspects of socialization, namely, the formation of value systems and worldviews. Yet in other respects the RCT approach seems too widely applicable. Since only the choice is presumed to be rational, one is as free to apply it to the behavior of impulsive and delusional people as to more typical instances of rational decision making, such as that of a champion blackjack player or master politician.

Nevertheless, RCT remains one of the most popular explanatory approaches in social science. Philosophers who endorse the approach include Popper (1994a) (who somewhat surprisingly does not worry about the lack of falsifiability of the rationality principle) and Jon Elster (2000).

Functional Analyses

This approach was especially popular with anthropologists in the first half of the twentieth century. In its starting points it seems quite different from RCT because it deals with societies, not individuals, and is most impressive in cases where the stated motives of the participants are quite different from the so-called latent, or hidden, functions of the behavior (see Function). A standard example of a functionalist explanation goes something like this: Given the obvious ineffectiveness of rain dancing, why do certain cultures persist in this activity? The functionalist reply is that the custom has the latent function of promoting social cohesion, an important social need for people facing a drought.

But as philosophers have pointed out, by its very nature a functionalist account cannot tell why rain dancing, as opposed to some other functionally equivalent ritual such as a sacrifice to the gods, was adopted. Functionalists would relegate that question to the historian. Neither can it explain the persistence of customs long after their original function is no longer operative. For example, even if one grants that religious taboos against eating pork served a health function in days before trichinosis was understood, that explanation cannot explain the persistence of the taboo, and one must once again invoke something like social cohesiveness. It is very difficult to bring evidence to bear on the question of whether some long-standing custom contributes to social cohesiveness or not. If not a tautology that such customs at least do not detract seriously from social well-being, positing that they cement society together is certainly not very explanatory.

Even functional attributions that are more specific may be difficult to test. Consider for example the feminist claim that one reason that rape in this society persists is that it has the function of "keeping women in their place." Thus the threat of rape is said to benefit male graduate students by posing obstacles to the use of the library at night by female graduate students. How could one even begin to bring evidence to bear on this allegation? This latter example also refutes the common claim that the use of functionalist analyses is inherently conservative because it seems to underwrite the optimality of the status quo. Social critics also use functional methodology when they explain the presence of an unfair social practice by asking *Cui bono?* Who benefits most from the practice, and who would have the most to lose if it were discontinued? Thus, although the original explicit functionalism of anthropologists such as Malinowski (see Jarvie 1984) and sociologists such as Talcott Parsons (see Merton 1973) has been abandoned, this general type of analysis continues to be widely employed in more informal accounts of social phenomena.

Cultural Materialism

In the functionalist approach, institutions are seen as serving a wide variety of social needs. In what Marvin Harris (2001) has called the cultural materialist approach, the focus of the analysis is on much more carefully defined material requirements, such as the need for dietary fats and protein. Harris also systematically looks for data to support or refute his analyses. Many of his cases offer explicit alternatives to explanations in terms of vague needs such as social cohesion. For example, he proposed that the taboo on eating beef in traditional Indian culture is held in place by straightforward material considerations, not religious needs. He then conducted a census of infant mortality among male calves in various parts of India and found that in areas where oxen were used primarily as draft animals, more male calves than females survived. In areas where cattle were particularly valued for their milk, male calves were less likely to survive. Yet the villagers' support for the sacredness of all bovines was equally enthusiastic in both locales.

As was the case with explanations in terms of latent functions, cultural materialists often end up completely overriding the actors' own accounts of the variables that are influencing their behavior. In these postcolonial times when ethics codes often require the researcher to share results with the people studied, it becomes increasingly difficult to ignore *ab initio* the participants' own commentaries on their cultural practices. By limiting their explanatory factors to properties of the infrastructure, cultural materialism gains in specificity, but it becomes increasingly implausible when the explananda are far removed from basic needs such as food and shelter.

Interpretive Methodologies

All of the approaches sketched above search for modes of explanation that bear strong similarities to those employed in the natural sciences. Cultural materialists talk un–self-consciously about causes. Functional explanations are very common in biology (see Wright 1976 for an early account of functional explanations that uses both biological and cultural examples). Popper (1994a) compares explanations in terms of the rationality principle to those provided by Newtonian mechanics: One describes a more or less complex set of initial conditions (positions and momenta or beliefs and desires) and then "activates" the model to obtain the state of the system at a later time.

But there is also a strong tradition within philosophy of social science of arguing that an entirely different mode of understanding (*verstehen*) must be brought to bear on social phenomena. In his influential *The Idea of a Social Science,* Peter Winch (1958) claimed that the correct description of social phenomena, such as an individual saying a prayer or a mob attacking the citadel, has to include the actors' intentions and cultural categories. A wink is not a blink, a rosary is not a string of beads, and a riotous sports event is not an uprising of the proletariat. It is impossible to separate the description of an act from the variables that supposedly explain it. What is called for instead is a mode of quasi-empathetic understanding and interpretation, which is brought to bear on a richly detailed social characterization of the activity, sometimes called a *thick description.*

A widely cited example of this kind of interpretive activity is Clifford Geertz's (1973) essay in *The Interpretation of Cultures* on Balinese cockfights, in which he draws parallels between the structure of the participants' reactions and other parts of Balinese culture. One is not looking for the causes or rational roots or beneficial consequences of this activity. Rather, one is using a hermeneutical method common to those employed by a literary critic analyzing a poem in order to find the *meaning* of the practice. The end result is not explanation as is found in natural science; rather, it is an understanding of a particular activity in terms of the cultural system as a whole.

It is easy to agree that the phenomena that social scientists study are typically meaningful in a sense that goes beyond superficial physical descriptions. A cricket bat in the hands of a batsman is not just a strange sort of constrained bar pendulum. Furthermore, the proper understanding of such activities may well require one to consider their broad cultural ramifications. The anguish in Britain when armed conflict first broke out between two cricket-playing nations provides a window into the wider cultural significance of this sport. So, part of what the batsman is doing is participating in an activity intended to instill the gentlemanly virtues.

This seems worth mentioning. But just how thick should the descriptions be, and how can one evaluate interpretations that go well beyond common sense?

Commentators searching for symbolic meaning have remarked on the parallels between American football, where gaining yards, thereby decreasing the turf controlled by the other team is a central strategy, and the crucial project in American history of pushing the frontier steadily west and occupying native American lands. Others with a Freudian bent find great significance in the way that linesmen expose their bottoms to their own backfield, with the center even going so far as to let the quarterback grab the ball from between his legs. One need not take such examples of interpretation very seriously (although it would be interesting to understand why soccer has been so slow to gain popularity in the United States), and they are sometimes offered in a playful fashion. But they illustrate a basic dilemma for interpretive studies of social phenomena. If the interpretation is close to the meanings to which the actors readily assent, then it goes little beyond the commonsense understandings of the actors. (It might still be quite informative to someone outside the culture, of course.) But if the interpretation posits structures and symbols that are novel to the participants, then one may well wonder whether there is any empirical basis for the proposed "reading" of the cultural text.

Foundational Problems

The philosophical work described above responds to methodological issues arising within specific research programs found in current social science. There is another important strand of philosophy of social science that takes as its departure classic works in philosophy of mind and philosophy of language, such as Wittgenstein's (1957) *Philosophical Investigations*, and applies them to the analysis of the concepts that seem to underlie much of the thinking about society.

For example, it is often said that part of socialization is learning to follow both the explicit and the tacit rules of the culture—an especially vivid example is learning to speak a language (see Linguistics, Philosophy of). Or one speaks of the shared values, beliefs, attitudes, and norms of a society. It is difficult enough to explicate the notion of *following* a rule (as opposed to merely behaving in accordance with the rule), but when philosophers try to unpack the notion of a *tacit* rule and to speculate about how it might be instantiated in either the mind or the brain, these familiar and useful concepts start looking very suspect.

There is also a vast philosophical literature trying to make sense out of the idea of *shared* meanings. Can one speak meaningfully of *collective* intentionality? All of the standard problems in philosophy of mind are exacerbated when one starts theorizing about societies instead of separate individuals (see Consciousness; Intentionality).

As discussed above, there are ongoing disputes about whether the methods of social science differ fundamentally from those appropriate to the natural sciences. A closely related debate concerns the ontological unity of the two sorts of science. Are there such things as social facts that somehow exist above and beyond the properties of individuals? Proponents of methodological individualism call for explanations of all social phenomena in terms of the doings of individual people while admitting that recourse to large-scale social factors is necessary in the short run (see Methodological Individualism). Searle (1995), however, argues that both institutional facts, such as the value of money, and simple group actions, such as going for a walk together, can be analyzed only in terms of shared intentions. (He finds the roots of this kind of collective intentionality in animal behavior, such as that of a pack of dogs hunting together.) Philosophical arguments in social science about reductionism, emergence, and supervenience (Rosenberg 1995) often parallel discussions about the relationship between biology and chemistry (see Reductionism). McIntyre (1996) presents a thoroughgoing naturalist approach to social science.

New Directions

Although the development of science sometimes generates new philosophical enigmas, scientific research can also shed new light on perennial philosophical puzzles. In a recent book subtitled *Social Theory after Cognitive Science,* Stephen Turner (2002) argues that the connectionist account of how learning affects the brain provides a clearer way to think about "shared" social practices. If each person's path to linguistic competence, for example, is unique, then there are no transcendent rules that are somehow embodied in individuals within that society. Other models, such as computational theories of mind, would have different implications. The important moral is that social theory needs in general to be aware of results from cognitive science.

Biology continues to be a valuable correlative discipline for scientists interested in explaining human behavior. Although the early claims coming from sociobiology faced severe criticism for a variety of reasons, more recent studies in what is now called evolutionary psychology are directly applicable (see Evolutionary Psychology). For example, the biological notion of *reciprocal altruism* helps clarify how human responses to the Prisoner's Dilemma vary with the situation. Comparative studies in animal behavior are also indispensable for understanding cross-cultural similarities and differences in, for example, the expression of human emotions.

Since scientific inquiry is an excellent example of highly organized social behavior, social scientists and historians have sometimes engaged in empirical and interpretive studies of science itself. Robert Merton's (1973) work on the norms of science launched the field of sociology of science. Philip Kitcher's (1993) *The Advancement of Science*, which presents economic models of the organization of cognitive labor and the attribution of scientific authority, can be viewed either as a piece of theoretical social science or as a part of naturalistic philosophy of science. As epistemologists become more interested in models of group rationality and societies of knowers, such work in sociology of science becomes increasingly relevant to philosophical models of the development of science.

There has also been a resurgence of interest in the sociology of knowledge tradition associated with Karl Mannheim (see Couvalis 1997). Feminist critics of science claim that gender ideology influences the content of science as well as its values, methods, and organization (for a critical survey, see Pinnick, Koertge, and Almeder 2003). Proponents of the so-called Strong Program look for the causal influence of interest groups on the content of science (see Social Constructivism). Scholars working in the new field of *science, technology, and society* (STS) try to combine interpretive accounts of scientific practice with political criticisms of both the applications of scientific findings and the directions of scientific inquiry. They find that traditional scientific values have produced a science that serves powerful elites at the expense of ordinary people. Some STS proponents argue that the aim of science should be emancipation, not explanation, and call for a revolution in both the methodology and the value structure of science.

Such dramatic proposals for both science policy and science education have led to a public debate sometimes called the "Science Wars," in which scientists and philosophers have objected vehemently to the portrayal of science coming out of recent work in the sociology of knowledge tradition (see Koertge 1998). They argue that the move from the underdetermination of theories by evidence to relativism is philosophically unsound (see Underdetermination of Theories). They also fault many of the case studies coming out of STS on the grounds that there are typically good empirical or theoretical reasons for the acceptance of scientific results that are ignored by those who look primarily for ideological factors or narrow professional interests. The defenders of traditional scientific methods warn that privileging political considerations over empirical considerations in the evaluation of scientific claims would seriously undermine both the explanatory power and the pragmatic usefulness of science.

The field of research ethics is an area of growing interest and one in which philosophers of social science might appropriately play a more important role (see Popper 1994b.) Professional organizations of social scientists took the lead several decades ago in adopting codes of ethics for their members. Issues of informed consent, the humane treatment of animal subjects, and the responsibility of researchers to share their results with human subjects have been brought to the forefront, and ethical procedures for dealing with them have been institutionalized by scientific societies. Various countries have also put in place governmental review boards to regulate research. As the prolonged debate over cloning illustrates, philosophers can be of service by providing scholarly analyses of the ethical and moral issues that have an impact on scientific research.

In conclusion, the problems studied by philosophers of social science are both intellectually challenging and socially relevant. New scientific developments, both in the social sciences as traditionally defined and in new areas such as cognitive science, lead to new issues and new perspectives on old problems. There are increasingly sophisticated historical analyses of the contributions of the founders of modern social science, such as Marx, Durkheim, and Weber (Gordon 1991). The growth of interest in issues arising out of the individual sciences has led to new specialized societies and journals.

Noretta Koertge

References

Couvalis, George (1997), *The Philosophy of Science: Science and Objectivity*. Thousand Oaks, CA: Sage Publications.

Elster, Jon (2000), *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge and New York: Cambridge University Press.

Gasper, Philip (1991), "The Philosophy of Social Science," in Richard Boyd et al. (eds.), *The Philosophy of Science*. Cambridge, MA: MIT Press, 713–718.

Geertz, C. (1973), *The Interpretation of Cultures*. New York: Basic Books.

Gordon, H. Scott (1991), *The History and Philosophy of Social Science*. London: Routledge.

Harris, Marvin (2001), *Cultural Materialism: The Struggle for a Science of Culture*. Walnut Creek, CA : AltaMira Press.

Jarvie, Ian C. (1984), *Rationality and Relativism: In Search of a Philosophy and History of Anthropology*. London: Routledge and Kegan Paul.

Kitcher, P. (1993), *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford: Oxford University Press.

Koertge, Noretta (ed.) (1998), *A House Built on Sand: Exposing Postmodernist Myths about Science*. New York: Oxford University Press.

Little, Daniel (1991), *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Boulder, CO: Westview Press.

Martin, Michael, and McIntyre, Lee C. (eds.) (1994), *Readings in the Philosophy of Social Science*. Cambridge, MA: MIT Press.

McIntyre, Lee C. (1996), *Laws and Explanation in the Social Sciences: Defending a Science of Human Behavior*. Boulder, CO: Westview Press.

Merton, Robert K. (1973), *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.

Pinnick, Cassandra L., Noretta Koertge, and Robert F. Almeder (eds.) (2003), *Scrutinizing Feminist Epistemology: An Examination of Gender in Science*. New Brunswick, NJ: Rutgers University Press.

Popper, K. R. (1957), *The Poverty of Historicism*. Boston: Beacon Press.

——— (1994a), "Models, Instruments, and Truth: The Status of the Rationality Principle in the Social Sciences," in Mark Notturno (ed.), *The Myth of the Framework: In Defence of Science and Rationality*. London: Routledge, 154–184.

——— (1994b), "The Moral Responsibility of the Scientist," in Mark Notturno (ed.), *The Myth of the Framework: In Defence of Science and Rationality*. London: Routledge, 121–129.

Rosenberg, Alexander (1995), *Philosophy of Social Science* (2nd ed.). Boulder, CO: Westview Press.

Salmon, Merrilee H. (1992), "Philosophy of the Social Sciences," in Salmon et al. (eds.), *Introduction to the Philosophy of Science*. Englewood Cliffs, NJ: Prentice Hall, 404–425.

Searle, John R. (1995), *The Construction of Social Reality*. New York: Free Press.

Turner, Stephen P.(2002), *Brains, Practices, Relativism: Social Theory after Cognitive Science*. Chicago: University of Chicago Press.

Winch, P. (1958), *The Idea of a Social Science*. London: Routledge and Kegan Paul.

Wittgenstein, L. (1957), *Philosophical Investigations*. London: Blackwell.

Wright, Larry (1976), *Teleological Explanations: An Etiological Analysis of Goals and Functions*. Berkeley and Los Angeles: University of California Press.

See also **Behaviorism; Economics, Philosophy of; Game Theory; Methodological Individualism; Psychology, Philosophy of; Reductionism**

SOCIOBIOLOGY

See **Adaptation and Adaptationism; Altruism; Evolutionary Psychology**

SPACE-TIME

Space-time is the set of all actual and possible locations, in space and time, of all actual and possible events. The notion was formally introduced by Minkowski (1908 and 1909), who noted that "no one has ever observed a time except at a place, nor a place except at a time." In other words, the occurrence of an event is not completely "located" except when both the place and the time are specified. Therefore, locations of all possible and actual events necessarily form a four-dimensional set, with three spatial dimensions and one temporal. Against this background, the history of any individual object—idealized, for simplicity, as a point-particle—may be represented as a trajectory in space-time, or a "worldline" connecting all the spatiotemporal locations that the object successively occupies. If space and time are assumed to be continuous, it is obvious that space-time can be represented as a model of the real-number manifold $\mathfrak{R}^4$, so that each point is represented by an ordered quadruple of real numbers, namely, its three spatial and one temporal coordinates.

Newtonian Space-Time

The notion of space-time is thus a very simple and natural one that accords well with ordinary intuitions about events that occur in space and time. One might even wonder that it did not develop earlier than it did. It is true that Kant took a step in this direction by noting that "if time is represented by a straight line produced to infinity, and simultaneous things at any point of time by lines drawn perpendicular to it, the surface so generated would represent the *phenomenal world* in respect both of substance and of accidents" (Kant [1770] 1911, 401n). But it was not until Minkowski's work (1909), which represented Einstein's (1905) special theory of relativity as the theory of a four-dimensional "world-structure," that the notion became an essential part of physics.

The reason for this is fairly straightforward: Until Einstein's special theory of relativity, both science and common sense viewed space-time events as standing in *relations* that effected a clear conceptual separation of space from time. The most important of these relations are *sameness* of time and place. That events could be identified as having happened at different spatial locations at the same time, or at the same spatial location at different times, seems intuitively completely obvious. Regarding sameness of time, there would appear to be direct sensory evidence of simultaneous events, or at least of events whose temporal separation is too small to discriminate. This is because, relative to the distances separating the events of ordinary experience, the signals carrying information about these events are generally extremely fast; in the case of visual perception through the propagation of light, the velocity is immeasurably great, and the time delay between the occurrence of an event and the perception of it could always be neglected. In the context of Newtonian mechanics, it was natural to think that even this delay was merely a practical problem, surmountable in principle by means of signals that may travel arbitrarily fast. In the first place, the laws of motion state that the acceleration produced in a given body by a given force is independent of the body's initial velocity, which entails that velocities may be increased arbitrarily; in the second place, universal gravitation appeared to be an actual instance of instantaneous action at a distance, or an effect that is simultaneous with its cause. Therefore, the Newtonian picture of physical causation divides space-time into space and time in a way that seems to justify the intuitive picture: For any moment, there is an objective partition of all events into those that have already happened, those that will happen, and those that are now happening; all events in the past are capable of influencing the present, but not vice versa; future events may be influenced by events in the present and the past, but not vice versa; present events may influence other present events, but only in case there are instantaneously propagating signals. In other words, Newtonian physics seems to confirm an intuitive trichotomy: For any two events *A* and *B*, either *A* precedes *B*, *B* precedes *A*, or *A* and *B* happened at the same time (see Figure 1).

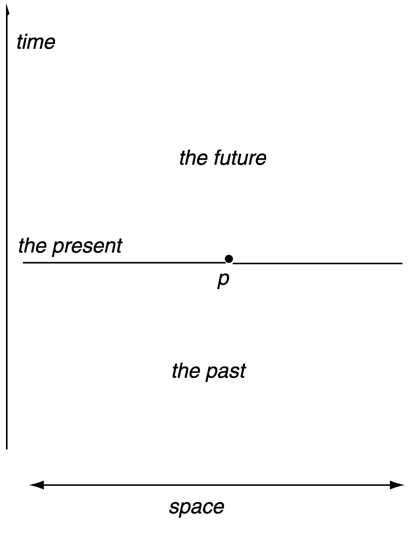


Fig. 1. The Newtonian view of simultaneity. The "present" represents an objective division among all events in space-time, separating "what is happening now" from what will happen in the future and what has happened in the past. It is a causal division, because it thereby separates what may still be influenced by events in "the present" from those that are already determined.

Thus, the notion of simultaneity may be thought of as decomposing four-dimensional space-time into an infinity of three-dimensional "slices" (spatial "hypersurfaces"), one corresponding to each moment of time and representing "space" at that moment. This "foliation" makes it possible to view space-time not as an integrated geometrical structure, but as a succession of states of space. Absolute simultaneity thus effects a "projection" of space-time onto time (i.e., a function that identifies three-dimensional subspaces of space-time with the same temporal coordinate), and the simultaneity slices are the fibers of the projection (see Figure 2).

Regarding sameness of place, Newtonian physics corrects the intuitive picture and shows it to be without physical foundation—the effect of mistaking a local perspective for an objective point of view. It is apparent to ordinary perception that a purely spatial change can always be canceled by a movement of the perceiver. That is, unlike a change in temporal perspective, a change in spatial perspective can be done and undone arbitrarily, so that for every spatial displacement of an object, there is an "inverse" displacement that restores the object to "the same place." Even though such motions necessarily take some time to accomplish, their possibility is central to the conceptual separation of space from time. The question is whether such combinations of displacements can ever be said to return an object to the same part of space, as opposed to merely restoring it to its former material surroundings. If the Earth could be assumed to be at rest, or if the laws of motion provided some dynamical measure of velocity, it could be inferred that some combination of displacements can lead from a given spatial point back to the same point at a later time. Analogously to the case of simultaneity, there would exist what amounts to a projection of space-time onto space, identifying one-dimensional subspaces each with the same spatial coordinates (so that the fibers of the projection would be the worldlines of particles at rest) (see Figure 3).

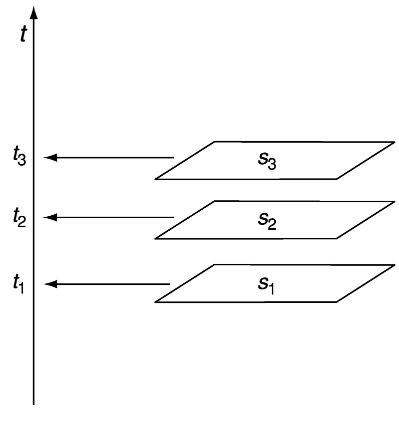


Fig. 2. Absolute simultaneity. By identifying events at different places at the same time, the relation of simultaneity is a "projection" of space-time onto time, identifying each hypersurface *s* as "all of space" at a corresponding time *t*.

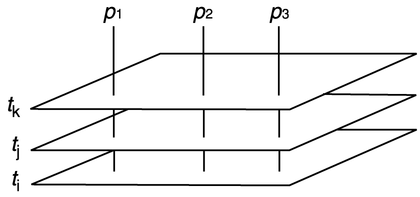


Fig. 3. Absolute space: If it were possible to discern the state of absolute rest, there would be a projection of space-time onto space, identifying certain worldlines as the histories of particular places in space. Each point *p* would represent a particular point in space at all times *t*.

As Galileo first argued convincingly, however, experience is incapable of discriminating rest from quasi-uniform motion (Galileo [1632] 1996). This principle, now known as the principle of Galilean relativity, was established in Newtonian physics as a corollary to the laws of motion. If this were not so, it might be supposed that the projection of space-time onto space had some objective empirical foundation. In that case, space and time could be thoroughly disentangled and space-time could be dispensed with altogether. But this is not possible, according to Newtonian physics, and the distinguished state of motion is uniform unaccelerated motion rather than rest. Therefore, space-time must be viewed as a four-dimensional affine space, that is, a space with a distinguished class of straight lines and a well-defined notion of parallelism: Straight lines (geodesics) represent the trajectories of bodies in uniform motion, not subject to forces; and parallel lines represent trajectories of particles that are relatively at rest (cf. Stein 1967; Ehlers 1973).

So, the relativity principle means that any family of parallel space-time geodesics may be regarded as being at rest, while none can be distinguished as absolutely at rest. Any family of parallel geodesics defines an "inertial frame," and all inertial frames are equivalent for the description of physical phenomena; the equivalence of inertial frames (the Galilean relativity principle) expresses the arbitrariness of the projection of space-time onto space, while the agreement of all inertial frames on which events are simultaneous expresses the

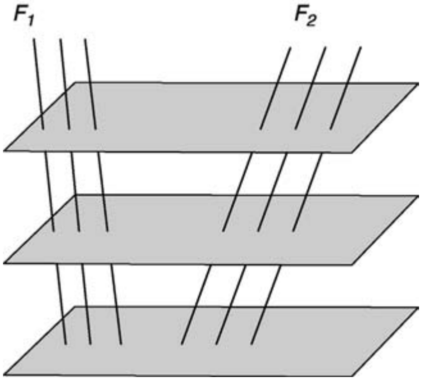


Fig. 4. Galilean relativity: F_1 and F_2 are distinct families of parallel space-time geodesics, that is, distinct inertial frames. They are in uniform motion relative to one another, but they agree on simultaneity, and therefore on the division of space-time into spatial hypersurfaces.

nonarbitrariness of the projection of space-time onto time (see Figure 4). The spatial and temporal coordinates of any two inertial frames, x, y, z, t and x', y', z', t', are related by the "Galilean transformations":

$$\begin{aligned} x' &= x - vt \\ y' &= y \\ z' &= z \\ t' &= t \end{aligned}$$

These relations express the absolute character of time, and the relative character of velocity, in a straightforward way.

The concept of space-time ought to place the philosophical controversies about space and time attendant on the birth of Newtonian physics in a clear perspective. Some of Newton's contemporaries were justifiably troubled by his concepts of absolute space and time, which he viewed as integral to his dynamical theory. From a modern perspective, in Newton's own construal of the term "absolute," his physics requires only "absolute space-time" (the aforementioned affine space) and absolute time, while absolute space involves the entirely superfluous assumption that some particular inertial frame is truly at rest. The space-time structure thus expresses only what is implicit in the dynamical assumptions shared by Newton, Huygens, Leibniz, and their contemporaries. Deeper philosophical analyses of space and time would, in later centuries, follow the critical analysis of only those assumptions. Thus it took profound transformations in the theories of electrodynamics and gravitation to displace this picture of space-time from its central place in physics.

Space-Time in Special Relativity

From the spatiotemporal perspective, the projection of space-time onto time—absolute simultaneity—might appear to be just as much in need of justification as the projection onto space. Yet it passed almost without comment until the early twentieth century, even from those with a broad philosophical commitment to "relativity" and a professed philosophical aversion to "absolute" structure in general; for example, Leibniz (1716) and Mach (1883) objected to Newton's conceptions of absolute space and time but never questioned the possibility of identifying simultaneous events or of specifying the relative positions of things "at a moment." Before Einstein, believers in the relativity of motion were, in effect, presupposing a continuous space-time, decomposable into instantaneous Euclidean spaces on which the momentary relative positions of bodies were well defined—otherwise their very conception of "relative motion" would have made no sense; their skepticism concerned, in effect, only whether space-time had the Newtonian affine structure as well. That is, they questioned only the existence of distinguished trajectories in space-time—the inertial motions—not the distinguished foliation of space-time into spaces at each moment of time. This is hardly surprising, given the apparent clarity of the intuitive distinction between past, present, and future. Moreover, Newtonian physics, while making no distinction between motion and rest, seemed to provide, at least in principle, a basis for determining simultaneous events. With Einstein's (1905) critique of the Newtonian concept of simultaneity, the basis for the intuitive separation of time from space was undermined. If simultaneity is relative, so that different inertial frames determine different spatial hypersurfaces, then the projection of space-time onto time is as arbitrary as the projection onto space, and for the same kind of reason.

Einstein introduced the special theory of relativity by proposing to unite two seemingly incompatible principles: the "relativity postulate," that is, that physical laws do not distinguish uniform motion from rest; and the "light postulate," that is, that the speed of light is the same in every inertial frame. In Newtonian mechanics, obviously, no velocity could possibly be invariant, since velocity is essentially a relative quantity. In classical electrodynamics, velocity relative to the ether was assumed to be a privileged velocity, but, here again, it could not possibly be an invariant velocity, as the velocity measured by any observer must depend on the observer's own velocity relative to the ether. Yet when experiments were devised (e.g., by Michelson and Morley) to measure such differences in the velocity of light, their results were consistently null. For Einstein, these results were only one aspect of the general pattern in electrodynamics—that much theory, but no phenomena, appeared to depend on motion relative to the ether. But it was not so straightforward in the case of electrodynamics, as it had been in the case of mechanics, simply to eliminate the privileged frame of reference and accept the equivalence of inertial frames. For the reasons already noted, the idea of an invariant velocity seemed absurd. To eliminate the absurdity, and so to reconcile his two postulates, Einstein had to uncover the assumptions that had made them seem incompatible in the first place.

The key assumption was that of absolute simultaneity. The claim that observers in different states of inertial motion cannot agree on an invariant velocity, Einstein realized, rests on the assumption that they must agree on which events are simultaneous. Conversely, they can agree on the invariant velocity of light if they differ on which events are simultaneous. The criterion for simultaneity is provided by light signals: In Einstein's definition, two events are simultaneous if light signals from each event, traveling equal distances, reach the same observer at the same time. If a signal is sent from *A* to *B* and reflected back to *A*, the reflection is defined to be simultaneous with that moment of time halfway between the transmission and the reception at *A* (see Figure 5). Since the velocity of light, like all electromagnetic phenomena, is apparently the same in every inertial frame, Einstein's criterion seems to have a sound basis in physical law. Moreover, it agrees with the most familiar intuitive criterion of simultaneity. Using the same criterion of simultaneity, however, observers in different states of inertial motion must determine different sets of events to be simultaneous. In short, simultaneity is relative. Frames of reference that agree on the velocity of light will be related not by the Galilean transformations, but by the Lorentz transformations:

$$\begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - v^2/c^2}} \\ y' &= y \\ z' &= z \\ t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} \end{aligned}$$

Evidently these transformations preserve the velocity of light, but time and length will vary according to the relative velocity of the frames.

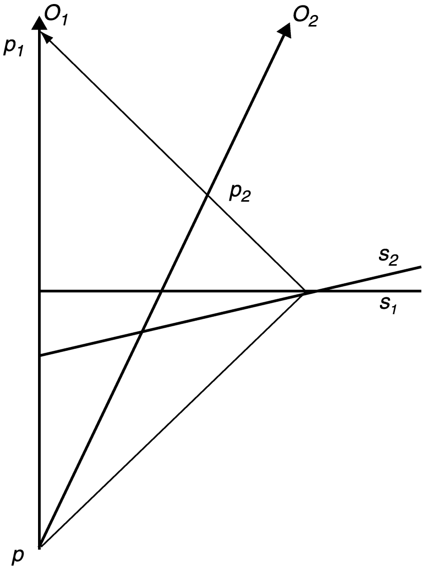


Fig. 5. The relativity of simultaneity: By Einstein's criterion, inertial observers and O_2 will disagree on which events are simultaneous. If a light signal is emitted from *p* (where their paths cross) and is reflected back to them, each will determine the time of reflection by halving the interval measured between the emission and the moment when the reflection is seen. For O_1 the interval will be from *p* to p_1; for O_2 the interval will be from *p* to p_2. Therefore for O_1 the events on the surface s_1 will be simultaneous, while for O_2 the events on the surface s_2 will be simultaneous.

The relativity of simultaneity is such a dramatic implication that one would seem to have good reason to reject the principle that implies it. The invariance of the velocity of light, before Einstein, was regarded as a mere appearance that must be explained by the interactions of bodies with the ether. Lorentz regarded the Lorentz transformations as expressing the contraction of measuring rods and the slowing of clocks in proportion to their velocities with respect to the ether. It could be argued, then, that there is a frame-independent fact of the matter about which events are simultaneous, but that light signals cannot determine it, and therefore they make a poor criterion. Einstein forestalled this argument by his conceptual analysis of simultaneity and its role in physical measurements. The very construction of a frame of reference must begin with some criterion for the measurement of time, and this is impossible unless one understands what one means by "simultaneous." Light signals were typically used for practical reasons; Einstein argued that their invariance properties make them uniquely suitable as the criterion of simultaneity. If this criterion is to be rejected, because of its surprising consequences,

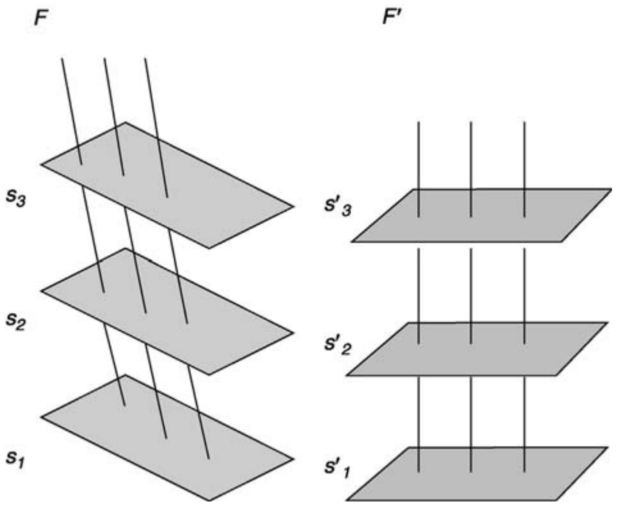


Fig. 6. Inertial frames in special relativity: Inertial frames F and F' disagree on simultaneity, so their frames correspond to two distinct decompositions of space-time into spatial hypersurfaces, s and s'.

there must be some other way of explicating the meaning of simultaneity, sufficient to make the concept applicable to the empirical world (cf. Einstein 1917, chap. 8). That is what neither Lorentz nor anyone else had been able to provide.

Einstein's theory leads directly to a new conception of space-time. The relativity of simultaneity means that for the worldline of each inertial observer (or family of parallel such worldlines), there is a distinct decomposition of space-time into space and time (see Figure 6). Relative to any given choice of a foliation, the spatial distance between any two points is given by Euclidean geometry, which implies that length is relative to the choice of inertial frame. Similarly, the relativity of simultaneity implies the relativity of time intervals. The invariant physical quantity is an inherently *spatio-temporal* quantity, the speed of light in vacuo, c. This follows simply from applying the familiar formula rate × time = distance: Letting c = the speed of light; x, y, z the spatial coordinates; and t the time, the formula becomes:

$$ct = \sqrt{x^2 + y^2 + z^2}$$

or, squaring both sides:

$$c^2t^2 = x^2 + y^2 + z^2.$$

In a relatively moving frame with coordinates t', x', y', z', the formula becomes

$$c^2t'^2 = x'^2 + y'^2 + z'^2.$$

Then the invariance of the speed of light can evidently be represented by the expression

$$c^2t^2 - x^2 - y^2 - z^2 = c^2t'^2 - x'^2 - y'^2 - z'^2.$$

(Note the purely conventional choice to represent the temporal component as positive and the spatial as negative; the objective fact is only that they are opposite in sign.)

This aspect of Einstein's theory suggested to Minkowski, schooled in Klein's (1872) view of geometry as a theory of structure and automorphisms (structure-preserving maps), that special relativity may be understood as a theory of space-time geometry. The physical invariant (in units where $c = 1$)

$$t^2 - x^2 - y^2 - z^2$$

has a natural interpretation as the metrical invariant of a four-dimensional pseudo-Euclidean space, that is, in a four-dimensional flat affine space with an indefinite metric. In fact, the metric appears in pseudo-Euclidean guise as

$$ds^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

through the identification $x_1 = ict$. This underscores the analogy between the Lorentz transformations between inertial frames, as the isometries of the four-dimensional space-time, and the group of Euclidean isometries that preserve the Pythagorean metric. The slices of simultaneous events for any inertial observer become the hyperplanes that are orthogonal, as determined by Minkowski's inner product, to the corresponding geodesic worldline; Einstein's criterion of simultaneity naturally corresponds to the orthogonality criterion in Minkowski's geometry. It should be noted, however, that Einstein's criterion is based on the *stipulation* that the speed of light is the same in every direction; the only empirically measurable quantity is the total travel time of light in two directions—for example, to a mirror and back. That the time in each direction is exactly one-half of the total has, therefore, been thought to be a matter of arbitrary choice. This would imply, over and above the relativity of simultaneity, a kind of conventionality of simultaneity (see Conventionalism).

The peculiar role of light propagation, as a finite invariant velocity, determines a distinctive causal structure for Minkowski space-time (see Causality). From any given point, the possible paths of light signals in all directions (or arriving from all directions) form a hypersurface, the light cone, that divides space-time analogously to the way in which "the present moment" divided Newtonian space-time; since the velocity of light is an invariant

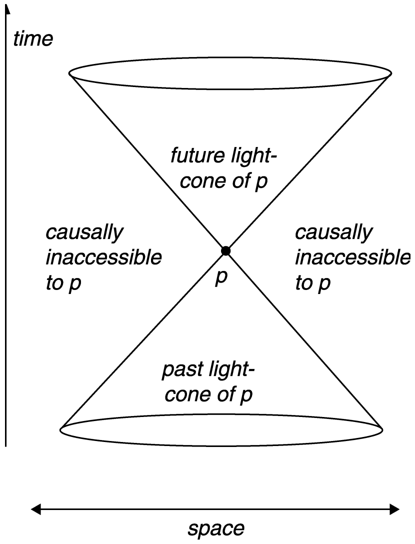


Fig. 7. The causal structure of Minkowski space-time: Since the speed of light is an invariant limit on the velocity of causal propagation, the regions that are causally accessible to point p are the past and future light cones at p.

limiting velocity, all causal propagation is confined to the surface (for electromagnetic signals) and the interior (for all slower motions, i.e., of massive particles) of the cone. In other words, the light cone divides space-time around a given point into causally accessible and inaccessible regions (see Figure 7). This clarifies the status of Newtonian space-time as a limiting case of relativistic space-time, since, in the limit as c is imagined to become infinite, the past and future light cones at a point collapse together into the plane of absolute simultaneity passing through that point. Mathematically, the cone structure corresponds to a division of all vectors at a point according to their lengths as measured by the Minkowski metric: "Timelike" vectors inside the light cone have positive length (assuming the sign convention adopted above), "spacelike" vectors outside the light cone have negative length, and "null" or "lightlike" vectors on the light cone have zero length. Curves whose tangents are everywhere timelike (spacelike, lightlike) are called timelike (spacelike, null) curves.

The difference between the Newtonian and relativistic space-times can now be easily summarized. In both cases, space-time is a flat four-dimensional affine space on the real numbers, where the distinguished geodesic trajectories represent the possible paths of free particles, and any family of parallel geodesics determines a global inertial frame. The differences concern metrical structure. Newtonian space-time has a preferred foliation into spatial hypersurfaces, with separate temporal and spatial metrics. Thus any two space-time points have a temporal separation—that is, the interval between the hypersurfaces in which they lie; but spatial separation is defined only for points that lie in the same spatial slice. In the relativistic setting, in contrast, there is only a single space-time metric, defining the space-time interval between any two points. The time elapsed between two events depends—much as the spatial separation of two points in space—on the path taken from one to another; instead of an invariant absolute time interval, there is only a measure of the "proper time" along a trajectory that connects the events.

The Minkowski metric has several counterintuitive features. One is that the trajectories of light rays have zero "length" and are orthogonal to themselves; another is that among timelike curves, the straight line represents the *greatest* distance between two points. (The latter is the root of the "twin paradox": Two observers can separate, and rejoin at some later time, only to find that more time has passed for one than for the other. The reason is that they have followed two different trajectories, and the proper time along one differs from the proper time along the other. So it is no more paradoxical than the fact that two people can take different routes from one place on Earth to another, and find that they have not traveled the same distance.) But the Minkowski metric also integrates the structures of space-time in a remarkably simple way. In the Newtonian case, the dynamical structure is completely extraneous to the kinematical structure of basic spatial and temporal measurement. This raised the question, for philosophers such as Ernest Mach, whether the dynamical space-time structures can be dispensed with in a theory that admits only changing "spatial relations" (see Mach, Ernest). In Minkowski space-time, however, the dynamical structure is inseparable from the kinematical structure, as the geometry of light propagation is the basis for all spatial and temporal measurement. And by Einstein's epistemological analysis of simultaneity, the classical notion of an immediately given structure of spatial relations, prior to any dynamical space-time structure, is revealed to be illusory.

Space-Time in General Relativity

The transition from special to general relativity is, in some respects, less dramatic than the one just outlined from Newtonian space-time to special relativity. For general relativity does not overturn special relativity's basic conception of space-time structure, but, rather, reveals it to be only the local structure of space-time, which, on larger scales, can become inhomogeneous. The dramatic difference is that the inhomogeneity of space-time is a function of the distribution of matter: Instead of Minkowski's metric as the global structure of space-time, there is Einstein's equation, which relates the geometry of space-time to the distribution of mass and energy. This comes about as a result of the connection that Einstein perceived between inertia and gravity: The gravitational field is no longer represented as a field disturbing the motions of particles in space-time, but as the curvature of space-time itself. In other words, a particle falling toward a large mass is not being forced by gravity to deviate from a space-time geodesic, as Newton's theory supposes; rather, it is following a geodesic in a space-time that is curved by the presence of that mass. Therefore, Newton's law of gravitation, according to which the gravitational field depends on the masses of bodies and their relative distances, was replaced by Einstein's field equation, according to which the geometrical structure of space-time depends on the distribution of matter and energy. The Newtonian law makes no mention of time; it implies that any variation in mass distribution immediately alters the gravitational field to arbitrary distances. By Einstein's equation, gravitational effects are "ripples" in the structure of space-time that propagate like waves at the speed of light. In short, where Newton had represented gravity as force acting instantaneously within a fixed space-time background, the general theory of relativity incorporates both gravity and the geometry of space-time in a single local field theory.

The motivation for this view comes from Einstein's analysis of the equivalence principle, that is, the principle of the equivalence of gravitational and inertial mass. The empirical significance of the principle is that the trajectory of a body in a gravitational field is independent of its mass and composition—a fact already known to Galileo and tested fairly precisely by Newton through the timings of pendulums of many different masses and materials. And as Newton also recognized, this implies that a system of bodies that is freely falling in a quasi-uniform gravitational field (e.g., Jupiter and its moons falling toward the sun) will be locally indistinguishable from one in uniform motion. Thus, free fall is locally indistinguishable from inertial motion. It also implies, however, that a frame of reference at rest in a gravitational field with acceleration *g* is indistinguishable from a frame that is uniformly accelerating with acceleration—*g*, since the weights of bodies in the first frame would not be distinguishable from their inertial resistance to acceleration in the second. Indistinguishability of inertial motion from gravitational free fall would undermine the entire Newtonian distinction between gravity and inertia, and therefore between the background structure of space-time with its privileged trajectories and the gravitational field that is supposed to force particles to deviate from those trajectories. Einstein boldly proposed, as subsequent experiments have shown to high precision, that the equivalence principle applies to all forces of nature. From here, the first step toward general relativity is to identify free-fall trajectories as the geodesics of space-time. This is analogous to the move Einstein had made with special relativity: As light had been held to be the only invariant criterion for simultaneity, free fall provided the only invariant instantiation of the idea of inertial motion; the equivalence principle implies that the geodesics of flat space-time are as inaccessible to observation as absolute simultaneity had proved to be. The concept of inertial frame is replaced by the concept of "local" inertial frame, that is, the frame defined in the neighborhood of a freely falling particle.

The second step is to acknowledge that the geodesics, thus identified, behave in ways that are characteristic of the geodesics of curved spaces. Unlike the geodesics of Newtonian or Minkowski space-time, these geodesics exhibit relative accelerations. This is because, relative to a given freely falling particle, another freely falling particle will generally not be moving uniformly; but this means that relative to one local inertial frame, another local inertial frame is in general relatively accelerated. In Newtonian or Minkowski space-time, flatness implied that any given inertial frame could be extended to a global inertial frame with respect to which all inertial trajectories remained inertial. In general relativity, however, this global extension is impossible, because of the relative accelerations of local inertial frames. In fact, these relative accelerations provide a measure of the curvature, just as, for example, the convergence of geodesics toward the poles provides a measure of the curvature of the Earth. Similarly, the impossibility of imposing a global inertial frame is analogous to the impossibility of imposing a single plane coordinate system on the surface of the Earth. Thus, instead of forcing bodies to deviate from geodesic motion, the presence of mass forces geodesics into non-Euclidean relations. In short, mass curves space-time.

Einstein initially thought that general relativity would establish a certain philosophical point about the "general relativity of all motion," eliminating Newtonian ideas about the objective character of space, time, and inertia. In particular, it would rid physics of the idea that there are privileged states of motion. Where Newtonian mechanics and special relativity had each identified an equivalence class of privileged frames of reference—the inertial frames—general relativity would place all frames of reference on an equal footing. In this manner space and time would lose "the last trace of physical objectivity" (Einstein 1916; Schlick 1917). This thought proved to be somewhat misleading. As was noted above, the notion of inertial frame was indeed fatally undermined—not because of the general relativity of motion, but because of the identity of inertia and gravity. There remained a privileged state of motion, the freely falling trajectory; what had to be abandoned was the privileged *frame*—again, because the local frame of a falling observer could no longer be extended. So it was the nonuniformity of space-time that made the inertial frame untenable, not the elimination of privileged states of motion. Questions about motion asked by Newton were still meaningful: Does the Earth rotate? Does the Earth revolve around the sun, or the sun around the Earth? But the answers were no longer supposed to depend on the background structure of flat space-time; instead, they depended on the relations between the geodesics of curved space-time and the distribution of matter and energy.

In retrospect, it was perhaps unreasonable to expect a "general relativization" from a theory in which the notion of a space-time geodesic plays such a fundamental role. Already Newton had pointed out that a dynamics that appeals to the notion of a distinguished trajectory postulates, *ipso facto*, some objective spatiotemporal structure. Nonetheless, the most profound philosophical consequences follow upon the transformation of space-time structure from a fixed background to something that is in dynamic interaction with its material contents. An especially important implication—not least for its bearing on contemporary research on quantum theories of gravity—is that space-time has a *history*, that is, a dynamical evolution of its states. This notion makes no sense in the previous theories, which countenance only the evolution of matter and fields *within* space-time. But the implication is a necessary consequence of Einstein's equation combined with what is known of the relative motions of the galaxies and the large-scale distribution of matter and energy. Moreover, the dynamic aspect of space-time in general relativity, and its consequent nonuniformity, means that space-time can no longer be adequately understood from the point of view of symmetry groups and their invariants. Yet the mathematical structure of general relativity places its relation to the earlier theories in an illuminating perspective. Instead of beginning with $\mathfrak{R}^4$, defining coordinates on it in the obvious way (see Figure 8), then defining simple geometrical relations among its points, general relativity begins with an arbitrary differentiable manifold (cf. Bishop and Goldberg 1980), which is (very roughly) a topological space that is locally homeomorphic to $\mathfrak{R}^4$; it can therefore be coordinatized by "charts" that map it onto $\mathfrak{R}^4$ in any number of overlapping pieces (see Figure 9).

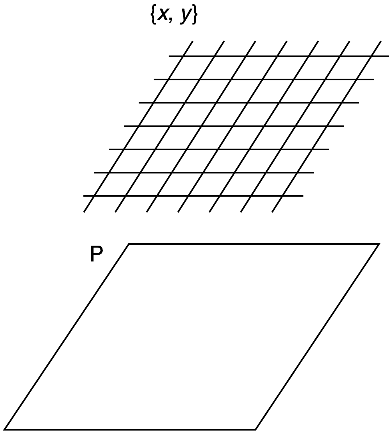


Fig. 8. Coordinates in a flat space: On the plane P, a rigid Cartesian coordinate system $\{x, y\}$ may be simply "set down" over the entire plane.

Then Newtonian or Minkowski space-time arises simply from introducing appropriate geometrical objects that impose symmetrical global structures upon it, making it globally homeomorphic to $\mathfrak{R}^4$.

In general relativity, however, space-time is not assumed to have any symmetries, but only the automorphisms of the manifold itself. Thus there is in principle no symmetry at all, but only general covariance, or covariance under the transformations that preserve the differentiable structure of the manifold. Additional geometric structure is not imposed in advance, but determined by mass distribution in accord with Einstein's equation. A symmetrical space-time geometry could arise in the case of a highly symmetrical distribution of matter, but that would be a contingent matter for empirical investigation. As Eddington (1920) pointed out, even though coordinate systems do not have any intrinsic physical meaning in general relativity, the possibility or impossibility of imposing coordinates reflects an important physical property, i.e., the nonuniformity of the curvature

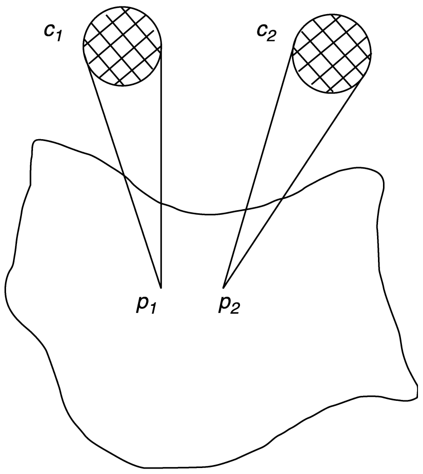


Fig. 9. Coordinates in curved space-time: In a non-uniform space, it will be impossible to lay a rigid Cartesian coordinate system over any finite region. The space must be coordinatized by overlapping "coordinate charts." For any two points p_1 and p_2 there will be local coordinate systems c_1 and c_2 that, in general, are "disoriented" relative to one another.

imposed by any nonsymmetrical distribution of mass.

Given these considerations, the general covariance of general relativity reflects a profound philosophical departure from earlier conceptions of space-time. At the same time, it is important to bear in mind what a contingent principle this is: It depends on the strict validity of the equivalence principle. An experimental violation of the equivalence principle—perhaps at higher levels of precision or energy than have been achieved up to now—would open the way for a distinction between inertial motion and free-fall motion, and so between an inertial frame and a freely falling frame. And that would open the way again for a separation of the gravitational field from the space-time background. This is only one of the respects in which present and future developments in physics promise to reopen questions about the nature of space-time that general relativity had seemed to resolve in such a compelling way. To Einstein, the unification of inertia and gravity in a single continuous field seemed to be a philosophically satisfactory as well as a physically convincing model for the treatment of all physical interactions, and this was the basis for his efforts toward a unified field theory. But those efforts failed, and since then the presumption has been that the next theory of space-time and gravity will be a quantum theory rather than a continuous field theory of the sort that Einstein was hoping for. The construction of such a theory may eventually lead to the replacement of general relativity's differentiable space-time manifold by some kind of discrete space. String theory holds out the possibility of a still more radical transformation, in which space-time as now conceived is not even a fundamental structure, but only the most obvious phenomenal expression of an underlying structure of as many as 26 dimensions. The radical changes that have occurred in the previous history of space-time theory, it should be recalled, were occasioned by radically new ideas about which physical processes reveal the structure of space and time: from the displacement of ordinary measuring rods, to electrodynamic propagation, to gravitational free fall. New approaches to this question are the likely result of further fundamental changes in physics.

Robert DiSalle

References

Bishop, R., and S. Goldberg (1980), *Tensor Analysis on Manifolds*. New York: Dover Publications.

Eddington, A. S. (1920), *Space, Time, and Gravitation: An Outline of General Relativity Theory*. Cambridge: Cambridge University Press.

Ehlers, J. (1973), "The Nature and Structure of Space-Time," in J. Mehra (ed.), *The Physicist's Conception of Nature*. Dordrecht, Holland: Reidel, 71–95.

Einstein, A. (1905), "Zur elektrodynamik bewegter Körper," *Annalen der Physik* 17: 891–921.

——— (1916), "Die Grundlage der allegemeinen Relativitätstheorie," *Annalen der Physik* 49: 769–822.

——— (1917), *Über die spezielle und die allgemeine Relativitätstheorie (Gemeinverständlich)*, 2nd ed. Braunschweig: Vieweg und Sohn.

Galileo ([1632] 1996). *Dialogo sopra I due massimi sistemi del mondo—Ptolemaico e Copernicano*. Milan: Oscar Mondadori.

Kant, I. ([1770] 1911), *De mundi sensibilis atque intelligibilis forma et principiis* [Kant's "Inaugural Dissertation"], in *Gesammelte Schriften, Akademie Ausgabe*, vol. 2. Berlin: Georg Reimer, 385–419.

Klein, F. (1872), *Vergleichende Betrachtungen über neuere geometrische Forschungen*. Erlangen: A. Duchert.

Leibniz, G.W. ([1716] 1960), "Correspondence with Samuel Clarke," in C. Gerhardt (ed.), *Die philosophischen Schriften von Gottfried Wilhelm Leibniz*. Hildesheim: Georg Olms, 345–440.

Mach, E. (1883), *Die Mechanik in ihrer Entwickelung, historisch-kritisch dargestellt*. Leipzig: Brockhaus.

Minkowski, H. (1908), "Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körper." *Nachrichten der königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physische Klasse*, 53–111.

——— (1909), "Raum und Zeit," *Physikalische Zeitschrift* 10: 104–111.

Newton, I. ([1726] 1999), *The Principia: Mathematical Principles of Natural Philosophy*. Translated by I. Bernard Cohen and Anne Whitman. Berkeley and Los Angeles: University of California Press.

Schlick, M. (1917), *Raum und Zeit in der gegenwärtigen Physik. Zur Einführung in das Verständnis der Relativitäts- und Gravitationstheorie*. Berlin: Springer.

Stein, H. (1967), "Newtonian Space-Time," *Texas Quarterly* 10: 174–200.

Suggested Readings

Geroch, R. (1978), *General Relativity from A to B*. Chicago: University of Chicago Press.

Taylor, E., and J. A. Wheeler (1978), *Space-Time Physics*. New York: Wiley.

Trautman, A. (1965), "Foundations and Current Problems of General Relativity," in A. Trautman, F. A. E. Pirani, and H. Bondi (eds.), *Lectures on General Relativity*. Brandeis 1964 Summer Institute on Theoretical Physics, vol. 1. Englewood Cliffs, NJ: Prentice-Hall.

Weyl, H. (1949), *Philosophy of Mathematics and Natural Science*. Translated by Olaf Helmer. Princeton, NJ: Princeton University Press.

See also **Causality; Conventionalism; Irreversibility; Time**

SPECIES

The most fundamental question with respect to *species* as this term functions in biology is whether or not a single level of organization exists, across all organisms, that counts as the species level. Is there a single definition of the concept 'species' that is equally applicable to all species, or must biologists resort to several different concepts to satisfy a variety of purposes? A second issue is the need for such a concept in the first place. What good are species? Why do biologists think they need a univocal concept of species that applies across all organisms?

Species Taxa

First and foremost, particular species (species taxa) must be distinguished from the species category. Species taxa are composed of particular organisms. *Homo sapiens* is the name of a particular species. It is made up of all human beings past, present, and future. Traditionally, the names of particular taxa such as *Homo sapiens* have been defined in terms of the phenotypic characteristics possessed by their constituent organisms. For example, people walk upright. They have a plantigrade foot, an opposable thumb, and a highly developed brain. For most of the history of biology, biologists have assumed that the names of all taxa, including species, can be defined in terms of phenotypic traits that are severally necessary and jointly sufficient. All human beings have each of the characteristics listed, and only human beings have this entire suite of characteristics. In addition, these characteristics incorporate the essence of the class. The conviction that all genuine class terms can be defined in this way is called 'typology' or 'essentialism'.

Typologists were well aware that species as characterized by biologists do not live up to this high standard of definition, but they were convinced that the only thing standing between them and essentially defined taxa was more work and greater knowledge. With the advent of evolutionary theory, this conviction became difficult to sustain. Before that, the phrase 'past, present, and future' did not represent much of an impediment to defining species taxa. Species come and species go, but not in ways that imply that they change gradually in the process. Once biologists came to realize that species can change through time, one species evolving into two or more, sometimes quite gradually, they were forced to abandon their typological assumptions. More study produced more, not fewer, problems. The phenotypic boundaries between species are sometimes quite fuzzy.

One modification of the essentialist position was to treat species taxa as polythetic; that is, the phenotypic characteristics used to define species covary only statistically. To repeat, the boundaries between species taxa in phenotypic space are fuzzy, not sharp. It is important to note that these boundaries exist in phenotypic space, not physical space. This can be made clearer by distinguishing between the range of a species and its distribution in character space. For example, a particular kind of tree might be limited to the Wabash Valley, which is formed by the Wabash River, flowing from Ohio through Indiana and along the border between Indiana and Illinois, to empty into the Ohio River. Systematists construct phenotypic space by mapping character distributions on a multidimensional graph. When they do, most organisms belonging to a particular species cluster in the center of the graph and become less prevalent as the conceptual boundaries of the species are approached. In short, *Homo sapiens* and the names of all other taxa are "cluster" concepts. They cluster not only at any one time but also through time, as species gradually change. The problem then becomes how to discern these fuzzy boundaries.

The foregoing considerations concern the distribution of ordinary phenotypic traits such as number of legs, color of hair, and the like. When characters—such as gene exchange and geographical distribution—that imply spatiotemporal relations are taken into account, the boundaries between species become real boundaries in space and time. The range of a species consists of the distribution of the organisms it comprises. Sometimes these spatiotemporal boundaries and the relations that define them are reasonably sharp, sometimes not—for example, there are degrees of reproductive isolation. However, the important change is the shift from ordinary phenotypic characters to relations that require spatiotemporal continuity and contiguity, a shift from character distributions to lineages. To be sure, character distributions are used to infer lineages, but the goal is to discern lineages. Taxa consist in chunks of these lineages.

The Species Category

All of the preceding discussion has concerned species as taxa, as groups of organisms defined in terms of their characteristics, but at bottom the problem of species concerns the category itself. How are we to define 'species' so that this term can fulfill the needs of biologists, in particular natural historians and evolutionary biologists? When one hears of the 'species problem,' this is the issue. Not all scientists, let alone all biologists, have the same concerns. Systematists who are interested in cataloging organismic diversity are willing to use rough-and-ready methods that make their job feasible. They did not in the past and do not at present have the workforce necessary to trace gene flow or establish ranges for all the taxa under investigation.

The labors of traditional systematists produced reasonably good classifications (Claridge, Dawah, and Wilson 1997). For many needs—Which species of mosquitoes can transmit malaria? How many species of fruit flies are there?—largely intuitive classifications were good enough. However, as time went by, systematists grew increasingly ambitious. They wanted their classifications to be more than just summaries of phenotypic variation. Some systematists wanted to retain the traditional goals of systematics but fulfill them more completely. Others wanted to expand the role of systematics in biology. They wanted to discern entities that functioned in natural processes, particularly the evolutionary process. As they viewed them, species were not just the basic units of classification, but also the basic units of evolution. Species were what evolved, speciated, and became extinct (Otte and Endler 1989).

The Concept of Phenetic Species

One group of systematists who proposed to change the orientation of biological systematics were the numerical pheneticists (Sneath and Sokal 1973). They are called "numerical" because they introduced mathematical techniques into systematics, especially the use of computers. Systematics is an ideal subject area for the use of computers. Millions of species exist, and each species has a different suite of characters. All these data can be maintained, readily accessible, in computer databases. Numerical pheneticists want their classifications to be constructed explicitly so that they can be tested by other workers. They want the classifications to be more objective, repeatable, and quantitative—in a word, more "scientific."

They are called "pheneticists" because they believe that organisms should be classified according to observable phenotypic characteristics. More important, these characteristics must be individuated without any theoretical input. One can discern the scales of fish and the scales of birds, and scales are scales. If classifications are to become objective, quantitative, and operational, then theory (or at least a technical approach such as evolutionary theory, and in the early stages of classification) has to be set aside as much as possible. As genetic information became more readily available, the notion of "phenetic" characters was expanded to include genetic characters as well (Sneath and Sokal 1973). However, the antipathy of numerical pheneticists toward scientific theories, especially evolutionary theory, remained.

The main problem with numerical phenetics was an embarrassment of riches. Early on, numerical pheneticists thought that something called "overall similarity" existed out there in nature. They assumed (or hoped) that alternative analytic procedures (e.g., parsimony, compatibility, neighbor joining, least-squares, likelihood) would zero in on the same classification or, if not, that one of these methods would prove superior to all others. But classifications reflecting overall similarity remained an elusive goal. Different clustering procedures produced different classifications even when applied to the same group of organisms, and there was no reason within numerical phenetics for preferring one procedure over another. Too many different distributions had an equal right to be described as overall similarity.

In addition, phenetic methods did not always mark the distinctions that biologists find absolutely essential. For example, in numerous cases, males and females of the same traditional species were placed in different taxa on the basis of character covariation alone; in response, the pheneticists allowed low-level biological distinctions to be introduced into systematics, even in the early stages of classification. No matter how different males and females may appear, and no matter how they differ genetically, they must be included in the same basal taxon. Allowing the results of phenetic methods to be overridden by issues such as sexual dimorphism seems to introduce the sort of theoretical speculation that pheneticists claimed to abhor. Granted that the distinction between male and female is very basic and frequently easy to discern, it is still more than just look, see, and cluster.

For pheneticists, is there a single level of phenetic covariation that is the same across all organisms? The answer is yes—and no. There are numerous levels of phenetic covariation. The trouble is that no level is any more "real" than any of the others. There is certainly no reason to choose any one level of phenetic covariation and call it the species level. Different clustering procedures produce different classifications, and there is no reason for choosing any one of these classifications over any of the others.

The Concept of Reproductive Isolation

Other definitions of the species category are more heavily laden with theory than the phenetic species concept is. The theories on which they are based concern both phylogeny and the evolutionary process. The most popular definition of the category during the past half century has been Ernest Mayr's ([1942] 1964) concept of biological species in terms of gene flow and reproductive isolation. In sexual species, boundaries of varying degrees of permeability exist in nature. According to Mayr:

> A species consists of a group of populations which replace each other geographically or ecologically and of which the neighboring ones intergrade or interbreed wherever they are in contact or are potentially capable of doing so (with one or more of the populations) in those cases where contact is prevented by geographical or ecological barriers. (120)

Or, in other words, species are groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups.

Mayr's definition is clearly committed to the evolutionary process. Gene flow or the cessation of gene flow is what really matters in sexual organisms. One common objection to this definition is that gene flow may not be as important in the evolutionary process as Mayr thinks. A second objection is that the definition can be very difficult to apply in particular cases. Gene flow is difficult to measure. One minor issue concerns ring species. A series of populations can be discerned in a circular course around some geographic barrier such as a large lake or mountain. Eventually, the termini of these courses meet. Population *A* can mate with population *B*, *B* can mate with population *C* around the geographic barrier until population *M* meets population *A*—and these two populations cannot mate! However, all these populations belong to the same species because mutations that occur anywhere in the ring can find their way into more distant populations.

"Potential interbreeding" is especially problematic. Species can be made up of numerous populations that at the moment are totally isolated from each other. Perhaps they formed contiguous sequences of populations in the past, as in the case of ring species, but they are now completely disjunct. No gene flow is taking place. However, later on they may come into contact again and reinstate gene flow. Cessation of gene flow alone is not enough to ensure total isolation. In order to belong to the same species, disjunct populations must retain the ability to mate successfully so that if they ever came into contact again, gene flow would resume. Discerning actual gene flow is difficult enough. Trying to estimate "potential" gene flow is even more difficult.

Aside from difficulties in application, the most fundamental objection to Mayr's definition is that gene flow really does not matter all that much in the evolutionary process. If this can be shown to be the case, Mayr's definition is in real trouble. In addition, it does not apply even in principle to organisms that never reproduce sexually. In his definition, these organisms do not form species. Hence, during the first half of life on Earth, no species existed. Evolution occurred, but in the absence of species.

In response to these and other objections, Mayr (1969, 26) modified his definition. He no longer refers to "potential interbreeding." Since that term is simply the negation of "reproductive isolation," he need not mention both in the definition. Mayr also limits his definition to species that are "nondimensional" in space and time: The definition applies only to sympatric and synchronous species, that is, those existing at the same place and time. However, nondimensionality is too strong a requirement. All species are extended to some extent in space and time. According to Mayr, species are not totally nondimensional, just minimally so (e.g., the chipmunks in my valley for the past few breeding seasons). These nondimensional species might well be organizable into more inclusive entities, but Mayr's definition applies literally only to nondimensional species per se. Although his definition remains popular among many taxonomists, other definitions have arisen to challenge it. Each turns on a different aspect of phylogeny or the evolutionary process.

The Concept of Evolutionary Species

As Mayr delineated them, species are time-slices of evolving lineages. G. G. Simpson (1961) presented a definition designed to portray species as evolving lineages: "An evolutionary species is a lineage (an ancestral-descendant sequence of populations) evolving separately from others and with its own unitary evolutionary role and tendencies" (153). The strength of Simpson's definition is that it is designed to delineate the basic lineages in phylogeny. Gene flow is one mechanism that can promote evolutionary unity, but it is not the only one. Simpson's definition also applies to asexual organisms. They too can form lineages that exhibit evolutionary unity, but without the aid of interbreeding. As might be expected, most objections to Simpson's definition concern difficulties in applying his notions of evolutionary roles and tendencies. It is not very operational. Discerning gene flow is difficult enough. Determining the influence of all the other mechanisms that promote evolutionary unity is even harder.

In sum, Mayr treats species as cross sections of evolving lineages, whereas Simpson treats them as temporally extended lineages. (See Wiley 1981 for an expanded version of Simpson's definition.)

Phylogenetics

In the past few decades a revolution has taken place in systematics, with the introduction of the views of Willi Hennig (1966). Both Mayr and Simpson wanted classifications to somehow reflect phylogeny, but for them these correlations were more than a little impressionistic. As Simpson stated numerous times, biological classification is as much an art as a science. Hennig proposed principles of classification that would be unequivocal. He decided that it was better to represent one relation clearly than numerous relations poorly. To do this he had to limit classification to reflect one and only one phylogenetic relationship:common ancestry as indicated by sister-group relations, for which he devised the notion of a cladogram (see Figure 1).

In the figure, the lines represent species. For example, if this cladogram is viewed as a phylogenetic tree, then an unnamed species splits into species *C* and a second, unnamed species. This second, unnamed species in turn splits into species *A* and *B*. Neither cladograms nor cladistic classifications can represent these relations. Instead, as a cladogram, the figure implies that *A* and *B* belong to a more inclusive taxon, such as a genus. All that the cladogram shows is that this genus is more general than any of the taxa mentioned previously. It does not imply that any species is ancestral to any other species. If species *A* is more closely related to species *B* than either is to species *C*, then *A* and *B* must be classified together at a more basic level before they can be classified with species *C*, because, on the basis of the preceding argument, *A* and *B* have a more recent common ancestor than either has with *C*. However, as shall be seen shortly, the role

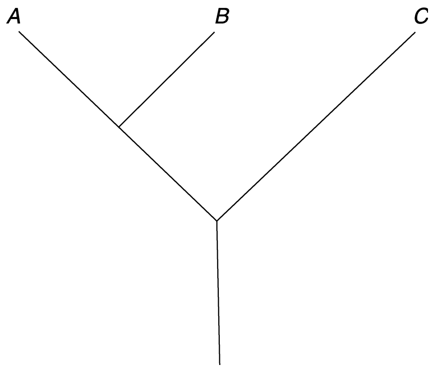


Fig. 1. A standard three-taxon cladogram. *A*, *B*, and *C* represent species. All this cladogram implies is that species *A* and *B* are more closely related to each other than either is to species *C*.

of common ancestors in phylogenetic classifications is problematic.

For cladists, cladograms and cladistic classifications are isomorphic to each other. They have the same information content. Hennig's crucial insight was that neither cladograms nor cladistic classifications can be isomorphic to phylogenetic trees. In trees, ancestral species split into descendant species. The forks in a phylogenetic tree represent speciation events. The lines represent species. For example, species *C* splits into species *A* and *B*. Thus, *C* is ancestral to *A* and *B*. Neither cladograms nor cladistic classifications represent this relation, nor can they. Genus *C* is more general than either species *A* or species *B*, but it is in no sense ancestral to these species.

At this juncture two sorts of cladists must be distinguished: *pattern cladists*, emphasizing patterns of the sister-group relation, and *phylogenetic cladists*, emphasizing the common ancestry that such relations imply. Both use Hennig's notion of cladograms but intend their cladograms to imply different relations—strict sister-group relations and common ancestry, respectively. For pattern cladists, nodes in a cladogram represent not ancestral species but increased generality, and nothing more. For phylogenetic cladists, such forks represent common ancestry but not actual common ancestors. The difference between pattern and phylogenetic cladists may not seem all that momentous. It stems from a different philosophical outlook about the goals of classification. Pattern cladists want their classifications to be as parsimonious as possible. The less information a classification presupposes, the less error can creep in. Pattern cladists limit their classifications to nested sets of characters. Such nested sets quite obviously imply something about common ancestry, but the pattern cladists are not willing to incorporate this additional element into their classifications. Phylogenetic cladists are willing to take that chance.

For higher taxa, Hennig reworked the notion of monophyly. A higher taxon is monophyletic if and only if all and only the species derived from an ancestral species are included in this single higher taxon. One result of this definition of monophyly is that such familiar taxa as Dinosauria and Reptilia cease to count as genuine taxa because they are not considered monophyletic. According to Hennig, all higher taxa must be monophyletic in his sense, but he was not willing to extend monophyly to the species level. However, several more recent authors have extended the notion of monophyly to cover species as well (Donoghue 1985; Mishler 1985; Mishler and Brandon 1987; Rosen 1979). According to the *monophyletic species concept*: A species is the least inclusive monophyletic group diagnosable by at least one autapomorphy, that is, a trait found in only one of two sister groups. This concept is designed to distinguish monophyletic groups and count them as species even if they are less inclusive than traditional species.

A second species concept that stems from Hennig's work is the *diagnostic species concept* (Cracraft 1983; Eldredge and Cracraft 1983; Nixon and Wheeler 1990). According to the diagnostic species concept: "A species is the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent." As in the case of the sophisticated phenetic species concept, more than diagnostic characters matter. A parental pattern of ancestry and descent must also be present.

One virtue of the monophyletic and the diagnostic species concepts is that they apply to both plants and animals. In addition, they are formulated to be as operational as possible; for instance, the presence of at least one autapomorphy is involved in the case of the monophyletic species concept, and a diagnosable cluster is involved in the case of the diagnostic species concept. However, neither of these species concepts always delineates traditional species. For example, when birds of paradise (Paradisaeidae) were reclassified according to the diagnostic species concept, the number of species went from roughly 40 to 90. Practicing taxonomists are a conservative lot. Stability of classifications is one of their major principles. As a result, they remain a bit leery of all these new species definitions.

A Unified Species Concept

The two final species concepts in the recent progression are Templeton's *cohesion species concept* and de Queiroz's *lineage species concept*. According to Templeton's concept (1989): "A species is the most inclusive population of individuals having the potential for phenotypic cohesion through intrinsic cohesion mechanisms" (12). Templeton contrasts phenotypic cohesion with the intrinsic (or genetic) mechanisms that produce it. For example, descent from common ancestors promotes phenotypic cohesion. The organisms belonging to a species fill the same niche because they are identical by descent. Templeton does not list all possible mechanisms that are responsible for the intrinsic potential in his definition. He thinks that they do not belong in the definition but should be appended to it, depending on the current state of empirical knowledge.

Among these intrinsic cohesion mechanisms are gene flow and natural selection, as well as ecological, developmental, and historical constraints. Asexual organisms form species, in this definition, through their adaptation to particular ecological niches.

De Queiroz (1999, 53) sees all preceding concepts as converging on a single unified species concept—the general lineage species concept, according to which species are segments of population-level lineages. De Queiroz acknowledges that his conception of the species category is very general, but this is its chief virtue. He sees no reason to include in the definition of 'species' all the causal processes that are responsible for their existence or the operational criteria used to discern them. Thus, he agrees with Simpson (1961) and Wiley (1981) that species at bottom are segments of population-level lineages. In addition, Mayr's nondimensional species are time-slices of these lineages. The monophyletic species concept is designed to distinguish phylogenetic structure within these lineages, and the diagnostic species concept emphasizes the need for distinguishing such lineages.

According to de Queiroz, too often what pass for species concepts are either specifications of the causal processes that produce these lineages or operational criteria used to recognize them in practice. As important as both these aspects of species are, they do not belong in the definition, any more than all the various ways to measure space and time belong in the definitions of these species concepts.

Pluralism

The preceding definitions are only a small sample of species concepts in the taxonomic literature (e.g., see Mayden's [1997] twenty-two species concepts and de Queiroz's [1998] thirteen). Philosophers have used this plethora of species concepts to support their preference for pluralism. They claim that no one species concept is useful for all the numerous legitimate contexts in biology; numerous species concepts are needed. Some of these authors go on to conclude that in this account, species are not real but a matter of convenience, depending on the context. Others respond that numerous different species concepts are necessary to understand natural phenomena, and for this reason all of them are equally real (Kitcher 1984).

As Mishler and Brandon (1987) have argued, one way to decrease the apparent multiplicity of species concepts is to distinguish between criteria used for grouping organisms into taxa and other criteria that serve to rank these taxa. According to these authors, the appropriate criterion for grouping organisms into taxa is monophyly, but monophyletic taxa can be found at numerous levels of inclusiveness. Quite different criteria are needed to decide which level of monophyletic taxa counts as the species level. Similarly, Templeton (1989) insists that what matters for species is internal cohesiveness. Pluralism comes into play only with respect to the various mechanisms that bring about this cohesiveness. Finally, de Queiroz (1999) presents a parallel position. All species are lineages, but the causal forces that produce these lineages and the criteria used to discern them vary. The species concept is monistic; pluralism enters in elsewhere.

If any and all species concepts formulated through the years are treated as equally legitimate, then the presence of so many species concepts certainly supports pluralism; but even the most pluralist philosophers acknowledge that some species definitions are philosophically inadmissible, such as those that claim to be totally operational and theoretically pure (Ereshefsky 1992; Kitcher 1984; Wilson 1999). Even among the species concepts that pass philosophical muster, some fail as far as scientists are concerned. Except for practical species concepts, the remainder of recent definitions of species seem to converge on a single concept: Species fundamentally are lineages. They are produced by a variety of mechanisms and are recognized by an equally wide range of techniques, but this variety need not be reflected in the definition itself.

Are all lineages equally extensive and cohesive? The answer is clearly no. Some sexual species have huge ranges and form extremely large lineages. At the other extreme, some sexual species consist of no more than a single population. They are a good deal more cohesive than their larger counterparts but much smaller in scope. As always, asexual organisms pose a major problem. They form lineages; but in the absence of gene exchange, other factors such as ecological constraints must produce the cohesiveness that these lineages may or may not exhibit.

Recent species concepts do seem to be converging on a unity. But even if all the concepts that biologists take seriously are held to delineate different species concepts, they do not make a long list, nor do they produce species that are very different from each other. In science a little bit of pluralism seems to go a long way. Most present-day philosophers revel in pluralism. If one species concept is good, two dozen are better. Scientists, on the contrary, are much less enthralled by being told that they must work with numerous different concepts delineating slightly different entities and groups of entities. Physicists put up some resistance when two different senses of 'mass' were proposed. How would they respond to two *dozen* different senses of 'mass?'

Why Does It Matter?

The literature on the species concept is huge, stretching from Aristotle to the present. Why the continuing controversy? Why the continuing fascination with species? Why does it matter? One reason is that different biologists want species to fulfill different roles. Many practicing systematists want simply a clear and easy way to identify their specimens. In which drawer does this specimen belong? No highly technical, theoretical species concept is needed to answer such a question.

Biologists as well as others are interested in conservation. One would like to slow down the current rate of mass extinction. For better or worse, the response of the U.S. Congress was to pass the Endangered Species Act, emphasizing the survival of individual species rather than more theoretically appropriate objects of concern such as entire ecosystems. But even ecosystems are not enough. Conserving biodiversity is also important. However, given current law, how inclusive we make species matters. "Splitters" produce numerous species, each of which has some claim to protection. "Lumpers" produce very few species, resulting in the vulnerability of what splitters consider species and what lumpers consider varieties or subspecies. Species definitions do matter even at this practical level.

Even more importantly, species play a role in the evolutionary process. Perhaps the situations in which species can serve as units of selection may be quite rare, but at the very least species are a result of various evolutionary mechanisms, including selection. Evolutionary biologists want to estimate such things as how common sexual reproduction is. To do so, they must be able to count entities that are comparable. If different systematists classify organisms in quite different ways, the results of any counting will be misleading. For example, if the level at which reticulation gives way to splitting is considered basic, then sexual species are comparable to asexual organisms. Suddenly, sexual reproduction becomes quite rare, instead of common. Species do matter.

The Evolution of Species: Philosophical Implications

People, including scientists, recognize all sorts of kinds, but not all these kinds are equally fundamental. For example, physicists recognize more than a hundred elements. Helium, hydrogen, lead, and gold are all elements. Numerous general statements can be made about these kinds, which have been viewed traditionally as natural kinds, important because they function in numerous general statements. 'Element' is more general than 'gold' or 'lead.' Even so, the names of individual elements are also general enough to function in laws of nature. For a long time, finding new elements aroused considerable attention. As we reach the level of radioactive elements, this quest has become greatly reduced. Such elements can exist for only nanoseconds and are produced under extremely artificial conditions.

Traditionally, species taxa have been viewed as natural kinds, akin to the physical elements. Species taxa, however, differ from elements in two important respects. First, there are many more species than elements—roughly a hundred or so elements versus millions of species—and new species continue to be discovered. Discovering a new species is rarely a significant occurrence. No one is going to become famous for discovering yet another species of fruit fly.

A second difference between the physical elements and biological species is that elements are spatiotemporally unrestricted, while species as evolvers are not. Physical elements are found scattered throughout the universe. One need not know where an atom of gold is before deciding that it is gold. Time and place do not matter. Hence, physical elements are very good candidates for natural kinds of the sort that function in natural laws. Just the opposite is the case with species as evolving lineages. They are spatiotemporally localized, and must be if they are to evolve.

Because of these differences, several authors have argued that species are not natural kinds, in fact not kinds at all, but spatiotemporal individuals, historical entities, or particulars (Ghiselin 1974; Hennig 1966; Hull 1976). The proponents of several of the more recent species concepts discussed above take this view of the metaphysical nature of species. From this perspective, particular species should not count as kinds, let alone natural kinds. Nor should their names function in any genuine laws of nature—and they do not. One reason for the continuing "species problem" is that it has taken professionals more than 2,000 years to realize that they have been putting taxa names in a metaphysical category to which such names do not belong. Instead of being highly aberrant classes, they are typical individuals (see Individuality).

But this discussion does not entail anything about the metaphysical nature of the species category itself. Species taxa are spatiotemporally restricted; the species category is not. It has all the generality needed to count as a kind. Species as evolvers are not restricted to Earth. In all probability, they have evolved numerous times throughout the universe. A particular lineage cannot evolve more than once, but lineages as such can recur. In addition, if species are that which evolves, then they function in an important scientific theory. The net effect is that species as such are a natural kind. *Homo sapiens* as a taxon is not a natural kind; the species category is.

David Hull

References

Claridge, M. F., H. A. Dawah, and M. R. Wilson (eds.) (1997), *Species: The Units of Biodiversity*. London: Chapman and Hall.

Cracraft, J. (1983), "Species Concepts and Speciation Analysis," in *Current Ornithology*, vol. 1. New York: Plenum Press, 159–187.

de Queiroz, K. (1998), "The General Lineage Concept of Species, Species Criteria, and the Process of Speciation," in D. J. Howard and S. H. Berlocher (eds.), *Endless Forms: Species and Speciation*. Oxford: Oxford University Press, 57–75.

——— (1999), "The General Lineage Concept of Species and the Defining Properties of the Species Category," in R. A. Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press, 49–89.

Donoghue, M. J. (1985), "A Critique of the BSC and Recommendations for a Phylogenetic Alternative," *Bryologist* 83: 172–181.

Eldredge, N., and J. Cracraft. (1983), "Species Concepts and Speciation Analysis," *Current Ornithology* 1, 159–187.

Ereshefsky, M. (1992), *The Units of Evolution: Essays on the Nature of Species*. Cambridge, MA: MIT Press.

Ghiselin, M. T. (1974), "A Radical Solution to the Species Problem," *Systematic Zoology* 25: pp. 536–544.

Hennig, W. (1966), *Phylogenetic Systematics*. Chicago, IL: University of Illinois Press.

Hull, D. L. (1976), "Are Species Really Individuals?" *Systematic Zoology* 25: 174–191.

Kitcher, P. (1984), "Species," *Philosophy of Science* 51: 308–333.

Mayden, R. L. (1997), "A Hierarchy of Species Concepts: The Denouement in the Saga of the Species Problems," in M. F. Claridge, H. A. Dawah, and M. R. Wilson (eds.), *Species: The Units of Biodiversity*. London: Chapman and Hall, 381–424.

Mayr, E. ([1942] 1964), *Systematics and the Origin of Species: From the Viewpoint of a Zoologist*, 2nd ed. New York: Dover.

——— (1969), *Principles of Systematic Zoology*. New York: McGraw-Hill.

Mishler, B. D. (1985), "The Morphological, Developmental, and Phylogenetic Basis of Species Concepts in Bryophytes," *Bryologist* 88: 207–214.

Mishler, B. D., and R. N. Brandon. (1987), "Individualism, Pluralism, and the Phylogenetic Species Concept," *Biology and Philosophy* 2: 397–414.

Nixon, K. C., and Q. D. Wheeler. (1990), "An Amplification of the Phylogenetic Species Concept," *Cladistics* 6: 211–223.

Otte, D., and J. A. Endler (eds.) (1989), *Speciation and Its Consequences*. Sunderland, MA: Sinauer.

Rosen, D. E. (1979), "Fishes from the Uplands and Intermontane Basins of Guatemala: Revisionary Studies and Comparative Geography," *Bulletin of the American Museum of Natural History* 162: 267–376.

Simpson, G. G. (1961), *Principles of Animal Taxonomy*. New York: Columbia University Press.

Sneath, P. H. A., and R. R. Sokal. (1973), *Numerical Taxonomy*. San Francisco: Freeman.

Templeton, A. R. (1989), "The Meaning of Species and Speciation," in D. Otte and J. A. Endler (eds.), *Speciation and Its Consequences*. Sunderland, MA: Sinauer, 3–27.

Wiley, E. O. (1981), *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. New York: John Wiley.

Wilson, R. A. (1999), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press.

See also **Conservation Biology; Evolution; Individuality; Natural Selection**

PHILOSOPHY OF STATISTICS

Philosophy of statistics may be seen to encompass the epistemological, conceptual and logical problems revolving around the use and interpretation of the methods of mathematical statistics. In contrast to the better known philosophies of science, physics, and mathematics, work in philosophy of statistics is as likely to be engaged in by practicing statisticians as by philosophers of science. Accordingly, contributions to philosophy of statistics might be regarded just as much as contributions to statistics as to philosophy of science. To make this entry useful and of manageable length, it focuses on the main philosophical debates relating to the modern methodology for *statistical inference:* significance tests, hypothesis testing, confidence interval estimation, likelihood, and Bayesian methods. This still leaves a huge territory marked by seventy years of debates widely known for reaching unusual heights both of passion and of technical complexity. To get a handle on the movements and cycles without too much oversimplification or distortion, three main waves of debates in philosophy of statistics will be distinguished: 1930–1960, 1960–1980, and 1980 to the present.

A core question that underlies the debates is: What is the nature and role of probabilistic concepts, methods, and models in making inferences in the face of limited data, uncertainty, and error? The different answers to this question have immediate ramifications for all of the central issues around which much of the debates revolve: what tasks do mathematical methods of statistics perform? And what criteria or principles are appropriate for evaluating them?

Two Roles for Probability in Inference

There are two distinct philosophical traditions regarding the role of probability in statistical inference in science. In one, probability is used to provide a post-data assignment of degree of probability, confirmation, support, or belief in a hypothesis, while in a second, probability is used to assess the probativeness, reliability, trustworthiness, or severity of a test or inference procedure.

Confirmation Theory

Conceding that all attempts to solve the problem of induction (see Induction, Problem of) suffered from circularity (Salmon, 1967), philosophers of induction (e.g., in the 1970s) turned their attention instead to constructing *logics of induction* or *confirmation theories* that would, ideally, reflect "inductive intuition." The goal would be to supply means to compute the degree of *evidential relationship* between given evidence statements, e, and a hypothesis, H (see Confirmation Theory). A natural place to look for such a computation is the definition of conditional probability, or *Bayes's theorem:*

$$P(H \mid e) = P(e \mid H)P(H)/P(e)$$

where $P(e) = P(e \mid H)P(H) + P(e \mid \neg H)P(\neg H)$.

Computing $P(H \mid e)$, the *posterior probability*, requires starting out with a probability assignment to all of the members of $\neg H$, and a major source of difficulty through all three waves is how to obtain, justify, and interpret these prior probabilities. Insofar as the computed degrees of confirmation are viewed as analytic and a priori, their relevance for predicting and learning about empirical phenomena is problematic; insofar as they measured subjective degrees of belief, their relevance for giving objective guarantees of reliable inference is unclear (see Bayesianism; Confirmation Theory; Inductive Logic).

The Error-Probability Philosophy ('Sampling Theory')

A distinct philosophical tradition uses probability to characterize a procedure's overall reliability in a series of (actual or hypothetical) experiments or in *repeated sampling* (hence, 'sampling theory'). These probabilistic properties of statistical procedures are called *error frequencies* or *error probabilities* (e.g., significance levels, confidence levels). Deliberately designed to reach conclusions about statistical parameters without invoking prior probabilities in hypotheses, error probabilistic methods use probability to quantify how *frequently* methods discriminate between alternative hypotheses and how *reliably* they facilitate the detection of error. As with logics of confirmation, there are connections with philosophy of induction, as in Peirce, Braithwaite, and, to some extent, Popper (see Popper, Karl Raimund). These two contrasting philosophies of the role of probability in statistical inference correspond to the core issues at the heart of the debate in all three waves of philosophy of statistics.

The First Wave

Quantitative methods of statistical inference involve drawing conclusions about parameters on the basis of the observed values of random variables. Statistical methods may be seen to connect questions about the phenomenon or data-generating source to questions about distributions of random variables that model the data-generating source or population. Thus the conception of a statistical model wherein these parameters are defined is an important component of statistical inference methods. The area of model specification and model selection has its own set of philosophical issues that will not be taken up here. A statistical hypothesis cannot be just any claim, but must give probability assignments to the different experimental outcomes or sample space Ξ, typically in terms of the parameters of the model. That is, for any x in Ξ, H assigns "the probability of x *under H,*" written $P(x;H)$. This notation helps avoid confusion between a probabilistic computation under a model and conditional probabilities needed for Bayes's theorem, $P(x \mid H)$, without prejudging issues. (An alternative notation some find useful is $P(x \| H)$; see Friedman 1995).

Fisherian "Simple" Significance Tests

The modern approach to statistical inference was initiated by Fisher, who introduced the main concepts and procedures of statistical significance tests. Fisher's strong objections to Bayesian inference (Fisher 1935, 1955), and in particular to the use of prior distributions, led Fisher to develop ways to express the uncertainty of inferences without deviating from frequentist probabilities.

The significance test is a procedure with the following components: there is a null hypothesis H_0 that is an assertion about the distribution of the sample $X = (X_1, \ldots, X_n)$, and a function of the sample, $d(X)$, th*e test statistic,* which measures the difference between the data $x_0 = (x_1, \ldots, x_n)$, and null hypothesis H_0. The larger the value of $d(x_0)$, the further the outcome is from what is expected under H_0, with respect to the particular question being asked (x_0 represents a particular realization of X). For an observed difference $d(x_0)$, the test computes the p-value, or the probability of a difference larger than $d(x_0)$, computed under the assumption that H_0 is true:

$$p(x_0) = P(d(\mathrm{X}) > d(x_0); H_0).$$

The p-value may be regarded as a measure of discordancy from H_0: the smaller the significance level, the greater the discordance between x_0 and H_0 (Kempthorne and Folks 1971).

Fisher described the significance test as a procedure for rejecting the null hypothesis and inferring that the phenomenon has been "experimentally demonstrated" (Fisher 1935, 14), where the latter inference corresponds to finding a small p-value, such as .05 or .01. How to justify this is a point of philosophical debate. One highly influential example is this. Suppose that x_0 is evidence against H_0 just in case x_0 is statistically significant at a small level p (or smaller). Then p is the maximal probability of rejecting H_0 when H_0 is actually a correct description of the underlying data-generating mechanism. So there is only a small probability of erroneously rejecting H_0, i.e., committing what Neyman-and Pearson call a *type I error* (Cox, 1958). Commonly used significance tests—Pearson's chi-square goodness of fit, the Student t test, the F test in analysis of variance—are regularly used to distinguish real effects of importance from apparent effects actually due to random sampling or uncontrolled variability.

The Alternative or "Non-Null" Hypothesis

Evidence against H_0 would seem to indicate evidence for some alternative, if only for a directional departure from the null value in a given direction. Although Fisherian significance tests strictly consider only the null hypothesis, Neyman and Pearson tests introduce as well an alternative H_1. Despite the bitter disputes with Fisher that were to erupt soon after their early developments of tests, Neyman and Pearson, at the outset, regarded their work as merely placing Fisherian tests on firmer logical footing by taking explicit account of an alternative to the null hypothesis.

Neyman-Pearson (N-P) Tests The N-P hypothesis test, mathematically considered, is a rule that maps each possible outcome $x = (x_1, \ldots, x_n)$ onto one of two hypotheses, the test or null hypothesis H_0 or an alternative hypothesis H_1. As in the Fisherian (simple) significance test, there is a test statistic $d(X)$, in terms of which the test rule is defined. In the N-P test, however, the values of $d(X)$ that will be taken to reject H_0 are fixed at the outset, by a predesignated choice of significance level. Most importantly, the null and alternative of an N-P test exhaust the parameter space of the statistical model, whereas in the Fisherian test there is the single null hypothesis, as against its logical complement.

The N-P error probabilities are computed under the assumption that the statistical model is adequate; what is being tested are the values of one or more parameters governing the distribution. For simplicity, illustrations here keep to the case of only one unknown parameter. Although N-P theory provides distinct tests of the assumptions of the statistical model, and the whole issue of model validation is important philosophically, the matter will not be explicitly discussed here.

Example, Test T(α): Consider a random sample of size *n*, $\boldsymbol{X} = (X_1, \ldots, X_n)$, where it is assumed that each X_i is normal $N(\mu,\sigma^2)$, independent and identically distributed (IID). Test $T(\alpha)$ denotes the familiar test of H_0: $\mu \leq \mu_0$ against H_1: $\mu > \mu_0$, where H_0 is the null, and H_1 the alternative hypothesis. Because H_1 includes only positive discrepancies from H_0, this is called a one-sided test. For simplicity, let the standard deviation σ be known—for instance, let $\sigma = 1$. The test statistic for T(α) is: $d(\mathbf{X}) = (\bar{X} - \mu)/\sigma_x$, where $\bar{X}$ is the sample mean with standard deviation $\sigma_x = (\sigma\sqrt{n})$. The N-P test with *significance level* α rejects H_0 with data x_0 if and only if $d(x_0)$ reaches the preset significance level α—for instance, $c_\alpha = 1.96$ for $\alpha = .025$, so

$$\text{Test T}(\alpha)\text{: if } d(\boldsymbol{x}_0) > c_\alpha, \text{ reject } H_0,$$
$$\text{if } d(x_0) \leq c_\alpha, \text{ accept } H_0,$$

The set of all outcomes that lead to "reject H_0" is called the *rejection region*. "Accept" and "reject" should be regarded as parts of the mathematical apparatus whose interpretation must be separately considered.

The test is specified so that the probability of a type I error, α, is fixed at some small number, such as .05 or .01, the *significance level* of the test:

$$\text{Type I error probability} = P(\text{Test T}(\alpha) \text{ Rejects } H_0; H_0) \leq \alpha.$$

Since "Test T(α) Rejects H_0" iff $\{d(X) > c_\alpha\}$, it follows that

$$\text{Type I error probability} = P(d(X) > c_\alpha; H_0) \leq \alpha.$$

N-P test principles then seek out the test that at the same time has a small probability of committing a type II error, β. Since the alternative hypothesis H_1, as is typical, contains more than a single value of the parameter, it is *composite*, the type II error probability is evaluated at a specific point $\mu = \mu_1$, and thus is abbreviated $\beta(\mu_1)$:

P(Test T(α) does not reject H_0; $\mu = \mu_1$) = $P(d(X) \leq c_\alpha; H_0) = \beta(\mu_1)$, for $\mu_1 > \mu_0$.

The "best" test with significance level β (if it exists) is the one that at the same time minimizes the value of β for all $\mu_1 > \mu_0$, or equivalently, maximizes the *power*:

$$\text{POW}(\text{T}(\alpha); \mu_1) = \text{P}(d(\text{X}) > c_\alpha; \mu_1), \text{for all } \mu_1 > \mu_0.$$

T(α) is said to be a *uniformly most powerful* (UMP) α significance level test. Letting $\alpha = .025$, $T(\alpha)$ If $d(x) > 1.96$, reject H_0. The rejection region for the corresponding two-sided .05 test, H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$, abbreviated as T(2α) is: $\{x : |d(x_0)| > 1.96\}$.

Error Probabilities Versus Conditional Probabilities

Confusion often results from interpreting the type I error probability: $P(d(\boldsymbol{X}) > c_\alpha; H_0)$ as a conditional probability statement of the form: $P(d(\boldsymbol{X}) > c_\alpha|H_0)$. From the definition of conditional probability it follows that

$$P(d(\boldsymbol{X}) > c_\alpha|H_0) = [P(d(\boldsymbol{X}) > c_\alpha, \mu = \mu_0)]/P(\mu = \mu_0).$$

However, neither the numerator $P(d(X) > c_\alpha, \mu = \mu_0)$ nor the denominator $P(\mu = \mu_0)$ of this ratio are meaningful unless the parameter may be assumed to be a random variable, as in a Bayesian approach (see Bayesianism).

In the N-P testing paradigm, there is no probability assignment to the conjunctive event ($d(\boldsymbol{X}) > c_\alpha$, $\mu = \mu_0$) or to ($\mu = \mu_0$). The statement $P(d(X) > c_\alpha; H)$, should be interpreted as the probability of rejecting H_0 when evaluated under the hypothetical scenario that the observed outcome x_0 has arisen from the distribution described in H_0. Within the error probability (frequentist) framework, a statistical hypothesis *H* either does or does not adequately describe the process generating the data. There is no suggestion that any *H* is precisely true; indeed, the purpose of tests is to evaluate discrepancies of specified sorts. But probabilities enter in this evaluation only as error probabilities.

Inductive Behavior Philosophy

Philosophical issues and debates arise once one begins to consider the uses to which these formal statistical tools might be put, the interpretations of the formal apparatus, and the justifiability of associated principles of tests. The proof by Neyman and Pearson of the existence of best tests set the stage for the mathematical development of statistical tests as rigorous rules for "deciding" to accept or reject hypotheses. In this conception, to infer the conclusion of the significance testing argument, 'data x_0 is evidence against H_0' or 'x_0 indicates the falsity of H_0,' is to take a decision of a sort, with a calculable risk. Wishing to draw a stark contrast between this conception of tests and those of Fisher as well as Bayesians (Jeffreys), Neyman declared that the goal of tests is not to adjust beliefs but rather to "adjust behavior" to limited amounts of data. Tests, accordingly, are not rules of inductive inference but rules of behavior. The value of tests as rules of behavior is that "it may often be proved that if we behave according to such a rule ... we shall reject H when it is true not more, say, than once in a hundred times, and in addition we may have evidence that we shall reject H sufficiently often when it is false" (Neyman and Pearson 1933, 142).

Debates Between Fisher and Neyman and Pearson: The 1950s

The dispute between "inductive behavior" and "inductive inference" coming on top of the break between Fisher and Neyman, which began in 1935, commingled philosophical, statistical, and personality clashes. Fisher (1955) denounced the way that Neyman and Pearson transformed "his" significance tests into "acceptance procedures," wherein tests are viewed as mechanical rules or recipes for deciding to accept or reject statistical hypothesis H_0, and the concern has more to do with speeding up production or making money than in learning about phenomena. In responding to Fisher, Pearson clearly distanced himself from Neyman's "inductive behavior" jargon, calling it "Professor Neyman's field rather than mine" (Pearson 1955, 207). However, Pearson protested that neither he nor Neyman were "speaking of the final acceptance or rejection of a scientific hypothesis on the basis of statistical analysis. … Indeed, from the start we shared Professor Fisher's view that in scientific enquiry, a statistical test is 'a means of learning'" (204–205).

Neyman, too, despite promoting "inductive behavior as a major concept in philosophy of science" (1957a), clearly denounced "mechanical" uses of significance tests (in responding to Fisher), and had no hesitation in using N-P tests for "inference" or reaching "conclusions." Tracing out the thrust and parry between Neyman, Pearson, and Fisher in the 1950s will amply reward those interested in what the key players "really thought." Later on, the N-P tests became so formally entrenched in the decision-theoretic framework of Wald (1950) that many of the qualifications by Neyman and Pearson in the first wave have been overlooked in the philosophy of statistics literature.

Confidence Interval Estimation Procedures

Statistical inference can take the form of estimation procedures as well as tests. In confidence interval (CI) estimation procedures, a statistic is used to set upper or lower (one-sided) or both (two-sided) bounds. The concept of a confidence interval with a frequentist interpretation was first introduced by Neyman (1935) as a way to extend point estimation to interval estimation, with a predesignated error rate. For a parameter, say, μ, a (1-α) confidence interval estimation procedure leads to estimates of form:

$$\mu = \bar{X} \pm \mathrm{e}$$

Different sample realizations x lead to different estimates, but one can ensure that $(1 - \alpha)$ 100% of the time the true parameter value μ, whatever it may be, will be included in the interval formed.

Dualities Between One- and Two-Sided Intervals and Tests

There exists a duality relationship between CIs and hypothesis tests that can be used to derive optimality properties for CIs analogous to those of tests. The general correspondence between a $(1 - \alpha)$ confidence intervals and tests is this: the confidence interval contains the values that would not be rejected by the given test at the specified level of significance (Neyman 1935); they would not be rejected because they would not be statistically significant (from the observed x_0) at significance level α, by the corresponding test. Consider test T(α). It follows that the $(1 - \alpha)$ one-sided interval corresponding to test T(α) is $\alpha > \bar{X} - \mathrm{c}_\alpha(\sigma\sqrt{\mathrm{n}})$. In particular, the 97.5% confidence interval estimator corresponding to test T(α) is:

$$\mu > \bar{X} - 1.96(\sigma\sqrt{\mathrm{n}}).$$

To grasp the duality, one must think not of a fixed null hypothesis, e.g., $\mu = 0$, but rather of different values for μ_0 that might have been tested. In particular, were the test of null hypothesis.

H_0: $\mu < (\tilde{\bar{x}}\mathrm{c}_\alpha(\sigma\sqrt{\mathrm{n}}), \mathrm{H}_0$ would have been rejected at level α. Similarly, the 95% CI for μ corresponding to the two-sided test, $\tilde{\mathrm{T}}(.05)$ is:

$$(\bar{X} - 1.96(\sigma\sqrt{\mathrm{n}}) \leq \mu < \bar{X} - 1.96(\sigma\sqrt{\mathrm{n}})).$$

These dualities will figure importantly in wave III.

Fisher's Criticism of Confidential Intervals: Fiducial Intervals

Calling $(1 - \alpha)$ the "confidence level" of the estimation procedure was infelicitous. It encourages the supposition that $(1 - \alpha)$ is the degree of confidence to be assigned the particular interval *estimate* formed, once $\bar{X}$ is instantiated with $\bar{x}$. That would be fallacious. Once the estimate is formed, either the true parameter is or is not contained in it. One can say only that the particular estimate arose from a procedure which, with high probability, $(1 - \alpha)$, would contain the true value of the parameter, whatever it is.

Fisher, in what is regarded as one of the most puzzling episodes in philosophy of statistics, seemed to advocate this fallacious instantiation for certain contexts. Fisher (1955) claimed N-P confidence interval methods are guilty of violating the principles of deductive logic by allowing

$$\mathrm{P}(\bar{x} - \mathrm{c}_\alpha(\sigma\sqrt{\mathrm{n}}) \leq \mu < \bar{x} - \mathrm{c}_\alpha(\sigma\sqrt{\mathrm{n}})) = 1 - \alpha \quad (1)$$

and yet upon observing a particular $\bar{x}$, denying that the probability holds for the resulting CI estimate:

$$(\bar{x} - \mathrm{c}_\alpha(\sqrt{\mathrm{n}}) \leq \mu < \bar{x} - \mathrm{c}_\alpha(\sqrt{\mathrm{n}})) \quad (2)$$

Fisher claimed that, at least in certain special cases, it *was* possible to assign a probability or “fiducial distribution” to the interval statement about μ, while keeping within the sampling distribution perspective, a move that Savage (1962) described as “an attempt to make the Bayesian omelet without breaking the Bayesian eggs.” Although the possibility of nonfallaciously instantiating into statement (1), to arrive at (2), without introducing a prior probability distribution, has tantalized researchers in philosophy of statistics, Fisher’s fiducial argument is generally regarded as a lapse, even by Fisher’s most ardent admirers (Hacking 1965; Seidenfeld 1979).

The Second Wave

The set of issues that swirled around the philosophy of statistics debates from the early 1960s through the late 1970s echoed the earlier debates but reflected as well changing problems in philosophy of science, statistics, and the statistical practices in the social sciences. Foundational debates of this period are noteworthy for the amount of direct interactions between philosophers of science and statistics; as is in evidence in the two significant collections of Godambe and Sprott (1971) and Harper and Hooker (1976).

As the most impressive mathematical developments of N-P theory occurred in a decision-theoretic framework, generalized further by Wald (1950)—the Neyman-Pearson-Wald (NPW) approach—it was the behavioristic-decision paradigm that bore the brunt of criticism from philosophy. Critics aimed at two central features of the “accept-reject” behavioristic conception of N-P tests: first, the justification of tests in terms of low (long-run) error rates alone, and second, the function of tests as routine, mechanical, or automatic accept-reject routines. While these features, taken strictly, give a caricature of tests—even as their founders intended and used them—they are at the heart of the philosophical criticisms of N-P testing. Not all critics call for tools that are more inferential and less decision-theoretic; some complained that N-P theory was at best a halfway house to a full-blown decision theory, with explicit loss functions, and prior probabilities that would be combined with measures of evidence (see Decision Theory). Because critics from both these camps hold a degree of confirmation stance, while error statisticians look to probability for objective measures of reliability of procedures, the disputants often talk past each other.

Error Probability Principle Versus Likelihood Principle

Hacking (1965) framed the main lines of criticism by philosophers in charging “Neyman-Pearson tests as suitable for before-trial betting, but not for after-trial evaluation” (99). Analogous charges are put in terms of distinctions between “initial precision” versus “final precision,” and “before-data vs. after data” evaluation. According to such “post-data criticisms,” N-P tools license inferences that while satisfactory from the pre-data viewpoint, seem unsatisfactory according to one of the post-data measures of (absolute or relative) evidential strength. The more general point may be put as follows:

- Data sets $\boldsymbol{x}$ and $\boldsymbol{y}$ may have exactly the same evidential relationship to hypothesis H, on a given degree of support measure, yet warrant different inferences according to significance test reasoning because x and y arose from tests with different *error probabilities*.

Such charges have weight, of course, only to the extent that one accepts the particular degree of support measure involved, the most common being based on the likelihood function of H, often written $\mathrm{L}(H; \boldsymbol{x})$, where $\mathrm{L}(H; \boldsymbol{x}) = \mathrm{P}(\boldsymbol{x}; H)$. There is often confusion about likelihoods. Unlike the probability function, which assigns probabilities to the *different* possible values of the random variable of interest X, under some *fixed* value of the parameter(s) such as μ, the likelihood function gives the probability (or density) of a *given* observed value of the sample under the different values of the unknown parameter(s) such as μ.

Hacking (1965) championed an account of comparative support based on his “law of likelihood”: Data $\boldsymbol{x}$ support hypotheses H_1 more than H_2 if the latter is *more likely* than the former, i.e.,

$P(\boldsymbol{x}; H_1) > P(\boldsymbol{x}; H_2)$. When there are many hypotheses, one takes the one that maximizes the likelihood. A problem is that there is always the rival hypothesis that things had to turn out the way they did. If such an alternative can always be constructed, then it will be possible to find H less well supported than some other hypothesis, even if H is true. Hacking (1965) rejected this likelihood approach on these grounds, but likelihoodist accounts are advocated by others and remain the focus of active interest (Birnbaum 1961; Royall 1997).

The likelihood function has an important role in all of the statistical accounts, but for those who endorse the likelihood principle, likelihoods suffice to convey "all that the data have to say." That is the gist of the *likelihood principle*—a pivot point around philosophy of statistics discussions:

> According to Bayes's theorem, $P(x|\mu)$...constitutes the entire evidence of the experiment, that is, it tells all that the experiment has to tell. More fully and more precisely, if y is the datum of some other experiment, and if it happens that $P(x|\mu)$ and $P(y|\mu)$ are proportional functions of μ (that is, constant multiples of each other), then each of the two data x and y have exactly the same thing to say about the values of μ .... (Savage 1962)

By contrast, the error probabilist must consider, in addition, the sampling distribution of the likelihoods (under hypotheses of interest). Thus, as Savage (1962) argued, significance levels and other error probabilities all violate the likelihood principle, leading to one of the most crucial philosophical controversies.

Debate Over the Relevance of the Stopping Rule

The conflict between significance levels and the LP is often illustrated by a variation on the two-sided test $T(2\alpha)$: a random sample from a normal distribution with mean μ and standard deviation 1, that is, $X_i \sim N(\mu, 1)$; with H_0: $\mu = 0$, and H_1: $\mu \neq 0$. However, instead of fixing the sample size n in advance, n is determined by a *stopping rule*:

$$\text{Keep sampling until } |\bar{x}| \geq 1.96/\sqrt{n}.$$

The probability that this rule will stop in a finite number of trials is 1, regardless of the true value of μ; it is a *proper* stopping rule. Whereas with n fixed in advance, such a test has a type 1 error probability of .05, with this stopping rule, the actual significance level differs from, and is greater than .05. Significance levels are sensitive to the stopping rule; and there is considerable literature on error probability adjustments for "optional stopping," that is, on *sequential tests* (e.g., Armitage 1961). By contrast, since likelihoods are unaffected by this stopping rule, the LP proponent denies there is an evidential difference between the two cases. For some, this was yet further grounds to embrace a Bayesian account:

> The likelihood principle emphasized in Bayesian statistics implies,...that the rules governing when data collection stops are irrelevant to data interpretation. It is entirely appropprioate to collect data until a point has been proved or disproved. (Edwards, Lindman, and Savage 1963, 193)

For others it only underscored the point raised by Pearson and Neyman, that "knowledge of [the likelihood ratio] alone is not adequate to insure control of the error involved in rejecting a true hypothesis" (Pearson and Neyman 1930, 106). The literature here is vast; at best one can list sources (beyond those already mentioned) with fairly broad citations (Cox and Hinkley 1974; Mayo and Kruse 2001).

The key difference between the two perspectives is that the holder of the LP considers the likelihood of the *actual* outcome, that is, just $d(\boldsymbol{x})$, whereas the error statistician considers the likelihoods of values *other than the one observed* in order to assess the properties of the test procedure. The calculation of error probabilities, the sampling distribution, all depend on the relative frequency of outcomes other than the one observed, for example, outcomes as or more statistically significant—the "tail area." This remains a pivot point around which controversy in philosophy of statistics revolves. It is not a matter of one side being right and the other wrong, it is a matter of holding different aims, which in turn grow out of different philosophies of statistics.

The Significance Testing Controversy

Morrison and Henkel (1970) stands as a hallmark to the foundational issues wrestled with by social and behavioral scientists of this period. Where philosophers directed most of their criticisms to N-P tests, the focus here tended to center on simple Fisherian significance tests that had been widely adopted in psychology and other social sciences. Chastising social scientists for applying significance tests in slavish and unthinking ways, contributors call attention to a cluster of pitfalls and fallacies of testing. These fallacies are at the center of the philosophical controversies in this and later waves:

(i) *Large N Problem*: With large enough sample size, an α significant rejection of H_0 can be very probable, even if the underlying discrepancy from μ_0 is substantively trivial. In fact, for any discrepancy from the null, however small, one can find a sample size such that there is a high probability (as high as one likes) that the test will yield a statistically significant result (for any *p*-value one wishes). Nevertheless, as Rosenthal and Gaito (1963) document, statistical significance at a given level is often (fallaciously) taken as more evidence against the null the larger the sample size (*n*). In fact, it is indicative of *less* of a discrepancy from the null than if it resulted from a smaller sample size. The "large *n* problem" is also the basis for the "Jeffrey-Good-Lindley" paradox brought out by Bayesians: even a highly statistically significant result can, as *n* is made sufficiently large, correspond to a high posterior probability accorded to a null hypothesis (see Bayesianism). Some suggest adjusting the significance level as a function of *n*, others, introducing some measure of the size of the discrepancy or "effect size" indicated. These issues return in the third wave.

(ii) *Fallacy of Non-Statistically Significant Results*: Test T(α) fails to reject the null, when the test statistic fails to reach the cutoff point for rejection, that is, $d(\boldsymbol{x}_0) \leq c_\alpha$. A classic fallacy is to construe such a "negative" result as evidence of the correctness of the null hypothesis. The problem is that merely surviving the statistical test is too easy, occurs too frequently, even when the null is false. One can always find a sufficiently small discrepancy δ from the null such that the test has low power to detect it. Thus, it would be fallacious to regard insignificant results as evidence that the discrepancy is less than δ, much less that there is no discrepancy at all. With publishers demanding at least a .05 significant result for publication, many of these studies remain tucked away, the so-called "file-drawer problem" (Meehl 1990).

The Power Analytic Movement of the 1960s

In their attempt to inculcate the calculation of power in psychology Cohen (1988) and others began, in the 1960s, the "power analytic" movement. The attention to power, of course, was a key feature of N-P tests, but apparently the prevalence of Fisherian tests in the social sciences, coupled, perhaps, with the difficulty in calculating power, resulted in power receiving short shrift.

Although this was less well advertised, the power analysts used power not only for planning but for interpreting nonsignificant results post-data: If a non-statistically significant result occurred with a test with low power to detect discrepancies of interest, the power analysts urged, then such a nonsignificant result should not be taken to rule out such departures from the null. In so doing, one is codifying a means to avoid the fallacy of taking "no evidence against" the null as "evidence for" the null.

It may be surprising to include Neyman, but one finds just such a post-data use of power in the occasional papers of Neyman in the 1950s. In one, Neyman addresses Carnapian confirmation: "In some sections of scientific literature the prevailing attitude is to consider that once a test, deemed to be reliable, fails to reject the hypothesis tested, then this means that the hypothesis is "confirmed". Calling this "a little rash" and "dangerous," he claims "a more cautious attitude would be to form one's intuitive opinion only after studying the power function of the test applied" (Neyman 1955, 41).

One is advised to consider: (i) how large a discrepancy from the null is considered "important" or non-trivial on substantive grounds (to be determined by the tester) $\delta_{\text{non-trivial}}$, and (ii) the power of detecting a $\delta_{\text{non-trivial}}$ with the test actually used, for example, Power($T(\alpha)$,$\delta_{\text{non-trivial}}$). If the power is low, "the fact that the test failed to detect the existence of δ "does not mean very much. In fact, [$\delta_{\text{nontrivial}}$] may exist and have gone undetected" Neyman (1957b, 16). So here in Neyman are the basic outlines of the post-data "power analytic" movement, admittedly, largely lost in the standard decision-behavior model of tests.

However, even the post-data use of power retains an unacceptable coarseness: power is always calculated relative to the cutoff point $c_\square$ for rejecting H_0. Consider test T($\alpha = .025$), $\sigma = 1$, $n = 25$, and suppose $\delta_{\text{non-trivial}} = .2$ is deemed "substantively important". To determine if "it is a little rash" to take a nonsignificant result, say $d(\boldsymbol{x}) = -.2$, as reasonable evidence that $\delta < \delta_{\text{nontrivial}}$ (i.e., an important discrepancy is absent), one is to calculate POW(T($\alpha = .025$), $\delta_{\text{nontrivial}}$) which is only .16! But why treat the particular non-significant result the same no matter how close it is to μ_0 (i.e., 0)? In fact $P(d(\boldsymbol{x}) > -.2; .2) \approx .93$. That is, were μ as large as .2, the test very probably would have detected a more significant result. This suggests that rather than calculating

$$P(d(\boldsymbol{X}) > c_\alpha; \mu = .2), \qquad (A)$$

one should calculate

$$(B)P(d(\boldsymbol{X}) > d(\boldsymbol{X}_0); \mu = .2). \qquad (B)$$

Even if (A) is low, (B) may be high. Whether Neyman and Pearson did or would have endorsed this modification of the pre-data error probabilities is an open question. The issue reappears in the "reforms" of the third wave.

The Third Wave: Relativism, Reforms, Reconciliations

Statistics in Meta-Methodology

In the 1980s and 1990s statistical inference began to figure in rational reconstructions of scientific episodes, in appraising methodological rules e.g., the value of novel evidence, the prediction versus accommodation debate (e.g., Howson and Urbach 1989; Glymour 1980; Mayo 1991) and in attempts to solve classic philosophical problems, such as Duhem's problem (Howson and Urbach 1989). The recognition that science in general, and statistical inference in particular, involves subjective judgments and values, the statistical method, most often appealed to here is largely one or another subjective Bayesian account. One can explain historical cases wherein anomalies are blamed on background rather than a hypothesis *H*, some argue, by showing how plausible prior beliefs could still permit *H* to have a reasonably high posterior degree of belief. Others charge that the very flexibility Bayes's theorem offers in reconstructing cases as rational is to sidestep the question at hand: Which hypothesis *ought* to be blamed for an anomaly? (Mayo 1997; Worrall 1993).

Bayesian Advances and Controversy

The heat of the old debates is less in evidence in the third wave. For the most part statisticians are comfortable with an eclecticism, wherein different methods may be suitable for different functions, for example "pure" (Fisherian) tests in some cases, N-P "decision procedures" in others, along with good-sense, informal recommendations for their interpretation. To others, particularly nonstatistician practitioners (e.g., in psychology, ecology, medicine), the situation seems less one of joyful eclecticism, and more one of "unholy hybrids" yielding a mixture of ideas from N-P methods, Fisherian tests, and Bayesian accounts that is "inconsistent from both perspectives and burdened with conceptual confusion" (Gigerenzer 1993, 323). Because increasingly philosophers of science come to these issues by way of subject matter fields, they are more likely to be users of the latest methods rather than occupy their historical role as outside critic.

The use of Bayesian methods has grown exponentially both because of the philosophical problems with error statistical methods as well as the development of effective computational tools such as a Markov Chain Monte Carlo (MCMC). The rise in statistical computer packages means that Bayesian and non-Bayesian methods are readily available, encouraging the practitioner to view them as simply enriching the statistical toolkit rather than as reflecting different perspectives on philosophical foundations. In this sense the use of high-powered statistical tools increases the distance between the use and philosophical foundations of the methods. But when competing interpretations arise, as they often do, the philosophical questions from the first and second waves re-emerge. Most especially are debates about the role and justification of Bayesian prior probabilities. Operating with mathematically convenient priors is common but, as Bayesians are well aware, more is needed to justify them. One important argument put forward shows that with sufficient data, posterior probabilities will converge even if they are based on different priors (see Bayesianism). As Kyburg (1993, 146) shows, however, for any body of data there are non-extreme prior probabilities that will result in posteriors that differ by as much as one wants.

A related argument defending the use of priors shows that it is possible to ascertain the influence different priors may have, and so long as the posterior remains relatively insensitive the Bayesian inference is *robust* to the prior. A question that arises is this: if when the choice of prior is found to matter one must seek a different procedure, and if there are sufficient data such that the choice of prior scarcely matters, then why is the prior relevant at all? Does not this revert to the goal that drove Neyman, namely, to find procedures whose validity does not depend on the priors? The appeal of error statistical methods, despite problems, is that they apply for the kinds of uncertain cases scientists often face. Granted it is appealing to enlist the beliefs of "experts," but the question is how to retain the ability to critique and hold them accountable—a growing concern in evidence-based policy. Error probabilities can be calibrated against empirical frequencies, but can one equally well calibrate the opinion of the experts?

Reforms Within Error Statistics

There is an extensive movement to retain error statistical tools and yet reform them in order to avoid the well-known fallacies and shortcomings. The significance test fails to convey the effect of discrepancy warranted, and thus many journals require they be supplemented with measures of effect size. The most fruitful idea seems to be to appeal to two sided CI estimation procedures, even in interpreting the one-sided test T(α).

Consider interpreting non-significant results. Since all elements of the CI "fit" or are consistent with the outcome at the given level, the interpreter is deterred from thinking there is evidence for 0. But, as critics note, this will not go far enough to block fallacies of acceptance in general. For example, the $(1 - 2\alpha)$ CI for the parameter μ in test T(α) with $\alpha = .025$ is: $[\bar{x} - 1.96(\sigma\sqrt{\mathrm{n}}), \bar{x}1.96(\sigma\sqrt{\mathrm{n}})]$ $\sigma = 1, n = 25, (\sigma\sqrt{n}) = .2$. Outcome $\bar{x} = .39$ just fails to reject H_0 at the .025 level, and correspondingly 0 is included in the two-sided 95% interval: $(-.002 < \mu < .782)$ (see duality between tests and CIs, above). Consider now the inference $\mu < \mu_1$ for μ_1 within the CI, say, $\mu < 0.2$. The hypothesis $\mu < 0.2$ is non-rejectable by the test—it is a survivor, as it were. But the construal is dichotomous: In or out, plausible or not; all values within the interval are *on par*, as it were (Mayo 1996). This does not adequately prevent fallacious interpretations of non-significance (fallacies of acceptance). Although $\bar{x}$ is not sufficiently greater (or less) than any of the μ values in the confidence interval to reject them at the α-level, this does not imply there is evidence *for each* of the values in the interval (Mayo and Spanos 2005).

Severity Assessments

The power analyst would seem to do better here. For each value of μ_1 in the confidence interval, there would be a different answer to the question: What is the power of the test against μ_1? Thus the power analyst makes distinctions that the CI interval theorist does not. The power analyst blocks the inference $\mu < 0.2$ since POW(T(α = .025), .2) is low (.16). But, as seen in the second wave, there is an important weakness of the use of power to avoid fallacies of acceptance. Were the result not $\bar{x} = .39$, but rather $\bar{x} = -.2$, the test again fails to reject H_0, but the power analyst, looking just at $c_\alpha = 1.96$ is led to the same assessment denying there is evidence for $\mu < 0.2$. (power analysts commonly recommend a power of .8 as high). Although the "prespecified" power is low, .16, it seems clear that the interpretation, post-data, should reflect the actual outcome, and there is a high probability for a more significant result than the one attained, were μ as great as 0.2! Rather than construe "a miss as good as a mile," parity of logic suggests that the post-data power assessment should replace the usual calculation of power against μ_1:

$$\mathrm{POW}(\mathrm{T}(\alpha), \mu_1) = P(d(\mathbf{X}) > \mathrm{c}_\alpha; \mu = \mu_1),$$

with what might be called the *power actually attained* or, to have a distinct term, the *severity (SEV):*

$$\mathrm{SEV}(\mathrm{T}(\alpha), \mu_1) = \mathrm{P}(d(\mathbf{X}) > d(x_0); \mu = \mu_1),$$

where $d(x_0)$ is the observed (nonstatistically significant) result (Mayo and Cox 2005). SEV(T(α), $d(\boldsymbol{x}_0)$, $\mu < \mu_1$) is a shorthand for "the severity of the test which $\mu < \mu_1$ has passed on the basis of the insignificant result $d(x_0)$ from test T(α)." This is the post-data measure of a test's *severity* for detecting discrepancies as large as $\gamma = \mu_1 - \mu_0$. Since T(α)'s probativeness would be even higher for greater values of μ, it follows that SEV(T(α), $\mu < \mu_1$) > P($d(\mathbf{X}) > d(x_0)$; $\mu = \mu_1$) Mayo and Spanos 2005).

The philosophical position here is that error probabilities serve a function in a post-data interpretation of statistical inferences, by characterizing the probativeness of the particular test result with respect to a particular interpretation or particular inference on may wish to consider: Pre-data, one is balancing the two types of errors; but post-data, the concern shifts to evaluating if particular inferences are warranted. Figure 1 compares power and severity.

Conversely, for any non-significant result from test T(α), one may find the value of μ against which the test has high severity, say .975. This is solved by $\mu_1 = \bar{x} + 1.96\sigma_x$, which is noticed to be the same value as the upper bound of a two-sided .95 level CI, μ. However, unlike the use of CIs, the severity analysis discriminate between inferences $\mu < \mu_1$ for different values of μ_1 within the interval. The computations related to delineating a series of observed CIs at different levels can be found in Kempthorne's "consonance intervals" (Kempthorne and Folks, 1971) and "confidence curves," "p-value functions" (Birnbaum 1961; Poole 1987). These strategies are motivated by the desire to move away from (i) having to choose a particular confidence level (or corresponding p-value), (ii) the dichotomous, "up"/"down" interpretations of tests. There would appear to be an important difference with these approaches, at least in emphasis. If one is thinking of values "consistent" with the observed data x_0, then a value μ' near the center of the CI is more in accord with x_0 than is μ'' near the upper CI bound;

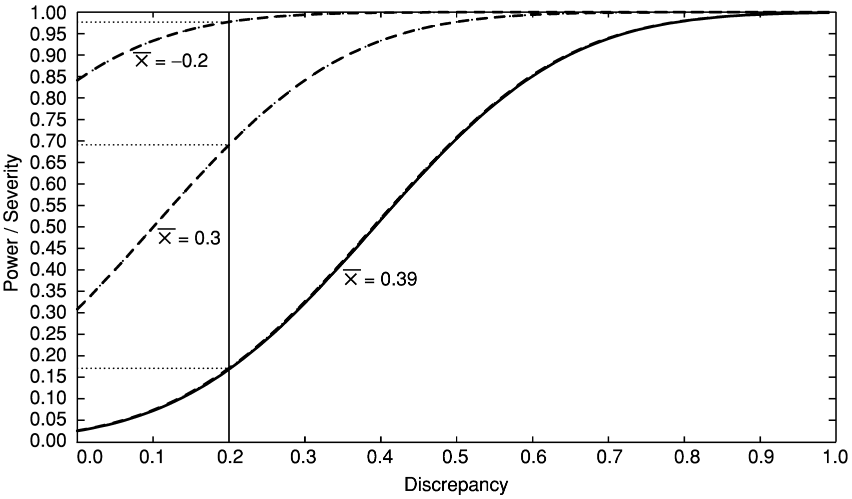


Fig. 1. The graph shows that whereas $\mathrm{POW}(\tilde{\mathrm{T}}(.025), \mu_1 = .2) = .168$, irrespective of the value of $d(\boldsymbol{x}_0)$ (or $\bar{\mathrm{x}}$); see solid curve, the severity evaluations are data-specific: for $\mathrm{d(x_0)} = 1.95(\text{or } \bar{\mathrm{x}} = .39), \mathrm{SEV}(\tilde{\mathrm{T}}(.025), \mu < .2) = .171$; for $\mathrm{d(x_0)} = 1.50(\text{or } \bar{\mathrm{x}} = .30), \mathrm{SEV}(\tilde{\mathrm{T}}(.025), \mu < .2) = .691$, and for $\mathrm{d(x_0)} = -1.0(\text{or } \bar{\mathrm{x}} = -.2), \mathrm{SEV}(\tilde{\mathrm{T}}(.025), \mu < .2) = .977$.

however the inference $\mu < \mu''$ has passed a more probative test than has $\mu < \mu'$.

Fallacies of Rejection: The Large n *Problem*

While with a nonsignif*icant* result, the concern is erroneously inferring that a discrepancy from μ_0 is absent; with a *significant* result x_0, the concern is erroneously inferring that it is present. Rejection need not be discussed separately here (see Mayo 1996), since for any H: $\mathrm{Sev}(\neg H) = 1 - \mathrm{Sev}(H)$, it follows for the particular case of H_1: $\mu > \mu_1$ $\mathrm{Sev}\ (\mu > \mu_1) = 1 - \mathrm{Sev}\ (\mu < \mu_1) = 1-$ (actual power at $(\mu = \mu_1)$.

The "large n" problem already made its splash in the second wave: With large enough sample size, an α significant rejection of H_0 can be very probable for any discrepancy α from μ_0, even if it is *substantively* trivial. Utilizing the severity assessment, an α-significant difference with n_1 passes $\mu > \mu_1$ less severely than with n_2 where $n_1 > n_2$.

Figure 2 compares test $\mathrm{T}(\alpha)$ with three different sample sizes: $n = 25$, $n = 100$, $n = 400$, denoted by $\mathrm{T}(\alpha, n)$; where in each case $d(x_0) = 1.96$ – reject at the cutoff point.

More generally, if two (otherwise identical) tests with different sample sizes give rise to rejections of H_0 at the same p-value, the result from the smaller sample experiment indicates a greater extent of a discrepancy from H_0 than from the larger. This immediately scotches the "large n problem," and simultaneously provides a way to supply p-values with assessments of population discrepancy (or effect size) that can be compared across different tests.

P-*values and Bayesian Posteriors*

Severity is an error probability calculation *based on* the actual data (and inference of interest) but it must be distinguished from what has sometimes been called the *conditional* "error probability" understood as a posterior probability. The most well-known fallacy in interpreting significance tests is to equate the p-value with a posterior probability on the null hypothesis. The p-value assessment refers only to the sampling distribution of the test statistic $d(X)$; and there is no use of priors. The Jeffrey-Good-Lindley "paradoxical" examples (see above) shows that attaining a fixed p-value, with a sufficiently large n, can correspond to large posterior probabilities for H_0. More recent work generalizes the result (Berger and Sellke 1987). Although from the degree-of-confirmation perspective, it follows that p-values come up short as a measure of evidence, the significance testers balk at the fact that use of the recommended priors can result in highly significant results being construed as no evidence against the null—or even evidence for it! An interesting twist in recent work is to try to "reconcile" the p-value and the posterior (e.g., Berger 2003).

The conflict between p-values and Bayesian posteriors often considers the familiar example of the two-sided $\mathrm{T}(2\alpha)$test, H_0: $\mu = \mu_0$ versus H_1:$\mu \neq \mu_0$. The difference between p-values and posteriors

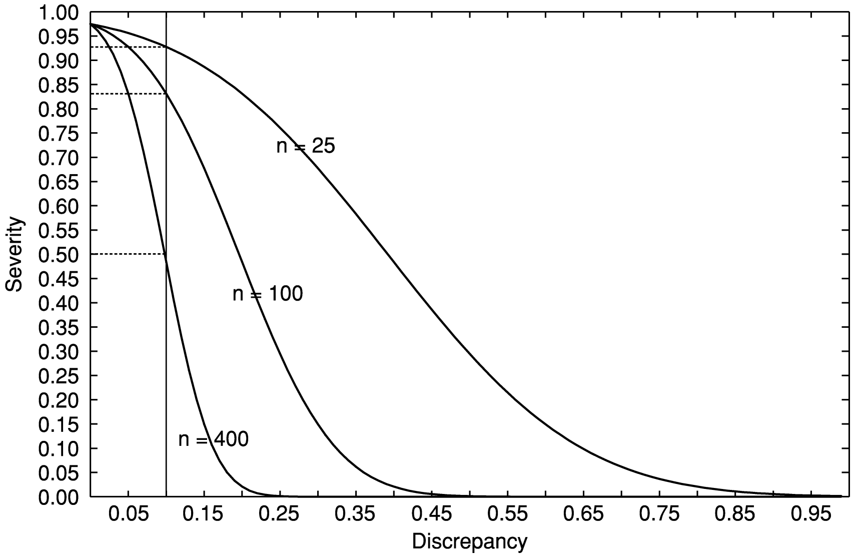


Fig. 2. In test $T(\alpha)$, ($H_0 : \mu \leq 0$ against $H_1 : \mu > 0$, and $\sigma = 1$), $\alpha = .025$, $c_\alpha = 1.96$ and $d(\boldsymbol{x}_0) = 1.96$. Inference under evaluation:
$\mu > 0.1 : \mathrm{SEV}(T(\alpha, 25), \mu = 0.1) = .93; \mathrm{SEV}(T(\alpha, 100), \mu = 0.1) = .83; \mathrm{SEV}(T(\alpha, 400), \mu = 0.1) = .5$

are far less marked with one-sided tests (e.g., Pratt 1977). "If $n = 50$ one can classically 'reject H_0 at significance level $p = .05$,' although $P(H_0|x) = .52$ (which would actually indicate that the evidence favors H_0)" (Berger and Sellke 1987, 113, replace *Pr* with *P* for consistency). Thus, data that the significance tester would regard as evidence against H_0, would, on the Bayesian construal being advocated actually indicate that the evidence favors H_0. If $n = 1000$, a result statistically significant at the .05 level leads to a posterior to the null of .82!

What makes the example so compelling to many is its use of an "impartial" or "uninformative" Bayesian prior probability assignment of .5 to H_0, the remaining .5 probability being spread out over the alternative parameter space, e.g., as recommended by Jeffreys (1939). Others charge that the problem is not *p*-values but the high prior. Moreover, the "spiked concentration of belief in the null" is at odds with the prevailing view "we know all nulls are false." Note too the conflict with CI reasoning since *0* is outside the corresponding CI.

Some examples strive to keep within the frequentist camp: to construe a hypothesis as a random variable, it is imagined that there is random sampling from a population of hypotheses, some proportion of which are assumed to be true. The percentage "initially true" serves as the prior probability for H_0. This gambit is common across all philosophy of statistics literature, and yet it commits a fallacious instantiation of probabilities:

> 50% of the null hypotheses in a given pool of nulls are true.This particular null hypothesis H_0 was randomly selected from this pool. Therefore $P(H_0 \text{ is true}) = .5$.

Faced with conflicts between error probabilities and Bayesian posterior probabilities, the error probabilist would conclude that the flaw lies with the latter measure. This is precisely what Fisher argued, and it seems fitting to end up this retrospective with a return to him.

Discussing a test of the hypothesis that the stars are distributed at random, Fisher takes the low *p*-value (about 1 in 33,000) to "exclude at a high level of significance any theory involving a random distribution" (Fisher 1956, 42). Even if one were to imagine that H_0 had an extremely high prior probability, Fisher continues—never minding "what such a statement of probability a priori could possibly mean"—the resulting high posteriori probability to H_0, he thinks, would only show that "reluctance to accept a hypothesis strongly contradicted by a test of significance" (ibid, 44) "is not capable of finding expression in any calculation of probability a posteriori" (ibid, 43). It is important too to recognize that sampling theorists do not deny there is ever a legitimate frequentist prior probability distribution for a statistical hypothesis: one may consider hypotheses about such distributions and subject them to probative tests. Indeed, if one were to consider the claim about the a priori probability to be itself a hypothesis, Fisher suggests, it would be rejected by the data!

Concluding Comment

Underlying the central points of controversy in the three waves of philosophy of statistics lie two contrasting philosophies of the role of probability in statistical inference. In one tradition, probability is used to provide a post-data assignment of degree of probability, confirmation, support or belief in a hypothesis (e.g., Bayesian and likelihood accounts); while in a second, probability is used to assess the probativeness, reliability, trustworthiness, or severity of a test or inference procedure (e.g., significance tests, N-P tests, CI). This basic contrast in underlying aims corresponds to conflicting principles for appraising methods: satisfying the likelihood principle, as opposed to controlling error probabilities. Whether statistical methodology should be regarded as supplying different tools depending on the task at hand, or whether the different methods can or should be reconciled in some way, are likely to remain questions of debate for a good while longer.

Deborah G. Mayo

The author acknowledges the helpful input of Aris Spanos and D. R. Cox.

References

Armitage, P. (1961), Contribution to the discussion in Smith, C.A.B., "Consistency in Statistical Inference and Decision," *Journal of the Royal Statistical Society, B* 23: 1–37

Berger, J. (2003), "Could Fisher, Jeffreys and Neyman Have Agreed?" *Statistical Science* 18: 1–12.

Berger, J. O., and T. Sellke (1987), "Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence," *Journal of the American Statistical Association* 82: 112–122.

Birnbaum, A. (1961), "Confidence Curves: An Omnibus Technique for Estimation and Testing Statistical Hypotheses," *Journal of the American Statistical Association* 56: 246–249.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cox, D. R., and D. V. Hinkley (1974), *Theoretical Statistics*. London: Chapman and Hall.

Edwards, W., H. Lindman, and L. Savage (1963), "Bayesian Statistical Inference for Psychological Research," *Psychological Review* 70: 193–242.

Fisher, R. A. (1935), *The Design of Experiments*. Edinburgh: Oliver and Boyd.

——— (1955), "Statistical Methods and Scientific Induction," *Journal of the Royal Statistical Society B* 17: 69–78.

——— (1956), *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

Friedman, D. (1995), "Some Issues in the Foundation of Statistics," *Foundations of Science* 1: 19–39.

Gigerenzer, G. (1993), "The Superego, the Ego, and the Id in Statistical Reasoning," in G. Keren, and C. Lewis (eds.), *A Handbook of Data Analysis in the Behavioral Sciences: Methodological Issues*. Hillsdale, NJ: Erlbaum, 311–339.

Glymour, C. (1980), *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Godambe, V., and D. Sprott (eds.) (1971), *Foundations of Statistical Inference*. Toronto: Holt, Rinehart and Winston.

Hacking, I. (1965), *Logic of Statistical Inference*. Cambridge: Cambridge University Press.

Harper, W., and C. Hooker (eds.) (1976), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science* (Vol. 2). Dordrecht, Netherlands: D. Reidel.

Howson, C., and Urbach, P. (1989), *Scientific Reasoning: The Bayesian Approach*. LaSalle, IL: Open Court.

Jeffreys, H. (1939), *The Theory of Probability*. Oxford: Oxford University Press.

Kempthorne, O., and L. Folks (1971), *Probability, Statistics, and Data Analysis*. Ames: Iowa State University.

Kyburg (1993), "The Scope of Bayesian Reasoning," in Hull, D. Forbes, L. and Okruhlik, K. (eds.), *PSA*, vol. 2. East Lansing, MI: PSA, 139–152.

Lehmann, E. L. (1993), "The Fisher and Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?" *Journal of the American Statistical Association* 88: 1242–1249.

Mayo, D. G. (1983), "An Objective Theory of Statistical Testing," *Synthese* 57: 297–340.

——— (1991), "Novel Evidence and Severe Tests," *Philosophy of Science* 58: 523–552.

——— (1996), *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

——— (1997), "Duhem's Problem, The Bayesian Way, and Error Statistics, or 'What's Belief Got To Do With It'?" and "Response to Howson and Laudan," *Philosophy of Science* 64: 222–224 and 323–333.

Mayo, D. G., and D. R. Cox (2005), "Frequentist Statistics as a Theory of Inductive Inference," Proceedings of the Second Erich L. Lehmann Symposium. *Institute for Mathematical Statistics Lecture Notes—Monograph Series* 70, forthcoming.

Mayo, D. G., and M. Kruse (2001) "Principles of Inference and Their Consequences," in D. Cornfield and J. Williamson (eds.), *Foundations of Bayesianism*. Dordrecht, Netherlands: Kluwer Academic Publishers, 381–403.

Mayo, D. G., and A. Spanos (2005), "Severe Testing as a Basic Concept in the Neyman-Pearson Philosophy of Induction," *British Journal of the Philosophy of Science* forthcoming.

Meehl, P. E. (1990), "Why Summaries of Research on Psychological Theories are Often Uninterpretable," *Psychological Reports* 66: 195–244.

Morrison, D., and R. Henkel (eds.) (1970), *The Significance Test Controversy*. Chicago: Aldine.

Neyman, J. (1935), "On the Problem of Confidence Intervals," *Annals of Mathematical Statistics* 6: 111–116.

——— (1956), "Note on an Article by Sir Ronald Fisher," *Journal of the Royal Statistical Society, B (Methodological)* 18: 288–294.

——— (1955), "The Problem of Inductive Inference." *Communications on Pure and Applied Mathematics* 8: 13–45.

——— (1957a), “Inductive Behavior as a Basic Concept of Philosophy of Science,” *Revue de l’Institut International de Statistique* 25, 7–22.

——— (1957b), “The Use of the Concept of Power in Agricultural Experimentation,” *Journal of the Indian Society of Agricultural Statistics* IX: 9–17.

Neyman, J., and E. S. Pearson (1933), “On the Problem of the Most Efficient Tests of Statistical Hypotheses,” *Philosophical Transactions of the Royal Society, A* 231: 289–337.

Pearson, E. S. (1950), “On Questions Raised by the Combination of Tests Based on Discontinuous Distributions,” *Biometrika* 37: 383–398.

——— (1955), “Statistical Concepts in Their Relation to Reality,” *Journal of the Royal Statistical Society B* 17: 204–207.

Pearson, E. S., and J. Neyman (1930), “On the Problem of Two Samples,” *Bulletin of the Academy of Political Science* 73–96, as reprinted in J. Neyman and E. S. Pearson (1967), 99–115.

Poole, C. (1987), “Beyond the Confidence Interval,” *American Journal of Public Health* 77: 195–199.

Rosenthal, R., and J. Gaito (1963), “The Interpretation of Levels of Significance by Psychological Researchers,” *Journal of Psychology* 64: 725–739.

Royall, R. (1997), *Statistical Evidence: A Likelihood Paradigm*. London: Chapman and Hall.

Salmon, W. (1967), *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

Savage, L. (ed.) (1962), *The Foundations of Statistical Inference: A Discussion*. London: Methuen.

Wald, A. (1950), *Statistical Decision Fuctions*. New York: Wiley.

Worrall, J. (1993), “Falsification, Rationality, and the Duhem Problem,” in J. Earman, A. Janis, G. Massey, and N. Rescher (eds.), *Philosophical Problems of the Internal and External Worlds, Essays on the Philosophy of Adolf Gruenbaum*. Pittsburgh: University Pittsburgh Press, 329–370.

See also **Bayesianism; Confirmation Theory; Decision Theory; Probability**

STATISTICAL MECHANICS

See **Kinetic Theory**

STRONG PROGRAM

See **Social Constructionism**

SUPERVENIENCE

The term ‘supervenience,’ as appropriated by the philosophical community, denotes a relation between two families of properties. Roughly stated, the *A*-properties supervene on the *B*-properties just in case there can be no difference in *A*-properties without some difference in *B*-properties. Equivalently, if two things are exactly alike in *B*-properties, they must be exactly alike in *A*-properties.

A simple and uncontroversial example of supervenience may help fix ideas: the case of aesthetic and nonaesthetic properties. If any two objects are exactly alike with regard to their nonaesthetic properties, they must be exactly alike with regard to their aesthetic properties; indiscernibility in nonaesthetic properties requires indiscernibility in aesthetic properties. Other normative properties display the same supervenience; for instance, any two actions that differ morally must differ in some nonmoral aspect as well.

The normative domain in fact was the original focus of philosophical musing on supervenience. The notion has more recently been found to be a useful tool in many areas of philosophy. In the philosophy of science in particular, supervenience is most prominent as one option for thinking about the ways in which various scientific domains might be related to each other. The idea that biological phenomena are entirely dependent on chemical phenomena, for example, might be clarified by talk of the supervenience of biological properties on chemical properties. Such claims are apt to be found attractive by those who want to endorse some sort of unity of science thesis without committing themselves to reductionism. Much of the contemporary discussion of supervenience has, after all, been prompted by Davidson's (1970) suggestion in his 'Mental Events' paper that supervenience might capture a kind of dependence of the mental on the physical that does not amount to reducibility. Whatever exactly reduction is, of course, is subject to considerable debate; and in certain views it is not clear that a supervenience thesis can avoid the commitment (see especially the work of Kim 1993) (see Reductionism).

The attractions of supervenience can be understood independently of any desire to avoid reductionism, however. Two aspects of the notion make for its appeal. First, when one family of properties supervenes on another, this often seems to reflect a more fundamental relationship of determination, where the supervenient properties are determined by the properties on which they supervene. The way in which these properties (sometimes called the "subvenient" properties) are distributed fixes the way in which the supervenient properties are distributed. Second, a supervenience thesis is quite minimal in character; it says very little about exactly how the one family of properties is a function of the other. This minimal character is responsible for the avoidance of reduction, but it can be counted as a virtue on its own.

Supervenience, Determination, and Explanation

Philosophers are often engaged in a distinctive explanatory project wherein they want to say that one sort of property is instantiated in virtue of the instantiation of others. A handful of typical philosophical questions makes this obvious: What makes a theory well confirmed? When does an action count as voluntary? What is it in virtue of which something has a mind? When two events are related as cause and effect, what constitutes their being thus related? The sort of explanation sought here is plainly not a causal one; this is especially evident in the last question (see Causality).

While these questions are ubiquitous, they are not very well understood. Nonetheless, one salient fact about them is this: They call for answers that would cite facts that fix or determine the fact to be explained, where this fixing is of a distinctively noncausal sort. Consequently, any proposed answer to a question of this distinctive sort carries with it an associated supervenience claim. Consider again the clear case of aesthetic supervenience: The supervenience of the aesthetic on the nonaesthetic is a reflection of the fact that whenever something has an aesthetic property, it has that property *because* of its nonaesthetic properties. Similarly, if the distribution of biological properties is indeed a result solely of the way things are chemically, then the supervenience of biological properties on chemical properties is assured.

Is this noncausal relation itself definable in terms of supervenience? Since supervenience is defined in terms that many philosophers count as relatively clear and unproblematic, such a feat would certainly help in addressing the sorts of philosophical questions at issue. Few now think, however, that any sort of supervenience is sufficient for the noncausal fixing relation in question. A certain historical irony is to be borne in mind. Philosophers' interest in supervenience can be traced back to meta-ethical discussions, especially those instigated by G. E. Moore. Famously, Moore held that ethical properties were "nonnatural"; he would deny that something's being good was constituted solely by its having certain nonethical properties. Nonetheless, he accepted the supervenience of the ethical on the nonethical. In light of this, the hope that supervenience might capture the right sort of dependence can seem fundamentally misguided.

History aside, there are independent reasons for thinking that no variety of supervenience can suffice for the determination relation in question. The simplest point to bear in mind is that supervenience is fundamentally nothing more than a holistic sufficiency relation. While no particular subvenient property is said to suffice for any particular supervenient property, each possible combination of all subvenient properties is sufficient for some possible combination of supervenient properties. Since it is well known that sufficiency by itself does not guarantee causation, it can seem unlikely that supervenience alone could guarantee the desired noncausal relation (Horgan 1993).

Even if no sufficiency relation can capture the right determination relation, the point remains that answers to the sorts of questions gestured at above seem to imply supervenience theses in a way that may help clarify the plausibility or significance of those answers.

Supervenience and Neutrality

The other aspect of supervenience that makes it appealing is its convenient neutrality, which comes into play when one generalizes over a variety of explanations of a similar sort. Consider the question of what makes something a law of nature. One may want to commit oneself to the claim that such a definition would be limited to facts about actual regularities. In doing so, one may want to avoid specifying exactly which facts about such regularities make the difference or how those facts make the difference; one may even want to avoid commitment to the claim that any particular facts about regularities can be picked out as especially salient to that determination. In this case, it is useful to clarify one's position by saying that any two worlds in which the very same actual regularities hold must be exactly alike with regard to their laws of nature—in other words, that the laws supervene on the regularities (see Laws of Nature). Supervenience thereby provides a relatively tidy and widely applicable way of capturing disputes about what determines what.

Occasional complaints about supervenience theses seem to ignore this feature. The infamous "wayward atom" example employed to raise doubts about the utility of "global" supervenience might be such a complaint (Kim 1987). Consider the thesis that any two worlds that are physically indiscernible are mentally indiscernible. This is consistent with there being a pair of worlds that differ physically in only one very small way (say, an atom in one location in the first world is displaced to a different location in the other), but differ radically in mental respects. The example is meant to cast doubt on the utility of the supervenience thesis. The proper response is to stress that the whole point of offering a supervenience thesis instead of a detailed account of which subvenient properties determine others is to avoid taking a stand on what, exactly, makes a difference. As a result, such theses will make room for bizarre theories about what makes a difference. Consider the example of laws and regularities suggested earlier. One might want to say that the laws supervene on the regularities; it is consistent with this claim, however, that there be a world differing from this one in its regularities in only the smallest fashion (say, there is just one fewer instance of the regularity that aspirin relieves headache) while differing radically in its laws of nature (say, there are no laws at all—everything happens by chance). If one wants to rule out such worlds, one needs to take a more definite stand on the initial explanatory claim, thus removing the motivation for appealing to supervenience in the first place.

An Approach to Taxonomy

One good way to approach the variety of supervenience relations is by first distinguishing three parameters that must be specified in formulating any particular supervenience thesis.

1. The *relata*: Which two families are at issue?
2. The *modal status* of the claim: When it is said that indiscernibility in subvenient properties requires indiscernibility in supervenient properties, what is the force of the requirement?
3. The mechanism of comparison: What sorts of objects are to be compared for (in)discernibility, and how exactly are the relevant pairs to be selected?

The first two parameters are relatively straightforward matters of stipulation. The options available for the third are less straightforward; the discussion will focus on these. Throughout, A will be the family of supervenient properties, and B will be the family of subvenient properties. For simplicity, the assumption will be that the modal parameter is absolute necessity, so that the range of possible worlds at issue is *all* of them.

There are two well-known options for the third parameter. One might compare particular individuals or entire possible worlds. A third option is to compare regions of space-time (Horgan 1982), but this may be left aside for the sake of space and because it introduces no fundamentally new issues.

Individual Comparison Supervenience

Say that two particular individuals are "F-indiscernible" just in case they are indiscernible with respect to family F of properties. More precisely, the notion may be defined for individuals at times as follows: x at t_1 and y at t_2 are F-indiscernible just in case, for every property P in family F, x has P at t_1 if and only if y has P at t_2.

The following is, then, the nonmodal core of what may be called *individual comparison* (IC) supervenience:

> For any individuals x and y and times t_1 and t_2, if x at t_1 and y at t_2 are B-indiscernible, then x at t_1 and y at t_2 are A-indiscernible.

Kim (1984), in his seminal "Concepts of Supervenience," has made famous three varieties of supervenience he dubs weak, strong, and global. His strong and weak supervenience relations are varieties of IC supervenience that differ in the way in which they select the individuals to be compared. The weak version compares individuals only *within* worlds, while the strong version compares individuals *across* worlds as well. Instead of weak and strong, one might call them

1. *intraworld IC supervenience*: For any possible world w, for any individuals x and y and times t_1 and t_2, if x at t_1 in w and y at t_2 in w are B-indiscernible, then x at t_1 in w and y at t_2 in w are A-indiscernible; and
2. *crossworld IC supervenience*: For any possible worlds w_1 and w_2, for any individuals x and y and times t_1 and t_2, if x at t_1 in w_1 and y at t_2 in w_2 are B-indiscernible, then x at t_1 in w_1 and y at t_2 in w_2 are A-indiscernible.

Intraworld IC supervenience ensures that any two objects within a given world are A-indiscernible if B-indiscernible. This is consistent with there being a pair of objects drawn from two distinct worlds that are B-indiscernible yet A-discernible. Crossworld IC supervenience rules out this latter situation.

Few philosophers have found occasion to appeal to an intraworld IC supervenience thesis. It is not difficult to see why. Suppose a supervenience thesis is wanted because it is thought that the A-properties are always instantiated in virtue of B-properties. Suppose there is only one A-property P_A and only one B-property P_B; now consider two worlds (Figure 1) such that each contains only two objects o_1 and o_2 existing for a span of time during which they undergo no change, having at all moments the following array of properties:

w_1	w_2
o_1 has P_B and P_A.	o_3 has P_B and lacks P_A.
o_2 lacks P_B and lacks P_A.	o_4 lacks P_B and has P_A.

Fig. 1. Intraworld without crossworld 1C supervenience.

The intraworld IC thesis is consistent with the existence of these worlds. Yet if objects o_1 and o_2 have their A-properties solely in virtue of their B-properties, it is hard to see what could explain the difference between the distribution of A-properties in w_1 and w_2. One presumably wants to rule out such a pair of worlds, but an intraworld IC thesis will not rule it out.

Crossworld IC theses, by contrast, may seem to rule out too much. Suppose the A-properties are the aesthetic properties pertaining solely to paintings, and the B-properties are all the visual properties an object might have. One might be tempted to the thesis that when something has an A-property it has it solely in virtue of its B-properties and subsequently venture the crossworld IC supervenience thesis: Any pair of individuals, drawn from the same or differing worlds, are A-indiscernible if B-indiscernible. Suppose, however, that perfect forgeries are aesthetically inferior to originals, so that the history of a painting makes a difference to its aesthetic value. If so, then the crossworld IC supervenience thesis is false: There could be two paintings that are visually indiscernible that could be aesthetically discernible—whether drawn from the same world or not.

It does not seem fair to blame this sort of problem on the choice of a crossworld IC supervenience relation. The problem, rather, seems to be the choice of supervenience *relata*: The subvenient family is insufficiently broad. A natural fix is to expand that family to include relational properties such as having such-and-such a history. Nonetheless, examples of this sort have directed attention to global supervenience theses—that is, supervenience theses that compare entire worlds.

A good question is whether every supervenience thesis can be, as it were, hammered into a crossworld IC form by building into the subvenient family more complex relational properties. It is clear that some cases could be thus accommodated only if one were willing to include very gruesome-looking subvenient properties. Beyond those—which may be especially unlikely to arise in the philosophy of science—there are others that demand a global treatment, *viz.,* those in which the properties in question are not even properly said to be possessed by individuals in a world but only by the worlds themselves. The law thesis considered earlier illustrates this. If one sets out the thesis that the laws of nature supervene on the regularities, one must take care to note that the two families of properties can be attributed only to worlds: being such that all *F*s are *G*s, or being such that it is a law that all *F*s are *G*s.

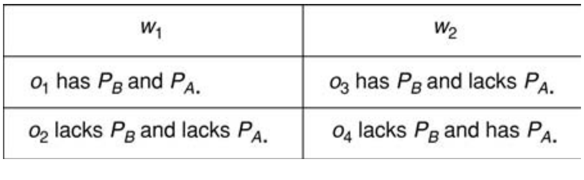
Global Supervenience

The sort of supervenience that compares entire worlds is usually called global supervenience. The claim that *A*-properties globally supervene on *B*-properties can be initially expressed by saying that any two possible worlds that are *B*-indiscernible are also *A*-indiscernible. If, however, the two families of properties include properties that are normally attributed to individuals, talk of two *worlds* being indiscernible with respect to those properties is not immediately comprehensible.

The most straightforward way to make sense of such talk presumes that the two worlds contain the very same individuals. In that case, one may say that w_1 and w_2 are *F*-indiscernible just in case, for each individual x that exists in w_1 and each time t, x at t in w_1 and x at t in w_2 are *F*-indiscernible. However, if one endorses a global supervenience thesis using this notion of global indiscernibility, one should be aware of one unhappy way in which it is weaker than might have been expected. Recall the scenario involving just two objects and two properties earlier set out in demonstrating the weakness of intraworld IC supervenience. Now consider a similar one (Figure 2), with the sole difference being that the two worlds contain numerically distinct individuals.

If the claim that *A*-properties were instantiated in virtue of *B*-properties earlier motivates rejecting the first scenario, it should presumably motivate ruling out this one as well. (A possible exception is when the supervenient properties include impure properties—those defined by reference to particular individuals.)

What is wanted is a one-one mapping between worlds that does not require numerical identity. There are two different routes one might take here. The first is to impose, from the outside as it were, a way of establishing what counts as a relevant mapping; the other is to allow the character of the subvenient family to do the work. As for the first route, one natural suggestion is to appeal to spatiotemporal location. One might define the relevant mapping between w_1 and w_2 as that which ensures that each individual x in w_1 is mapped to an individual in w_2 that exists at the same times and places in w_2 as does x in w_1. (Note that if spatiotemporally coincident entities can exist, this approach will not result in a unique mapping.)



w_1	w_2
o_1 has P_B and P_A.	o_3 has P_B and lacks P_A.
o_2 lacks P_B and lacks P_A.	o_4 lacks P_B and has P_A.

Fig. 2. *B*-discernibility by means of numerical supervenience.

If one adopts such a spatiotemporal mapping, it is presumably because one wants to allow that the spatiotemporal location of an individual is relevant to how its *B*-properties determine its *A*-properties. In general, however, the factors that one thinks relevant to the determination of the supervenient properties are meant to be built into the subvenient family. Of course, this means one option one might want to take is to return to a crossworld IC thesis with an appropriately expanded subvenient family. For example, if one thinks that spatiotemporal location is relevant, one may include in the subvenient family properties *encoding* such location—the property of having existed at such-and-such places and at such-and-such times. But if one prefers a global thesis, this point suggests an alternate approach. One might use only those mappings that match individuals in one world to those in another according to their being indiscernible with respect to those properties deemed relevant in determining the facts about the supervenient properties. These will, of course, be exactly those properties placed in the subvenient family. If one has included in the *B*-family those factors that one believes relevant to determining the instantiation of *A*-properties, it is natural to suppose that a relevant mapping is one that pairs up *B*-indiscernible individuals.

More precisely: Say that a one-one mapping *M* from w_1 to w_2 is a "*B*-isomorphism" just in case, for any property *P* in *B*, if an individual x in w_1 has *P* at a time t, then $M(x)$ in w_2 has *P* at t. One may then want to say that two worlds are *F*-indiscernible just in case there exists an *F*-isomorphism between them. Using this definition of indiscernibility to define global supervenience in the usual way results in this formulation:

> For any two possible worlds w_1 and w_2, if there is a *B*-isomorphism between w_1 and w_2, then there is an *A*-isomorphism between w_1 and w_2.

This thesis is much weaker than one might have expected. Indeed, it is compatible with the pair of worlds considered earlier as a way of showing how weak intraworld IC supervenience is (see Figure 1). In that case, there is a *B*-isomorphism between w_1 and w_2, i.e., the function that maps o_1 in w_1 to o_1 in w_2 and o_2 in w_1 to o_2 in w_1. There is also an *A*-isomorphism, i.e., the function that maps o_1 in w_1 to o_2 in w_1 and o_2 in w_1 to o_1 in w_2. If allowing this scenario made intraworld IC supervenience too weak to be of interest, then the present global thesis is also too weak to be of interest.

The sort of supervenience just considered has been dubbed "weak global supervenience" (Sider 1999; Shagrir 2002). Its weakness lies in the fact that it quantifies over two one-one mappings without relating them to each other. By contrast, the following "strong global supervenience" thesis relates them quite closely:

> For any two possible worlds w_1 and w_2, every *B*-isomorphism between w_1 and w_2 is also an *A*-isomorphism.

This latter thesis rules out the pair of worlds described in Figure 1 and seems to be the most promising notion of global supervenience.

D. Gene Witmer

References

Beckermann, Ansgar, Hans Flohr, and Jaegwon Kim (eds.) (1992), *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. Berlin: De Gruyter.

Chalmers, David (1996), *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Charles, David, and Kathleen Lennon (eds.) (1992), *Reduction, Explanation and Realism*. Oxford: Oxford University Press.

Davidson, Donald (1970), "Mental Events," in L. Foster and J. W. Swanson (eds.), *Experience and Theory*. Amherst: University of Massachusetts Press.

Horgan, Terence (1982), "Supervenience and Microphysics," *Pacific Philosophical Quarterly* 63: 29–43.

——— (1993), "From Supervenience to Superdupervenience: Meeting the Demands of a Material World," *Mind* 102: 555–586.

Kim, Jaegwon (1984), "Concepts of Supervenience," *Philosophy and Phenomenological Research* 45: 153–176.

——— (1987), "'Strong' and 'Global' Supervenience Revisited," *Philosophy and Phenomenological Research* 48: 315–326.

——— (1993), *Supervenience and Mind*. Cambridge: Cambridge University Press.

McLaughlin, Brian (1995), "Varieties of Supervenience," in Elias Savellos and Ümit Yalçin (eds.), *Supervenience: New Essays*. Cambridge: Cambridge University Press, 16–59.

——— (1997), "Supervenience, Vagueness and Determination," in James Tomberlin (ed.), *Philosophical Perspectives 11*. Cambridge: Blackwell, 209–230.

Melnyk, Andrew (1997), "On the Metaphysical Utility of Claims of Global Supervenience," *Philosophical Studies* 87: 277–308.

Paull, R. Cranston, and Theodore Sider (1992), "In Defense of Global Supervenience," *Philosophy and Phenomenological Research* 52: 833–854.

Post, John (1995), "'Global' Supervenient Determination: Too Permissive?" in Elias Savellos and Ümit Yalçin (eds.), *Supervenience: New Essays*. Cambridge: Cambridge University Press, 73–100.

Savellos, Elias, and Ümit Yalçin (1995), *Supervenience: New Essays*. Cambridge: Cambridge University Press.

Shagrir, Oron (2002), "Global Supervenience, Coincident Entities and Anti-Individualism," *Philosophical Studies* 108: 171–195.

Sider, Theodore (1999), "Global Supervenience and Identity Across Times and Worlds," *Philosophy and Phenomenological Research* 59: 913–937.

Stalnaker, Robert (1996), "Varieties of Supervenience," in James Tomberlin (ed.), *Philosophical Perspectives 10*. Cambridge: Blackwell, 221–241.

See also **Reductionism; Unity of Science Movement**

T

TACIT KNOWLEDGE

See **Experiment**

TAXONOMY

See **Individuality; Species**

TECHNOLOGY

See **Experiment**

TELEOLOGY

See **Explanation; Function**

TESTABILITY

See **Cognitive Significance; Demarcation, Problem of; Verifiability**

THEORIES

Scientific theories, as carriers of scientific knowledge, were at the focus of philosophy of science throughout the twentieth century. In particular, philosophers argued about how scientific theories should be formalized or reconstructed. At first, only the most successful scientific theories (such as Newtonian theory, relativity theory, and quantum theory in physics, and evolutionary theory in biology) were considered as candidates for reconstruction; more recently a much wider range of scientific constructs has come under consideration. This article will review the three most prominent approaches to the reconstruction of scientific theories in the philosophy of science. Each attempts to portray the content of theories using formalisms of various kinds. Criticisms of the approaches are also reviewed. First to be discussed is the leading approach to scientific theories in the first half of the twentieth century.

The Received View

Logical empiricist approaches to theory structure are known as the *received view* or the *syntactic view* (see Logical Empiricism). These "syntactic" approaches characterize scientific theories as linguistic entities. They explicate scientific theories as being axiomatic logical systems with a set of rules of interpretation or of correspondence tying the theoretical language in the system to the observation language. Suppe (1977) has given a useful characterization of the later versions of the logical empiricists' view, as presented by Hempel (1958) and Carnap (1956). Under this view, theories can be reconstructed as satisfying the following conditions:

1. The theory is formulated in terms of a first-order language L and a logical calculus K defined in terms of L. (L may be expanded by modal operators.)

2. The terms of L are divided into two exclusive categories: V_O, which contains just the observation terms, and V_T, which contains the theoretical terms.
3. This allows the language L to be divided into two parts: the observation language, L_O, which contains terms from V_O but no terms from V_T, with an associated logical calculus; and the theoretical language, L_T, which does not contain any V_O terms, also with an associated logical calculus. In addition, L contains mixed sentences, in which at least one term from both V_T and V_O occurs.
4. L_O and its associated logical terms, K_O, are given a semantic interpretation, in which the domain of interpretation consists of concrete observable events or things. This interpretation is a partial semantic interpretation of L and K.
5. Central to the reconstruction of the theory T in L are a special set of sentences of L_T, which express the axioms or laws of the theory. The axioms of the theory T, in which only terms of V_T occur, and the correspondence rules, C, which are mixed sentences, make up a partial interpretation of the theoretical terms and the sentences of L. The correspondence rules must be finite in number and logically compatible with T. In addition, each rule in C must contain at least one V_O term and one V_T term, and C must contain no extralogical term that does not belong to either V_T or V_O.

In summary, if L is the language and A is the conjunction of all the axioms of the theory, while C is the set of correspondence rules, the conjunction $T \wedge A$ is the scientific theory.

This last version of the received view differs from earlier versions. In particular, earlier notions of the correspondence rules were construed in terms of explicit definitions that identified the content of theoretical claims with complex observation conditions. Later, correspondence rules were formulated as reduction sentences that partially defined theoretical language in terms of particular experimental setups. Finally, as presented above, correspondence rules were seen as interpretive systems; under this approach, theoretical terms were not coordinated with individual observable conditions. Rather, the inclusion of theoretical terms in the theory had to make a difference to the theory's observable consequences.

Note that the syntactic view characterizes theories both syntactically and semantically. The syntactic characterization consists in the axiomatic calculus. Theorems of the theory are derived from the set of axioms. The syntactic specification is followed by a semantic interpretation, which involves an interpretation of the signs of the axiomatic calculus in terms of various empirical entities and properties.

The received view of theory structure has been used to represent various pieces of physical theory, including Newtonian mechanics, as well as economic theory and evolutionary theory (see Suppe 1989 for references).

Criticisms of the Received View

Although the received view had the merit of being relatively clear, its clarity, which it owed in large part to its association with the standards and ideas of formal logic, came at a price. The account came in for a number of criticisms from the 1960s onward. A sample of the most serious are listed below.

As noted above, correspondence rules require the use of both an observation term and a theoretical term. These parts of the language of theories were considered distinct and nonoverlapping. But the received view's distinction between observational and theoretical language was sharply criticized by Achinstein and Putnam in the 1960s Hanson and Feyerabend also attacked the logical empiricists' view, arguing that observations were theory-laden and that the observational/theoretical distinction was untenable. Kuhn argued that the connections between theory and phenomena could not be represented by explicit correspondence rules. He favored the view that exemplars of applications of theories were the basis for much of science (see Suppe 1977, 4).

Putnam and Achinstein also criticized the notion of partial interpretation as it was used in the liberalized correspondence rules. These rules were also criticized by Schaffner (1969), who attacked the logical empiricist view by arguing that it merged together experimental procedures, meaning, and causal relations. The received view was also criticized for failing to individuate theories correctly. Under the syntactic definition of theories, any change in the syntax of a theory is taken to introduce a new theory (Suppe 1977).

A general problem with the received view of theories is that the Löwenheim-Skølem theorem implies the existence of unintended models of any theory, which is a result of the use of the first-order languages that the logical empiricists insisted upon. For example, first-order theories of arithmetic on the natural numbers have uncountable models. These unintended models were, in turn, sources of potential counterexamples. As Suppe (1998) puts it, "Positivistic syntactical analyses of theories, confirmation and explanation persistently were plagued by problems of unintended models" (345) (this paper also has a useful and large bibliography concerning theories).

To make matters worse, standard formulations of most theories do not take the form of axiomatic calculi with attached interpretations. In fact it is doubtful whether some theories can be formalized in this manner. Van Fraassen (1980, 65–67) emphasizes the "enormous distance" between research on the foundations of science and any "syntactically capturable axiomatics." In addition to these pragmatic concerns, it has been argued (French and Ladyman 1999, following van Fraassen 1989, 211) that the received view is the wrong way to represent scientific theories altogether. It is not merely practically unworkable, it is technically impossible, in the sense that any scientific theory that makes use of the real numbers (i.e., theories that take space-time to be continuous) will not be axiomatizable in any first-order language. Given that the majority of scientific theories do make use of the real numbers, this result constitutes a major difficulty for the received view.

In sum, the received view has been beset since the 1960s with objections to the possibility of formalizing theories in the preferred manner. The important distinction between theoretical and observational vocabularies turned out to be untenable, as was the adherence to first-order languages. The problem of having unintended interpretations of the axiomatized theory haunted these accounts; a great deal of energy was spent simply trying to eliminate the unintended interpretations. These problems were taken to be serious by many philosophers of science, and when an alternative came along, the received view quickly fell into disfavor.

The Semantic View

The semantic view was originally developed by Suppes (1957 and 1967), and further developed by Suppe (1977), van Fraassen (1970 and 1972), and Giere (1988). Suppes proposed that when analyzing a theory, one directly specifies the intended models, without reference to a particular axiomatic calculus. In other words, one should characterize theories as specifications of the kinds of systems to which they can be applied. As van Fraassen (1972) puts it,

> The essential job of a scientific theory is to provide us with a family of models, to be used for the representation of empirical phenomena. On the one hand, the theory defines its own subject matter—the kinds of systems that realize the theory; on the other hand, empirical assertions have a single form: the phenomena can be represented by the models provided (310).

This fixes one problem with the logical empiricist account, namely, the problem of unintended models, but it also raises the issue of what a model is. Technically, a semantic model is an interpretation that makes all sentences in a theory true. But in the semantic account, the mapping relation is implicit, and the model is conceived as an independent structure. As van Fraassen (1989) makes clear, "A model consists, formally speaking, of entities and relations among those entities" (365).

The semantic approach is really a family of related approaches, all having two assumptions in common: first, that scientific theories are best conceived not primarily as axiomatized linguistic systems, but rather in terms of their models; and second, that the appropriate tool for the formal explication of scientific theories is not first-order logic and metamathematics, but rather mathematics. Advocates of the semantic view differ in the mathematics they use to present the models that make up theories. Choices range from the state space approach taken by van Fraassen (1989) and Suppe (1989) to set theory (Suppes [1962], Sneed [1971], Stegmuller [1976]) to set theory supplemented by category theory (Balzer, Moulines, and Sneed 1987). These latter two approaches are characteristic of a branch of the semantic view called the structuralist approach, which shall be treated in its own section below.

Models

In Suppe's and van Fraassen's views, models are presented by specifying a state space and the laws and parameters on it. The various variable values (or states of the system) are limited by the laws of the model, which portray either the possible states of the system (coexistence laws) or the successive states of a system over time (laws of succession). The structure of the theory is thus presented as a family of related state space models and their laws.

Van Fraassen (1980) differentiates between the model-theoretic use of "models" and scientists' use of the term. Model-theoretic models are "specific structures, in which all relevant parameters have specific values," while a physicist's model is "a type of structure, or class of structures, all sharing certain general characteristics" (44). And what physicists call a model is what van Fraassen would call a *model-type*. (See McMullin 1985, 257, on different uses of the term "model"; also Achinstein 1965.)

Although his approach has great overlap with other semantic-view theorists, for Giere (1988, 47–48) a model is both a structure *and* an interpretation, a mapping from elements of a linguistic formulation to elements of the structure. Other semantic-view theorists do not include the interpretation as part of the model, and see models only as structures that satisfy the linguistic formulation. Giere's approach is also distinctive in the way it characterizes the relation between the model and the system being modeled. In his formulation of the semantic view, models must be "similar" to the systems that they represent in nature. Scientific models are understood as idealized systems (and thus as abstract entities), while "real systems" refer to nature. The relationship between models and real systems is one of similarity in certain respects and to certain degrees. A theory is characterized by the models it uses, and a theoretical hypothesis asserts a similarity between a real system and some aspects of one of these models. Giere is interested in the similarity relation because it captures the use of idealization and approximatizations in scientific modeling. Besides, one cannot claim isomorphism between mathematical structures on one hand and real systems in nature on the other; it would be a category mistake.

The role of a hierarchy of models (Suppes 1967) that intervenes between the system in nature and the theoretical structure is especially important here. Suppes (1962) emphasizes that there is a hierarchy of representations between theories and phenomena. As van Fraassen (1985) puts Suppes' point, the "theory is not confronted with raw data but with models of the data and . . . the construction of these data models is a sophisticated and creative process" (271). As described by Suppes (1967, 62–63), the hierarchy of models constitutes a methodology for the transformation of observational and experimental data into a form that can be compared with theory. He includes models of the experiment, models of data, experimental design, and *ceteris paribus* conditions. In this view, data models are the low-level structural representations of the natural world. As such, they involve processed information regarding the natural system that is then formulated into a mathematical structure. It thus becomes possible to compare various models in the hierarchy with one another, because they all represent mathematical structures. Giere's similarity relation can be understood in terms of a subfamily of relations that stand in correspondence to certain of the relevant family of relations in the total model that completely represents the system (French and Ladyman 1999, 111).

There is still a problem of how to understand the relationship between the lowest-level data model and the world, but this problem is shared by all systems of representation and is not unique to the semantic approach. At any rate, the notion of similarity can be understood in terms of a partial isomorphism holding between the families of relations concerned. Bueno (1997) proposed a formalization of the hierarchy of partial structures. Thus, as French and Ladyman (1999) conclude, "The relationship between theory and empirical reality is mediated by a series of representations and so the use of isomorphism and related notions is perfectly legitimate" (113). What binds the Suppes, Suppe, Giere, and van Fraassen views together is the overall claim that theories are best thought of as families of models, rather than as partially interpreted axiomatic systems.

Several advantages of the semantic approach have been claimed. First, theories are "extralinguistic entities which may be described or characterized by a number of different linguistic formulations" (Suppe 1974, 221). In other words, the particular linguistic formulation does not affect the content of the theory. The semantic approach thus avoids the linguistic puzzles and problems of unintended models that dominated discussions within the received view of theories. Second, the semantic view is argued to be more easily and naturally applied to scientific theorizing than the received view. This can be seen in the numerous applications of the semantic view to various scientific theories, including quantum mechanics (van Fraassen 1991), evolutionary theory (Beatty 1980; Lloyd 1988; Thompson 1989), economic theory (Hausman 1981), ecological theory (Castle 2001), chaos theory (Kellert 1992), and sex and gender (Crasnow 2001).

Third, proponents of the semantic view claim that it is compatible with any epistemology of science, whether realist, empiricist, or instrumentalist. The received view, in contrast, was taken by many to be tied to the antimetaphysical ontology of logical empiricism. Finally, the semantic view "offer[s] the possibility of incorporating the various senses of the word 'model', as used in scientific practice, within a single, unitary account" (French and Ladyman 1999, 106). This is done through the semantic-view argument that being "true in some respects" can be characterized in terms of "partial" truth (Da Costa and French 1990, 259). The proponents of the semantic view argue that this accords well with scientific practice and captures various senses of the term "model" in science.

Before turning to challenges to the semantic view, something must be said about a variant of a semantic approach, the structuralist view.

The Structuralist View

The structuralist view was launched by Sneed (1971), in his set-theoretic analysis of physical theory. The view has since been elaborated, and numerous applications have been made to various scientific theories (Balzer, Sneed, and Moulines 2000). One of the fundamental aspects of the approach is that the close association of system specifications and empirical applications is taken into account. Theories are characterized in terms of exemplary or paradigmatic applications of the system that is specified. A theory consists of a structure and a set of intended applications.

Under the structuralist view, the first step in analyzing the structure of a scientific theory is to start by presenting a class of models in set-theoretic terms. The class of structures that settle the formal properties of the scientific concepts in the theory are called potential models, or M_p. The class of structures that also satisfy the substantial laws of the theory are called actual models, or M. The structuralists determine M_p and M by defining a set-theoretic predicate by means of the method of axiomatization. But the theory's identity is given by M, and not by the set-theoretical predicate. The theory's identity is associated with the theory's *empirical claims*, which start with a certain domain of empirical systems being investigated, called the domain of intended applications, I. A theory's central empirical claim, under structuralism, is that a certain M_p-conceptualized domain I can be subsumed under M. Thus, the claim has empirical content, and can be true or false. In this sense, structuralism also involves statements.

The "formal core" K of a theory is an array of structures that include M_p, M, partial potential models M_{pp}, constraints C, links with other theories L, and approximations A. Thus, the refined view of the empirical claim of a theory is that it consists in the global statement that the domain of intended applications I can be subsumed under K. Many intertheoretic relations can be represented within the structuralist approach, including specialization, reduction, equivalence, and approximation (Moulines 2002; for detailed presentations of the structuralist view, see Stegmuller 1979 and Balzer et al. 1987).

The structuralist approach is intended to be useful in addressing relationships among theories, "their connections to empirical data, the methods used to check them, the pragmatic aspects of their use by the scientific community, their evolution in historical time, and related matters" (Moulines 2002, 2). This approach has been especially important in work on intertheoretic reduction (Stegmuller 1976; Balzer, Pearce, and Schmidt 1984; Balzer et al. 1987). The accounts of intertheoretic reduction are regarded by some as having proved useful in very detailed reconstructions of episodes in the history of science (Balzer et al. 1984 and 1987). For example, Bickle (1993) uses this account of intertheoretic reduction to analyze the connectionist eliminitavist arguments (see Reductionism).

While both the semantic view and the structuralist view are "variants of semantic metatheory," the distinction between M_p and M_{pp} is treated differently by the two approaches (Sintonen 1990, 679). In addition, structuralists have argued that mature scientific theories form "theory nets," which are theories related to one another by certain intertheory relations, whereas semantic-view theorists do not make this claim.

Finally, the structuralists read Suppe and van Fraassen as denying any role for syntactic analysis. A more substantive disagreement with the other variants of the semantic view is that structuralists think that pragmatic analysis plays an important role in understanding the theoretical structure of science, whereas semantic-view theorists do not use pragmatic analysis.

Criticisms of the Semantic View

Despite the fact that the semantic view is today more or less the accepted view, it is not without its critics. There are at least three major areas of concern: Is the semantic view as free of the linguistic entanglements as it claims to be? Can the semantic view really account for the diversity of models seen in science? and, Does the semantic view misrepresent the place of models in science?

Hendry and Psillos (1998) note that the "semantic view" seems to be a misnomer, as theories are not really nonlinguistic in the semantic conception. Giere's theoretical hypotheses, wherein abstract structures are claimed to represent given classes of real systems, are linguistic entities. If models are seen purely abstractly, the theory's content is divorced from its applications, and theories are not representational. Thus, the theory cannot be seen as equivalent simply to a class of models. Giere's formulation avoids this problem and is endorsed by van Fraassen (1989, 222), and Suppe (1989, 4), but only by allowing linguistic components into the theory. In other words, the mapping claims made on behalf of the models are themselves linguistic in form.

A pervasive criticism of the semantic view is that it cannot account for the wide array of models used in scientific practice. The semantic view uses only a narrow logical meaning of modeling, while there are many other types of models in scientific use, such as scale models, analogue models, iconic models, Watson and Crick's wire model of DNA, Kuhnian exemplars, and Griesemer's biological "remnant" models, which are actual specimens in a museum (Griesemer 1990) (see Scientific Models). Downes (1992) argues that since these models cannot be accommodated within the semantic approach, its strong form should be rejected (Weinert 1999). This is an echo of Achinstein (1968), who rejected the semantic approach to theories, arguing that there could be no theory of scientific models.

The claim is that there is a difference between the mathematical structures that make up the model in the semantic view and scientific models more broadly. The concept of isomorphism is especially problematic, because while isomorphism may easily be applied to mathematical models, it is unclear how it could apply to the relationship between empirical systems and theoretical models (Downes 1992, 147). But French and Ladyman (1999, 107) take it that the challenge is whether a set-theoretical description can capture the kinds of models used in scientific practice. They argue that this is possible and demonstrate how to do it with various types of models. This usually involves a partial isomorphism between the partial structures used in representing the models, where a partial structure is a set-theoretical structure that satisfies a Suppes-style set-theoretical predicate. This approach even works with DNA models and remnant models (or so it is claimed), in that they represent structure, which can be formalized in the appropriate way.

An additional worry about models and the semantic view, though, is that models are relatively independent of theories (Morgan and Morrison 1999). When constructing models, more than theories are relied upon—model construction also involves data, technological design, and intuitive insight (Cartwright, Shomar, and Suarez 1995). Because models are relatively independent of each source, they can serve as mediators between them. In addition, it is possible to have a number of possibly mutually inconsistent models of a system or phenomenon covered by a theory. Morgan and Morrison (1999) emphasize that constructing models and manipulating them are important scientific activities that are not adequately addressed by a view of models purely in terms of their relation to theories. This is not actually a direct criticism of the semantic view, but it does emphasize the limits of its utility in understanding how models work in science.

Conclusion

The most fundamental implicit criticism is that perhaps theory structure is not the thing to study in the philosophy of science, and that more analytical effort should be spent on looking at the various uses of scientific models. Nevertheless, it would seem that the semantic view of theory structure has been applied successfully in a number of cases, and is still generating more solutions than problems. The same cannot be said of the logical empiricists' approach to theory structure, which collapsed under its own set of formal problems. The semantic view is widely accepted presently, and it is unclear that any of its critics have offered new ways to characterize scientific theories. There is a promising sketch of an alternative, "interactionist view" by Hendry and Psillos (1998): "As historical individuals, theories are complex consortia of different representational media: words, equations, diagrams, analogies and models of different kinds" (1). According to Hendry and Psillos, both the semantic and received views fail to tackle the complexity of the ways that scientific theories represent the world, which can only be represented by these complex groupings of representational media. So far, though, no developed alternative to viewing theories has been offered outside of the semantic and structuralist approaches.

Elisabeth A. Lloyd

References

Achinstein, P. (1965), "Models, Analogies, and Theories," *Philosophy of Science* 31: 329–350.

——— (1968), *Concepts of Science: A Philosophical Analysis*. Baltimore: Johns Hopkins University Press.

Balzer, W., D. A. Pearce, and J.-J. Schmidt (1984), *Reduction in Science*. Dordrecht, Netherlands: Reidel.

Balzer, W., C. Moulines, and J. D. Sneed (1987), *An Architectonic for Science*. Dordrecht, Netherlands: Reidel.

Balzer, W., J. D. Sneed, and C. Ulises Moulines (2000), *Structuralist Knowledge Representation: Paradigmatic Examples*. Amsterdam: Rodopi.

Beatty, J. (1980), "Optimal-Design Models and the Strategy of Model Building in Evolutionary Biology," *Philosophy of Science* 47: 532–561.

Bickle, J. (1993), "Connectionism, Eliminitavism, and the Semantic View of Theories," *Erkenntnis* 39: 359–382.

Bueno, O. (1997), "Empirical Adequacy: A Partial Structures Approach," *Studies in the History and Philosophy of Science* 28: 585–610.

Carnap, R. (1956), "The Methodological Character of Theoretical Concepts," in H. Feigl and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 1. Minneapolis: University of Minnesota Press.

Cartwright, N., T. Shomar, and M. Suarez (1995), "The Tool Box of Science," in W. E. Herfel, W. Krajewski, I. Niiniluoto, and R. Wojcicki (eds.), *Theories and Models in Science: Poznan Studies in the Philosophy of the Sciences and the Humanities*. Amsterdam: Rodopi, 137–149.

Castle, D. G. A. (2001), "A Semantic View of Ecological Theories," *Dialectica* 55: 51–65.

Crasnow, S. L. (2001), "Models and Reality: When Science Tackles Sex," *Hypatia* 16: 138–148.

Da Costa, N. C. A., and S. French (1990), "The Model-Theoretic Approach in the Philosophy of Science," *Philosophy of Science* 57: 248–265.

Downes, S. M. (1992), "The Importance of Models in Theorizing: A Deflationary Semantic View," in *PSA 1992*, vol. 1. East Lansing, MI: Philosophy of Science Association, 142–153.

French, S., and J. Ladyman (1999), "Reinflating the Semantic Approach," *International Studies in the Philosophy of Science* 13: 103–121.

Giere, R. (1988), *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Griesemer, J. R. (1990), "Modeling in the Museum: On the Role of Remnant Models in the Work of Joseph Grinell," *Biology and Philosophy* 5: 3–36.

Hausman, D. (1981), *Capital, Profits, and Prices: An Essay in the Philosophy of Economics*. New York: Columbia University Press.

Hempel, C. (1958), "Theoretician's Dilemma," in H. Feigl, M. Scriven, and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 2. Minneapolis: University of Minnesota Press.

Hendry, R. F., and S. Psillos (1998), *Theories as Complexes of Representational Media*, at www.umkc.edu/scistud/psa98/papers/hendry.pdf.

Hughes, R. I. G. (1989), *The Structure and Interpretation of Quantum Mechanics*. Cambridge, MA: Harvard University Press.

Kellert, S. H. (1992), "A Philosophical Evaluation of the Chaos Theory," *PSA 1992*, vol. 2. East Lansing, MI: Philosophy of Science Association, 33–49.

Lloyd, E. A. (1988), *The Structure and Confirmation of Evolutionary Theory*. New York: Greenwood Press.

McMullin, E. (1985), "Galilean Idealization," *Studies in History and Philosophy of Science* 16: 247–273.

Morgan, M., and M. Morrison (eds.) (1999), *Models as Mediators*. Cambridge, UK: Cambridge University Press.

Moulines, C. (2002), "Introduction: Structuralism as a Program for Modeling Theoretical Science," *Synthese* 130: 1–11.

Schaffner, K. (1969), "Correspondence Rules," *Philosophy of Science* 36: 280–290.

Sintonen, M. (1990), "Darwin's Long and Short Arguments," *Philosophy of Science* 57: 677–689.

Sneed, J. D. (1971), *The Logical Structure of Mathematical Physics*. Dordrecht, Netherlands: Reidel.

Stegmuller, W. (1976), *The Structure and Dynamics of Theories*. New York: Springer-Verlag.

——— (1979), *The Structuralist View of Theories: A Possible Analogue of the Bourbaki Programme in Physical Science*. Berlin: Springer-Verlag.

Suppe, F. (1974), "The Search for Philosophic Understanding of Scientific Theories," in F. Suppe (ed.), *The Structure of Scientific Theories*. Urbana: University of Illinois Press.

——— (1977), *The Structure of Scientific Theories*, 2nd ed. Urbana: University of Illinois Press.

——— (1989), *The Semantic Conception of Theories and Scientific Realism*. Urbana, IL: University of Illinois Press.

——— (1998), "Theories, Scientific," in *Routledge Encyclopedia of Philosophy*. New York: Routledge, 344–355.

——— (1957), *Introduction to Logic*. New York: Van Nostrand.

Suppes, P. (1962) "Models of Data," in E. Nagel, P. Suppes, and A. Tarski (eds.), *Logic, Methodology and the Philosophy of Science: Proceedings of the 1960 International Congress*. Stanford, CA: Stanford University Press, 252–267.

——— (1967), "What Is a Scientific Theory?" in S. Morgenbesser (ed.), *Philosophy of Science Today*. New York: Basic Books.

Thompson, P. (1989), *The Structure of Biological Theories*. Albany: State University of New York Press.

van Fraassen, B. C. (1970), "On the Extension of Beth's Semantics of Physical Theories," *Philosophy of Science* 37: 325–338.

——— (1972), "A Formal Approach to the Philosophy of Science," in R. Colodny (ed.), *Paradigms and Paradoxes*. Pittsburgh: University of Pittsburgh Press.

——— (1980), *The Scientific Image*. Oxford: Clarendon Press.

——— (1985), "Empiricism in the Philosophy of Science," in P. M. Churchland and C. A. Hooker (eds.), *Images of Science: Essays on Realism and Empiricism*. Chicago: University of Chicago Press.

——— (1989), *Laws and Symmetry*. Oxford: Oxford University Press.

——— (1991), *Quantum Mechanics*. Oxford: Oxford University Press.

Weinert, F. (1999), "Theories, Models and Constraints," *Studies in History and Philosophy of Science* 30: 303–333.

See also **Carnap, Rudolf; Hempel, Carl Gustav; Instrumentalism; Kuhn, Thomas; Logical Empiricism; Realism; Reductionism; Scientific Models**

TIME

What is the relationship between time and space? What times exist? What does it mean to assert that time had a beginning? What explains why causation proceeds from past to future? These questions will serve to introduce some central debates in the study of time. It is best to begin with the relationship between time and space in special relativity.

Time and Special Relativity

Special relativity asserts that the backdrop on which physical happenings are arrayed is a four-dimensional space-time manifold (see Space-Time). To get used to thinking about manifolds, consider the following prerelativistic example.

Flatland (Abbott 1884) is a world consisting of just two spatial dimensions. Think of Flatland's space-time as a three-dimensional block, and think of the block as a stack of two-dimensional time-slices. Each time-slice determines the state of Flatland at a particular time, and each point of a slice determines what happens at a particular place at a particular time (see Figure 1).

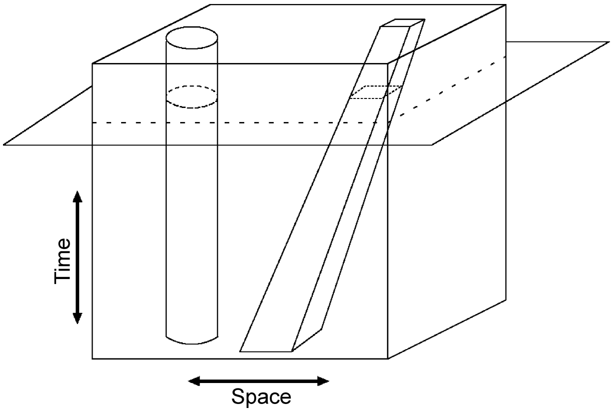


Fig. 1. Flatland. Flatland's space-time can be thought of as a three-dimensional block. Each two-dimensional slice corresponds to the state of the world at one time. In the history depicted, a circle sits unmoving on the left, while a square both moves to the right and shrinks over time. Also shown is the intersection between these objects and a sample time slice.

Flatland's space-time can be divided into slices in many different ways—horizontally; at another angle, into surfaces that each contain happenings at many different times. Notice that the horizontal time-slicing is special in that it alone guarantees that points on the same slice are simultaneous.

Like Flatland, a special relativistic world is associated with a space-time manifold. But there is no single privileged way of dividing a special relativistic manifold into time-slices. Instead, many candidate slicings are equally deserving of the title.

A contrast with Flatland is helpful at this point. Take any two instantaneous Flatland events—let them be flashes of light. Flatland's space-time determines which flash happened earlier (and the elapsed time between them). It also determines the spatial distance between their locations. Now consider two flashes of light in a relativistic space-time. Unlike Flatland's space-time, relativistic space-time does not in general have enough structure to determine which flash happened earlier (see Figure 2). Nor does it have enough structure to determine the spatial distance between the flash locations. Instead it determines a single magnitude of separation between the flashes: the space-time interval between them. So in the following (limited) sense, special relativity combines time and space: Special-relativistic space-time does not separately determine the spatial and temporal intervals between events.

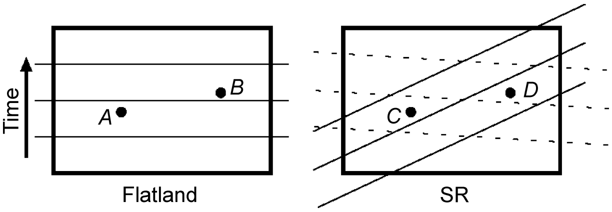


Fig. 2. Flatland Versus Special Relativity. Flatland space-time has a preferred decomposition into time-slices, compared with a special relativistic space-time, which does not. In each space-time, two instantaneous events are marked with dots. In Flatland, *A* occurs before *B*. In the relativistic space-time, relative to one system of time-slices (dotted lines), *C* occurs before *D*. But relative to another system of time-slices (solid lines), *C* occurs after *D*.

But in another sense, special relativity distinguishes sharply between time and space, for not all ways of slicing a special-relativistic manifold are on a par. Certain slicings (and associated labelings of the slices by spatial coordinates) are distinguished. These inertial frames are associated with lengths and time intervals as measured by unaccelerated observers (see Space-Time). Furthermore, although special relativity does not separately determine the time and space intervals between events, it does pick out the timelike curves, that is, those paths through space-time that correspond to slower-than-light travel. And it determines how much time elapses along any such path.

For example, consider a pair of clocks, initially synchronized and sitting next to each other. Suppose that clock *A* never accelerates but that clock *B* zooms far away and returns. The geometry of space-time determines the total elapsed time along the path of each clock. In this case, more time is elapsed along the path of clock *A* (the stay-at-home clock) than along that of clock *B*. As a result, when clock *B* returns, clock *A* will display a later time.

This example illustrates again that in a certain limited sense, special relativity mixes together time and space: The time elapsed for each clock depends on its trajectory through space. But the example also illustrates that special relativity does distinguish between time and space, for it picks out those paths along which it makes sense to speak of an elapsed time, and it determines how much time elapses along them.

One further issue deserves mention. As explained above, it makes no sense in special relativity to assert that two events are simultaneous in any absolute sense. It does make sense to assert that two events are simultaneous relative to some particular inertially moving observer. That much is uncontroversial. But a question remains: Once an observer has been fixed, does that alone determine which events are simultaneous? Or is it that in order to settle which events are simultaneous relative to an observer, one needs to adopt a *convention of simulteneity*? (For discussion of this question, see Malament 1977, Sarkar and Stachel 1999, and Janis 2002.)

Time and General Relativity

General relativity allows time and space to get mixed up in additional ways. Like special relativity, general relativity has it that space-time is a four-dimensional manifold. But manifolds in general relativity may be curved, in order to take gravitation into account. Furthermore, the way that matter is distributed through space-time constrains the way it is curved. For example, where there is a heavy and dense star, there is extreme curvature.

The constraint also goes in the other direction: The curvature of space-time constrains the way matter is distributed through it. For example, the extreme curvature of space-time near the above-mentioned star constrains the trajectories of particles floating in the star's vicinity. (It bends the trajectories of the particles toward the star.) Thus the dynamical laws of general relativity come in the form of a consistency condition—expressed by Einstein's field equations—between the curvature of space-time and its contents.

That form of law stands in sharp contrast to the form of the laws of classical mechanics, in which (as in special relativity), the geometrical structure of space-time is fixed, and the laws specify what states follow any given initial state. Such laws are compatible with time-slice generation: The state of one time-slice of the world produces or generates (via the laws) the states of subsequent time-slices (Maudlin 2002).

Since general relativity does not impose a fixed space-time geometry, it allows for universes with exotic temporal structures. The simplest example is that one can obtain circular time by pasting together two temporal ends of a special relativistic manifold (Earman 1986). The resulting "cylindrical" space-time has closed timelike curves: paths associated with slower-than-light travel that loop back on themselves. A particle traversing such a path returns to its own past.

Closed timelike curves also appear in rotating dust space-times (Gödel 1949) and in some space-times with wormholes—throats that provide shortcuts connecting one part of a manifold to another (Thorne 1995), all of which are consistent with the Einstein field equations. The physical possibility of such exotic space-times is in tension with the time-slice generation picture. Some such space-times—including rotating dust space-times and wormhole space-times—cannot be sliced into surfaces well suited to playing the role of time-slices. More precisely, they cannot be foliated into Cauchy surfaces: spacelike hypersurfaces that intersect every inextendible timelike curve exactly once (Arntzenius and Maudlin 2002). Other space-times cannot even be equipped with a temporal orientation: a structure that determines for each timelike curve which direction time elapses along that curve (Earman 1974; Sklar 1993).

One response to this tension is (a) to point out that the exotic universes are highly physically unrealistic and (b) to conclude that although the Einstein field equations allow them, they should be discarded as physically impossible on other grounds (Maudlin 2002; Callender 2000). Another response is to give up the time-slice generation picture and rest content with thinking that the laws of nature constrain what total histories of the world are allowable (Price 1996).

The Beginning of Time

Modern "big bang" cosmologies entail that time itself had a beginning. How is that claim to be understood? It should not be understood as the (contradictory) claim that there was a time before time began. It is better understood as there being an upper bound on the temporal distance between current events and any past event. (Here the temporal distance between two events is understood to be the time elapsed along a geodesic that connects the space-time locations of the events.)

How could it be that the past has finite extent? Consider a universe with such a past. For such a universe, it is tempting to think of the space-time as embedded in some larger, background space-time, just as a curved surface of finite extent might be embedded in an infinite Euclidean space. If so, then the larger space-time might have an infinite past after all.

But *this picture should be resisted.* The curvature and extent of space-time can be completely characterized by spatiotemporal relations that hold among its parts. It is *not* necessary to posit a "container" space-time, and modern cosmologies posit no such things. Nor is it necessary to posit a first time, for it may be that (though the universe has a finite past in the above sense) every moment is preceeded by earlier ones (Earman 1995, chap. 7).

Presentism

According to relativity theory, space-time is a four-dimensional manifold. Interpreted literally, this entails that there exists a four-dimensional manifold. It follows that there exist some things that do not presently exist (e.g., the part of the manifold corresponding to 1776). *Presentists* deny this, and claim that whatever exists at all, exists now (Markosian 2002). So, unlike their opponents, presentists cannot interpret literally relativity's claim that space-time is a four-dimensional manifold. (Relationalists about space-time also deny that there exists a four-dimensional space-time manifold, but do so for different reasons than presentists.)

In addition, presentism is apparently incompatible with the lesson of special relativity that there is no privileged notion of simultaneity. For if presentism is correct, it makes sense to speak of what exists now, without specifying a frame of reference. So presentism provides the following criterion of simultaneity: Two events are simultaneous if and only if they are both currently occurring.

Some respond to the apparent incompatibility between presentism and special relativity by concluding that presentism is false. (See Putnam 1967 for a treatment of this kind; Stein 1968 for a response; and Callender 2000 for a rejoinder.) One presentist response is to adopt a theory that is observationally equivalent to special relativity but in which there is a preferred reference frame (Markosian 2002; Bell 1989). Maudlin (1994) suggests that quantum nonlocality may provide independent motivation for such a replacement (see Locality).

Temporal Asymmetry

An important cluster of questions concern various asymmetries in time. One such asymmetry is the irreversible nature of certain thermodynamic processes (such as the cooling of a hot cup of tea) (see Irreversibility). Another asymmetry is causal asymmetry: earlier events cause later events, but not the other way around. What explains causal asymmetry?

According to one answer, time itself has a built-in past→future orientation that is constitutively tied to the direction of causation (Maudlin 2002). To illustrate what it means for time to have a built-in past→future orientation, consider the following specification of a Flatland space-time (based on an example in Maudlin 2002):

> The space-time contains nothing but two stationary spheres (*X* and *Y*) and a particle that makes a single trip from one to the other.

It might be thought that this specification is incomplete by failing to determine whether the particle travels from *X* to *Y* or the other way around. Whether the specification is incomplete in this way depends on whether each Flatland space-time has a past→future orientation. Such an orientation would determine which temporal end of the space-time is the past end and which end is the future. If Flatland space-times have past→future orientations, then the specification was incomplete in the manner described, for in that case, there are two space-times meeting the specification: one in which the particle travels from X to Y, and one in which it travels from Y to X.

On the other hand, if Flatland space-times lack past→future orientations, then the specification was *not* incomplete in the manner described, for in that case, the above two possible space-times collapse into a single one. This space-time has two temporal ends, and nothing determines which is the past and which is the future. At one temporal end, the particle is located at X. At the other end, it is located at Y. And at intermediate times, it is located at various places in between.

Of course, in speaking about a Flatland space-time with no past→future orientation, one might adopt the convention of treating one end of the space-time as past and the other as future. But nothing in the structure of the space-time itself privileges that convention over the one that labels the ends in the opposite way (Price 1996).

If time—actual time—has a built-in past→future orientation, then one might explain the asymmetry of causation by an analysis in which the direction of causation necessarily aligns with the past→future direction (Maudlin 2002). If not (if time has no built-in past→future orientation), then that explanation is unavailable. One might in this case adopt perspectivalism about causation. According to perspectivalism, the term "cause" picks up its temporal asymmetry from the manner in which one—a temporally asymmetric agent—is embedded in space-time (Price 1996).

To illustrate perspectivalism, consider a pair of agents who are temporally reversed with respect to each other. (For example, one agent's biological processes proceed in the opposite temporal direction from the other agent's.) According to one version of perspectivalism, the two agents possess different concepts of causation. The difference in their concepts arises because what one agent counts as the past→future direction of time, the other agent counts as future→past. Since in this view time has no built-in past/future asymmetry, both concepts of causation are on a par.

A third explanation of causal asymmetry agrees with perspectivalism that time has no built-in past/future asymmetry. But it insists that the asymmetry of causation is not just a consequence of the particular perspective that one adopts. Instead, it appeals to an analysis of causation in terms of the distribution of matter over space-time. According to that analysis, special features of the actual matter distribution entail that causation proceeds only in one direction. But other distributions of matter would have yielded causation proceeding in the opposite direction. And still other distributions would have yielded no pervasive causal asymmetry at all (Lewis 1986).

ADAM ELGA

References

Edwin Abbott (1884), *Flatland: A Romace of Many Dimensions*. Boston: Roberts Bros.

Arntzenius, Frank, and Tim Maudlin (2002), "Time Travel and Modern Physics," in Craig Callender (ed.), *Time, Reality and Experience*. Cambridge: Cambridge University Press.

Bell, John (1989), "How to Teach Special Relativity," in *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Callender, Craig (2000), "Shedding Light on Time," *Philosophy of Science* 67 (suppl): S587–S599.

Earman, John (1974), "An Attempt to Add a Little Direction to the 'Problem of the Direction of Time,'" *Philosophy of Science* 151: 15–47.

——— (1986), *A Primer on Determinism*, vol. 32 of the University of Western Ontario Series in the Philosophy of Science. Dordrecht, Holland: D. Reidel.

——— (1995), *Bangs, Crunches, Whimpers and Shrieks*. Oxford: Oxford University Press.

Gödel, Kurt (1949), "A Remark about the Relationship Between Relativity Theory and Idealistic Philosophy," in P. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*. LaSalle, IL: Open Court, 557–562.

Janis, A. (2002), "Conventionality of Simultaneity," in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/fall2002/entries/spacetime-convensimul

Lewis, David (1986), "Counterfactual Dependence and Time's Arrow," in *Philosophical Papers*. Oxford: Oxford University Press.

Malament, D. (1977), "Causal Theories of Time and the Conventionality of Simultaniety," *Noûs* 11: 293–300.

Markosian, Ned (2002), "A Defense of Presentism," in Dean Zimmerman (ed.), *Oxford Studies in Metaphysics*, vol. 1. Oxford: Oxford University Press.

Maudlin, Tim W. (1994), *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*. Oxford: Basil Blackwell.

——— (2002), "Remarks on the Passing of Time," *Proceedings of the Aristotelian Society* CII: 237–252.

Price, Huw (1996), *Time's Arrow and Archimedes' Point*. New York: Oxford University Press.

Putnam, Hilary (1967), "Time and Physical Geometry," *Journal of Philosophy* 64: 240–247.

Sarkar, S., and J. Stachel (1999), "Did Malament Prove the Non-Conventionality of Simultaneity in the Special Theory of Relativity?" *Philosophy of Science* 66: 208–220.

Sklar, Lawrence (1993), *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. New York: Cambridge University Press.

Stein Howard (1968), "On Einstein-Minkowski Space-Time," *Journal of Philosophy* 65: 5–23.

Thorne, Kip S. (1995), *Black Holes and Time Warps*. New York: W.W. Norton and Company.

See also **Causality; Irreversibility; Space-Time**

ALAN TURING

(23 June 1912–7 June 1954)

The work of the British mathematician Alan Turing stands as the foundation of computer science. Turing's (1936–1937) definition of computability remains a classic paper in the elucidation of an abstract concept into a new paradigm. His 1950 argument for the possibility of artificial intelligence is one of the most cited in modern philosophical literature. These papers, his best known, have led his contributions to be defined as theoretical. But his work was highly practical, both in codebreaking during World War II and in the design of an electronic digital computer. Indeed Turing's expression for the modern computer was "practical universal computing machine," a reference to his 1936 "universal machine." This combination of theory and practice meant that Turing's work fit no conventional category of "pure" or "applied." Likewise, his life involved many contradictions. Detached from social and economic motivations, and perceived as an eccentric, apolitical, unworldly innocent, he was swept into a central position in history as the chief scientific figure in the Anglo-American mastery of German communications.

The Matter of Mind

Amidst this complexity there is one constant theme: Turing's fascination with the description of mental action in scientific terms. His computational model can be seen as a twentieth-century renovation of materialist philosophy, with a claim that the discrete state machine is the appropriate level of description for mental states. However, it did not begin in that way: Turing's interest in the material embodiment of mind comes first in a private letter of about 1932 (Hodges 1983, 63), in which he alluded to the newly elucidated quantum-mechanical nature of matter and then, influenced by Eddington, speculated on "will" as a physical effect in a region of the brain. It was after studying von Neumann's axioms of quantum mechanics, then Russell on the foundations of mathematics, that he learned of logic, and so of the question that was to make his name.

The question, proposed by Hilbert but transformed through the 1931 discovery of Kurt Gödel, was that of the *decidability* of mathematical propositions. Is there a definite method or procedure that can (in principle) be applied to a mathematical proposition and will decide whether it is provable? Turing learned of this outstanding *Entscheidungsproblem* from the lectures of the Cambridge mathematician Max Newman. The difficulty of the question was that it demanded an unassailable definition of the concept of method, procedure, or algorithm. This is what Turing supplied in 1936 through the definition of what came to be called the *Turing machine.* Specifically, he modeled the action of a human being in following a definite method, either through explicit instructions or through following a sequence of "states of mind."

Turing's definition came shortly after another elucidation of "effective calculability" by the United States' logician Alonzo Church. Thus, in a narrow sense, Turing was preempted. Church's definition turned out to be mathematically equivalent to Turing's definition of computability. But Church (and Gödel) agreed that Turing's definition gave an intuitively compelling account of why this definition encompassed "effectiveness."

The Turing machine breaks down the concept of a procedure into primitive atomic steps. The exact details are somewhat arbitrary, and nowadays other equivalent formulations are often used. The essential point is that the machines should have finite descriptions (as "tables of behavior") but be allowed unlimited time and space for computation.

Church's thesis, that his definition of effective calculability would capture any natural notion of that concept, became the Church-Turing thesis, and opened a new area for mathematical "decision problems." But the work had much wider consquences: Turing's bold modeling of states of mind opened a new approach to what are now called the cognitive sciences (see Cognitive Science). It is often asserted that modeling the mind with a computer shows the influence of a dominant technology, but in fact Turing's work went in the reverse direction—for a striking and visionary aspect of Turing's paper was his definition of a "universal" Turing machine, which led to the computer. A machine is universal if it can read the table of behavior of any other machine and then execute it. This is just what a modern computer does, the instructions in programs being equivalent to tables of behavior. It was essential in Turing's description that the instructions be stored and read like any other form of data; and this is the idea of the internally stored program. It is now hard to study Turing machines without access to the programmers' mindset, and to remember that when Turing formulated them, computers did not exist.

Turing and Machines

Turing's work, being based on answering Hilbert's question, modeled the *human being* performing a methodical computation. However, the imagery of the teleprinter-like "machine" was striking. Newman (1955), in stressing the boldness of this innovation, said that Turing embarked on "analyzing the general notion of a computing machine" (256). This was criticized by Gandy (1988) as giving a false impression of Turing's approach. But Newman's account of the flavor of Turing's thought should not be entirely discounted; Turing certainly became fascinated by machines and engineered machines with his own hands in 1937–1939. Church also makes it clear that the notion of a computing machine was current at this time. Church (1937) wrote (while Turing was working with him at Princeton) that Turing

> proposes as a criterion that an infinite sequence of digits 0 and 1 be 'computable' that it shall be possible to devise a computing machine, occupying a finite space and with working parts of finite size, which will write down the sequence to any desired number of terms if allowed to run for a sufficiently long time. As a matter of convenience, certain further restrictions are imposed on the character of the machine, but these are of such a nature as obviously to cause no loss of generality—in particular, a human calculator, provided with pencil and paper and explicit instructions, can be regarded as a kind of Turing machine (42).

Yet neither Turing nor Church *analyzed* the general concept of a computing machine with "working parts." Turing (1939), giving a definitive statement of the Church-Turing thesis, used the expression "purely mechanical" without further analysis. Only in 1948 did he give some more discussion.

This topic has recently been made controversial by B. J. Copeland, who holds that the Church-Turing thesis is widely misunderstood, because it was never intended to apply to machines. Copeland overlooks Church's characterization of computability, as quoted above, which assumes that all finitely defined machines fall within the scope of computability. To support his claim, Copeland points to the "oracle" defined by Turing (1939), which supplies an uncomputable function, and holds that it gives a broader characterization of computation. But the whole point of Turing's oracle is that it facilitates the mathematical exploration of the *uncomputable.* The oracle, as Turing emphasized, cannot be a machine. It performs *nonmechanical* steps. His "oracle-machines," defined so as to call upon oracles, are not purely mechanical.

Turing's oracle is related to Gödel's theorem, which seems to show that the human mind can do more than a mechanical system when it sees the truth of formally unprovable assertions. Turing described this as mental "intuition." The oracle, as Newman (1955) interpreted it, can be taken as a model of intuition. But Turing left open the question of how intuition was to be considered as actually embodied. He was not at this stage committed to the computability of *all* mental acts, as came to be his position after 1945. His 1936 work had considered the mind only when applied to a definite method or procedure. Turing had to resolve this question before embarking on his artificial intelligence program.

Copeland has gone even further and has described Turing's oracle as heralding a new revolution in computer science, illustrating "what Turing imagined" by sketching an oracle supposed to operate by measuring a physical quantity to infinite precision. But Turing's oracle models what machines *cannot* do, and the question for him was, and always remained, whether machines can do as much as minds. He did not suggest the opposite idea, stated in Copeland (1997), that "among a machine's repertoire of atomic operations there may be those that no human being unaided by machinery can perform."

Naturally, one must distinguish between the historical question of what Turing thought and scientific truth. It is a serious (and unanswered) question as to whether actual physical objects do necessarily give rise to computable effects. Nowadays we demand a closer analysis of "finite size." Gandy (1980) arrived at conclusions that generally support Church's assumptions: The limitations of computability follow from quite general assumptions about the construction of a machine. But if the constraint of finiteness is interpreted so as to allow a "machine" with infinitely many subcomponents built on smaller and smaller scales, or working faster and faster without limit, then it is easy to show that Turing's computability can be surpassed in a finite time. In any imaginary universe, such a construction might be possible; such examples may therefore be said to show that the Church-Turing thesis has a physical content.

"Effective" means "doing" (as opposed to postulating or imagining), thus depending on some concept of realistic action, and hence on physical law. Quantum computation has already shown that the classical picture of "doing" is incomplete. The nature of quantum mechanics, still mysterious because of its nonlocality and the "reduction" during measurement, means that there may yet be more to be found out, and in recent years the work of Penrose (1989 and 1994), who like Turing focuses on the mind and brain, has drawn new attention to this question (see Locality; Quantum Mechanics; Quantum Measurement Problem).

Turing's Practical Machinery

It is perhaps surprising that Turing himself did not, in this 1936–1939 period, say anything about the physics of the machine concept, in view of his interest in quantum mechanics. He might have done so but for the war. War disrupted Turing's investigation of the uncomputable, along with his conversations with Wittgenstein on the foundations of mathematics (Diamond [1939] 1976), which never extended, as many now wish they had been, into the philosophy of mind. But war work gave Turing an intimate acquaintance with the power of algorithms and with advanced technology for implementing them, for Turing became the chief scientific figure at Bletchley Park, the British cryptanalysis center and a key location in modern history.

Turing had anticipated this development in 1936. He had applied his ideas to "the most general code or cipher," and one of the machines he made himself was to implement a particular cipher system (Hodges 1983, 138). This, together with his Cambridge connection with influential figures such as J. M. Keynes, may explain why he was the first mathematician brought into the codebreaking department.

Turing transformed the Government Code and Cypher School with the power of scientific method. His logic and information theory, applied with advanced engineering, achieved astounding feats, with particular effect in decrypting the U-boat Enigma signals, for which he was personally responsible. By 1944 the power, reliability, and speed of electronic technology showed Turing that his universal machine could be implemented. The plethora of advanced algorithms employed in cryptanalysis also supplied ample practical motivation.

In 1945, Turing was appointed to the National Physical Laboratory, with the commission of designing an electronic computer. Turing's plans soon emerged as the ACE (Automatic Computing Engine) proposal (Turing [1946] 1986). Again, Turing was preempted by work in the United States, for his publication had been preceded in 1945 by the report of EDVAC (the Electronic Discrete Variable Automatic Computer). Turing's plans, however, were independent, more detailed, and more far-reaching. Furthermore, a recent survey (Davis 2000) suggests that von Neumann needed his knowledge of Turing's work when shaping EDVAC. As a point of interest in the history of science, none of the mathematical leaders—Turing, von Neumann, Newman—clearly defined the stored program concept or its debt to symbolic logic in treating instructions as data (see von Neumann, John). Turing never published the book on the theory and practice of computation that he might have done, and so neglected his own good claim to be the inventor of the computer.

It is an unobvious fact, long resisted, but now familiar, that more complex algorithms do not need more complex machines, only sufficient storage space and processor speed. This was Turing's central idea. In a world familiar with the power of a universal machine, one can better appreciate the remark by Turing ([1946] 1986, 44) that "every known process has got to be translated into instruction table form at some stage." Turing emphasized that arithmetical calculations were only one aspect of the computer's role—partly the influence of nonnumerical Bletchley Park work, but more deeply, his base in symbolic logic. His hardware design was probably impractical in detail, but he far surpassed von Neumann in seeing the significance of software and that this would use the computer itself, a fact now familiar in compilers and editors:

> The work of constructing instruction tables should be very fascinating. There need be no real danger of it ever becoming a drudge, for any processes that are quite mechanical may be turned over to the machine itself. (Turing [1946] 1986, 44)

Turing's insight into programming by modifying instructions led to the idea of simulating learning, training, and evolution by an extension of these ideas. As he put it, human intelligence required something other than "discipline," namely "initiative." By 1945 Turing had convinced himself that human faculties of an apparently nonmechanical nature did not require explanation in terms of the uncomputable. Turing was well aware of the paradox of expecting intelligence from a machine capable only of obeying orders. But he believed that with sufficient complexity, machines need not appear "mechanical," as in common parlance.

A crucial point is that Turing had by this stage formulated an argument from human "mistakes" to explain why Gödel's theorem did not show the existence of an uncomputable human intuition; indeed, remarks in Turing ([1946] 1986) show the early influence of this view in his project for artificial intelligence (Hodges 1999). He now expected self-modifying machines to exhibit the apparently "nonmechanical" aspects of mental behavior.

With this as his strategic goal, Turing sketched networks of logical elements, to be organized into function by "training" or by an analogy with evolution. Although he did not develop his specific schemes, he anticipated the "connectionist" approach to artificial intelligence. Nowadays the term "nonalgorithmic" is confusingly used for systems in which the program is implicitly developed, rather than explicitly written by a programmer. Turing was, however, quite clear that such operations still lie within the realm of the computable. All these developments were sketched in Turing ([1948] 1969). He also in this paper gave his only systematic account of the concept of a machine. In doing so, he introduced the possiblility of random elements into the Turing machine model, but he made no reference to a need for uncomputable elements in randomness. Indeed he indicated that pseudo-random elements, clearly computable, would suffice.

Turing's Intelligent Machinery

Disappointed with the lack of progress at the National Physical Labotratory, Turing moved to Manchester in 1948. There Michael Polanyi stimulated Turing to write for a more general audience on the question of whether machines could in principle rival the human mind. The idea in Turing (1950) was to cut through traditional philosophical assumptions about 'mind' by the thought experiment now known as the Turing test, but which he called "the imitation game." A human and a programmed computer compete to convince an impartial judge that they are human, by textual messages alone. Turing's position was that thought or intelligence, unlike other human faculties, is capable of being fairly tested through a communication channel like a teleprinter.

Critics have raised questions about nonverbal communication, cultural assumptions, animal intelligence, and other issues. Turing's principal defense against all such arguments was that one can judge the intelligence of other humans only by making such external comparisons, and that it is unfair to impose more stringent criteria on a computer. But it may be held that when addressing human consciousness with moral seriousness, there is something inadequate about a definition of intelligence that depends upon deceit (see Consciousness). Turing also confused the issue by introducing the imitation game with a poor analogy: a parlor game in which a man has to pretend to be a woman under the same conditions of remote questioning. In such a game, imitation proves nothing, so the analogy is misleading and has confused many readers. However, Turing's drama has the merit of expressing the full-bloodedness of his program. His wit has attracted lasting popular interest. Turing's references to gender have also fascinated cultural critics, who speculate widely on biographical and social issues in their commentaries (Lassègue 1998).

A drier but stronger feature of Turing (1950) lies in his setting out the level of description of the discrete state machine, and his emphasis on explaining computability and the universal machine. Critics who point out that the brain is not structured like a computer miss his essential point that any algorithm can be implemented on the computer. This applies to explicit algorithms and to those arrived at by processes as in neural networks; Turing described both. Another strength of Turing's paper lies in his advocating both approaches, never seeing programming as standing in opposition to the modeling of intelligence by learning and adaptation. Artificial intelligence research has tended toward division between the two camps dominated by expert systems and neural networks, but recently hybrid approaches have appeared (see Artificial Intelligence). Thus Turing's ideas still have force. Also futuristic was Turing's prophecy that by the end of the century "one will be able to speak of machines thinking without expecting to be contradicted." It is probably true to say that this prophecy was not fulfilled in 2000, but Turing was prepared to take this risk:

> The popular view that scientists proceed inexorably from well-established fact to well-established fact, never being influenced by any unproved conjecture, is quite mistaken. Provided it is made clear which are proved facts and which are conjectures, no harm can

> result. Conjectures are of great importance since they suggest useful lines of research. (Turing 1950, 442)

Turing's conjecture was that the brain's action is computable. If this is true, then it is hard to refute his argument that given sufficient storage space, a computer can perform the function of the mind. Turing emphasized this argument by putting a figure on the storage capacity of the brain. But he still did not directly address that underlying question of whether physical objects necessarily have computable behavior. Artificial intelligence research has generally accepted without question the assumption that they do. But Penrose (1989 and 1994), linking the still ill-understood "reduction" in quantum mechanics with Gödel's theorem, has thrown the spotlight back on the problems that Turing himself found most perplexing.

Turing's Unfinished Work

Turing (1951) did refer to the problem of quantum mechanics when giving a popular radio talk. As Copeland (1999) has noted, he gave a more qualified assertion of the computability of mental action than in Turing (1950). But this was not because of believing oracles to reside in the brain; it was because of quantum unpredictability. Harking back to his schooldays reading, he referred to Eddington in connection with the mechanism of the brain. This brief allusion may explain why Turing thereafter gave fresh attention to quantum theory. His friend, student, and colleague Gandy ([1954] 2001) wrote in a letter to Newman of Turing's ideas, in particular of his attention to the question of the "reduction" or "measurement" process and the "Turing paradox" that according to standard theory, continuous observation should prevent a system from evolving.

These ideas were not developed into publication. They were curtailed by his suicide in 1954. One of many ironies of Turing's life, lived for science, was that he suffered in 1952–1953 a "scientific" treatment by estrogen supposed to negate his homosexuality. This treatment was the alternative to prison after his arrest in February 1952. His openness and defiance did not command the admiration of authority, and there was at least one further crisis when he found himself under watch. One may regret that he did not write more of his own sense of liberty and will. Indeed it is remarkable that he, so original and unconventional, should champion the possibility of programming the mind. But he left few hints as to the personal dilemmas on the roads to freedom. He was of course constrained by intense state secrecy, as the most privileged insider to Anglo-American secrets in a sphere then kept totally secret.

Besides his new enquiries in physics, he had a large body of incomplete theory and computational experiment in mathematical biology. This, neglected until the 1970s, is now the foundation of a lively area of nonlinear dynamics. Turing described his theory as intended to oppose the "argument from design." It was a theme parallel to (and through his interest in brain function connected with) his quest for a new materialism of mind. He had not exhausted his ideas, and their impact has not yet been fully absorbed.

ANDREW HODGES

References

Church A. (1937), "Review of Turing 1936–7," *Journal of Symbolic Logic* 2: 42–43.

Copeland. B. J. (1997), "The Church-Turing Thesis," in E. N. Zalta (ed.), *Stanford Encyclopaedia of Philosophy* http://plato.stanford.edu.

——— (1999), "A Lecture and Two Radio Broadcasts on Machine Intelligence by Alan Turing," in K. Furukawa, D. Michie, and S. Muggleton (eds.), *Machine Intelligence, 15*. Oxford: Oxford University Press, 445–476.

Davis, M. (2000), *The Universal Computer*. New York: Norton.

Diamond, C. (ed.) ([1939] 1976), *Wittgenstein's Lectures on the Foundations of Mathematics*. Hassocks, UK: Harvester Press.

Gandy, R. O. ([1954] 2001), "Letter to M. H. A. Newman," in R. O. Gandy and C. E. M. Yates (eds.), *Collected Works of A. M. Turing* (4 vols.): *Mathematical Logic* Amsterdam: North-Holland.

———(1980), "Principles of Mechanisms," in J. Barwise, H. J. Keisler, and K. Kunen (eds.), *The Kleene Symposium*. Amsterdam: North-Holland.

——— (1988), "The Confluence of Ideas in 1936," in R. Herken (ed.), *The Universal Turing Machine: A Half-Century Survey*. Oxford: Oxford University Press.

Hodges, A. (1983), *Alan Turing: The Enigma* London: Burnett, New York: Simon & Schuster.

——— (1999), *Turing, a Natural Philosopher*. New York: Routledge.

Lassègue, J. (1998), *Turing*. Paris: Les Belles Lettres.

Newman, M. H. A. (1955), *Alan M. Turing, Biographical Memoirs of Fellows of the Royal Society*, 1: 253–263.

Penrose, R. (1989), *The Emperor's New Mind*. Oxford and New York: Oxford University Press.

——— (1994), *Shadows of the Mind*. Oxford and New York: Oxford University Press.

Turing, A. M., ed. B. J. Copeland (2004) The essential Turing (Oxford: Oxford University Press).

Turing A. M. (1936–37), "On Computable Numbers, with an Application of the Entscheidungsproblem," *Proceedings of the London Mathematical Society* (series 2) 42, 230–265.

——— (1939), *Systems of Logic Defined by Ordinals, Proceedings of the London Mathematical Society* (series 2) 45: 161–228.

——— ([1946] 1986), “Proposed Electronic Calculator,” in B. E. Carpenter and R. W. Doran (eds.), *Turing’s ACE Report of 1946 and Other Papers*. Cambridge, MA: Tomash/MIT Press.

——— ([1948] 1969), “Intelligent Machinery,” *Machine Intelligence* 5: 3–23. http://www.turingarchive.org.

——— (1950), “Computing Machinery and Intelligence,” *Mind* 59, 433–460.

——— (1951), BBC Radio talk, script available at http://www.turingarchive.org; also in Copeland 1999.

See also **Artificial Intelligence; Cognitive Science; Consciousness; Hilbert, David; Russell, Bertrand; von Neumann, John**

U

UNDERDETERMINATION OF THEORIES

The *underdetermination of theory by data* allegedly reveals something important about science. Although a philosophical commonplace, the term 'underdetermination' is used in different ways by various authors. At the same time, many authors write as if it were a single, well-understood phenomenon. Some write as if underdetermination were the problem that every theory has empirically equivalent rivals. Others write as if it were the problem that theories entail predictions only in conjunction with auxiliary hypotheses. Others write as if it were the problem that methodological rules may be ambiguous, vague, or circular. Thus, the problem of underdetermination is best understood as an extended family of potential problems.

To put it crudely, underdetermination obtains when scientists are unable to decide responsibly which theory to believe, that is, the choice between rival theories is underdetermined if scientists cannot make a responsible choice of one over the others. So underdetermination is always relative to some standard for what will count as *responsible theory choice*.

There is a banal kind of underdetermination that obtains any time scientists are ignorant about something. They are, in that situation, unable to decide responsibly what to believe. So they do some research; they gather relevant evidence; and the question is resolved. The underdetermination of interest to philosophers is something more trenchant than this: suppose scientists were not able to decide between rival theories in a broad range of circumstances. Yet there is no general agreement about how broad this range of circumstances must be in order for a choice to be underdetermined. It is often stipulated that a choice is underdetermined if it could not be responsibly made even after all evidence had been collected (as in Quine 1970). Other authors (such as Stanford 2001) write of "transient underdetermination" that obtains only at certain times. To take in these and other permutations, say that underdetermination is always relative to a *scope* or range of circumstances across which responsible choice is impossible. The banal kind of underdetermination obtains for a scope that contains only present circumstances, but not for a broader scope containing circumstances in which the further research has been conducted.

In the most general terms, underdetermination obtains for a set of rival theories, a standard, and a scope. In order to be pernicious, a case of underdetermination must obtain between rivals worth taking seriously, according to a reasonable standard, and with a scope that includes not only present circumstances but also most any plausible, future circumstances. However, such a case of underdetermination would not show anything about the whole scientific enterprise. It would show only that science could not settle that particular question.

Philosophers have used two general strategies to try to show that underdetermination is a problem for all of science. Some arguments operate at the level of science generally, attempting to show directly that all or most scientific theory choice is underdetermined and thus that a successful scientific theory is not any better than its losing rivals. Other arguments begin from alleged examples of underdetermination and generalize, suggesting that all or most scientific theory choice is underdetermined in the same way. Since these bottom-up arguments diagnose science as rife with underdetermination, they are often used to support the same conclusions as direct, top-down arguments.

Empirical Equivalence

The problem of empirically equivalent theories is one variety of underdetermination problem. The worry takes this form:

(1) Every theory has indefinitely many empirically equivalent rivals;
(2) There is no good reason to believe a theory over its empirically equivalent rivals;
(3) Therefore, there are never decisive reasons to believe any theory—that is, theory choice is always underdetermined.

Premise (1) states that, for any theory, there are rival theories of a particular sort. This gives us a set of rivals for the underdetermination scenario. Premise (2) states that permissible standards of judgment are insufficient to distinguish between empirically equivalent theories. The conclusion brings these two together: The choice among the set of rivals given by (1) is, by (2), underdetermined. The implicit scope here is all observationally possible circumstances. The ultimate conclusion that one may accept but not believe successful theories turns on a substantive principle about what one should do in the face of such underdetermination. The argument thus presupposes a division between circumstances that are observationally possible and those that are not. This boundary is vague, historically variable, and—some have argued—unprincipled (see, e.g., Churchland and Hooker 1985 and Laudan and Leplin 1991).

Some arguments for (1) proceed by offering an algorithm for producing empirically equivalent rivals. For example: For any theory *T*, let T^* be the theory "*T* is false, but all observations are just as if *T* were true." This is like a Cartesian skeptical scenario, akin to worries about dreams or evil demons. With (1) interpreted in this way, (2) would be true only for an implausibly strict standard; all manner of ordinary belief choice would be underdetermined.

Many philosophers have argued against (2) by offering nonempirical criteria by which one might decide between empirically equivalent rivals: simplicity, breadth of scope, explanatory power, heuristic fecundity, practical success, and so on. It is possible to accept the further criteria but argue that they still result in a kind underdetermination. Rival theories might beat one another according to different criteria, such that no choice between them is possible. If scientists are aiming at accurate and simple theories, for example, then they will be stymied by cases where the simpler theory is less accurate. It is then essential to ask whether real scientific theories face this kind of underdetermination or whether scientific controversies are typically resolved in favor of a theory that is superior according to all or most criteria (contrast Doppelt 1978 and Kitcher 1993, esp. ch. 7). It is also possible to reject these further criteria. For instance, the constructive empiricist admits that these criteria are pragmatically important while denying that they have anything to do with the truth of theories. One should accept the theory that best satisfies these super-empirical desiderata, but—according to the constructive empiricist—one need not believe it (van Fraassen 1980).

Other philosophers have argued against (2) by noting that even when two theories are empirically equivalent considered in isolation, they may have divergent empirical consequences when conjoined with other accepted theories (Laudan 1990). This point is often answered by alleging that background theories are themselves underdetermined—this is a form of the Duhem-Quine thesis. A similar response is to concede that background theories resolve the underdetermination between theories, but to argue that underdetermination recurs between rival "total sciences" or complete systems of theories.

The argument from empirical equivalence is primarily a matter between realists and antirealists. At most, it shows that every theory faces underdetermination against some possible rivals. It does *not* show that underdetermination obtains between all rival theories or that any theory is as good as any other. It has few implications for broader questions about the authority enjoyed by science.

The Duhem-Quine Thesis

Quine is often mentioned as having established the force of underdetermination. Since he first introduces it without much positive argument (Quine 1953, 41, 17n) and attributes it to Duhem ([1914] 1954), one variety of underdetermination is called the Duhem-Quine thesis (see Duhem Thesis).

Scientific theories typically do not yield observable consequences on their own. Suppose biologists observe an amoeba under a microscope, and it looks as if it is performing tricks that contradict their theory of amoeba behavior. They may accept this as a refutation and decide that their theory of microbial behavior needs revision. The observation refutes their theory only insofar as they assume that the microscope gives an accurate glimpse into amoeboid life. From the standpoint of deductive logic, the observation shows only that their theory is incorrect *or* that the microscope is deceptive.

Of course, instruments sometimes produce erroneous results. The worry is that scientists might *always* plead instrument error to discount recalcitrant observations. Biologists could maintain a theory of microbial behavior in the face of any would-be refutations if they were willing to dismiss all those observations as experimental artifacts.

The underdetermination obtains between any rival theories that are sophisticated enough to require auxiliary assumptions in order to yield testable predictions, with a standard that demands that scientists accept deductive consequences of observations, and with a scope that includes all possible observations. Yet a standard that requires scientists to accept only deductive consequences would leave a great deal underdetermined. Given such a standard, it would be impossible to make any predictions about the future based on observations that are always of the present or of the past. Arguably, the underdetermination disappears given a more plausible standard. In the example, there is no absolute principle that requires biologists to blame the anomaly on their theory rather than on the microscope. Yet, when an accepted instrument is used in a usual way, its reliability is assumed. A pattern of strange results might lead them to worry about this assumption, but no science would be possible if everything were up for grabs at the same time (see Kitcher 1993, Ch. 7, §6).

Some philosophers, such as Longino (1990), have argued that this reliance on background assumptions, although unavoidable, makes science rest on implicit value commitments. They argue that these commitments should be ones that scientists can reflectively endorse.

Bottom-Up Arguments

The arguments about empirically equivalent theories and the Duhem-Quine thesis proceed in a top-down fashion, moving directly to the conclusion that all or most theory choice is underdetermined. Other arguments work in the opposite direction, starting from specific examples of underdetermination in science. Since underdetermination is present in the cases considered, then perhaps it is elsewhere as well. This final step is often implicit, and one may doubt whether a grab bag of cases should ignite worries about science generally. (Contrast Ellis' contribution to Churchland and Hooker 1985, 63, Earman 1993, 31, and Stanford 2001, 6.)

Standard examples have concerned the geometrical structure of space. For instance, general relativity treats space-time as non-Euclidean, such that seemingly curved paths like planetary orbits are described as straight paths in a curved space-time. However, objects would follow the same trajectories if space-time were Euclidean and "universal forces" distorted the paths of the objects. Note, of course, that a physics with universal forces would look very different from physics as it is usually formulated. Regardless, it is unclear whether the alleged underdetermination (the "conventionality of geometry") tells us anything about sciences besides mechanics (see Conventionalism).

An impressive panoply of historical cases has been offered by philosophers and other authors writing in the area of science studies: the rivalry between classical Newtonian and relativistic mechanics, between phlogiston and oxygen theories of combustion, between caloric and kinetic theories of heat, between particle and wave theories of light, between the germ theory of disease and its various rivals, and so on. One lesson of these studies has been that single experiments are rarely, if ever, completely decisive.

Certainly, there is a kind of underdetermination that obtains in cases like these. There is a time when the old theory is sagging under the weight of anomalies but before the new theory has accrued its eventual successes, during which there is legitimate disagreement about which theory to accept. Considering a scope that includes only this period of legitimate controversy, these are cases of underdetermination. The underdetermination is of historical interest only if it also obtains for a wider scope that includes the present scientific situation. Even vague standards may have been decisive in the long run, as adherents of the old theory were unable to salve its troubles and the new theory accrued successes.

Bottom-up arguments thus face two challenges: first, to show that the case is underdetermined with a wide enough scope that it has not been and cannot be resolved by further research; second, to show that the underdetermination present in specific cases generalizes to the rest of science.

P. D. Magnus

References

Churchland, Paul M., and Clifford A. Hooker (eds.) (1985), *Images of Science*. Chicago: University of Chicago Press.

Doppelt, Gerald (1978), "Kuhn's Epistemological Relativism: An Interpretation and Defense," *Inquiry* 21: 33–86.

Duhem, Pierre ([1914] 1954), *The Aim and Structure of Physical Theory*. Translated by P. P. Wiener. Princeton, NJ: Princeton University Press. Originally published as *La Théorie Physique: Son Objet, Sa Structure* (Paris: Marcel Rivière & Cie).

Earman, John (1993), "Underdetermination, Realism, and Reason," in *Midwest Studies in Philosophy* (Vol. XVIII). Notre Dame, IN: University of Notre Dame Press, 19–38.

Kitcher, Philip (1993), *The Advancement of Science*. Oxford: Oxford University Press.

Laudan, Larry (1990), "Demystifying Underdetermination," in C. Wade Savage (ed.), *Minnesota Studies in Philosophy of Science* (Vol. XIV). Minneapolis: University of Minnesota Press, 267–297.

Laudan, Larry, and Jarrett Leplin (1991), "Empirical Equivalence and Underdetermination," *The Journal of Philosophy* 88: 449–472.

Longino, Helen (1990), *Science as Social Knowledge*. Princeton, NJ: Princeton University Press.

Quine, Willard Van Orman (1953), "Two Dogmas of Empiricism," in *From a Logical Point of View* Cambridge, MA: Harvard University Press, 20–46.

——— (1970), "On the Reasons for the Indeterminacy of Translation," *Journal of Philosophy* 67:178–183.

Stanford, P. Kyle (2001), "Refusing the Devil's Bargain: What Kind of Underdetermination Should We Take Seriously?" *Philosophy of Science* 68 (Proceedings): S1–S12.

van Fraassen, Bas C. (1980), *The Scientific Image*. Oxford: Clarendon Press.

See also **Conventionalism; Duhem Thesis; Logical Empiricism; Scientific Change; Scientific Progress; Theories**

UNIFICATION

See **Explanation; Reductionism; Unity and Disunity of Science**

UNITY AND DISUNITY OF SCIENCE

What kinds of integration are manifested, or sought after, in the claims and practices of the different sciences? This question should be carefully distinguished from any of the different specific theses addressing it, and yet it should be stressed as the linking thread of a time-honored philosophical debate. The question belongs to a tradition of thought that can be traced back to pre-Socratic

Greek cosmology, in particular to the preoccupation with the question of the one and the many. In what senses are the world and, thereby, our knowledge of it one? A number of representations of the world in terms of a few simple constituents considered fundamental emerged: Parmenides' static substance, Heraclitus' flux of becoming, Empedocles' four elements, Democritus' atoms, Pythagoras' numbers, Plato's forms, and Aristotle's categories. The underlying question of the unity of our types of knowledge was explicity addressed, for instance, by Plato: "Surely science too is one, but that which ranges as a part over some bit of it, once it is made distinct (isolated), each severally gets a name peculiar to itself. It's for this reason that arts and sciences are spoken of as many" (Sophist, 257c, in Benardete 1986).

With the advent and expansion of Christian monotheism, the organization of knowledge reflected the idea of a world governed by the laws dictated by God, creator and legislator. From this tradition emerged encyclopedic efforts such as the *Etymologies*, compiled in the sixth century by Isidore, Bishop of Seville, the works of Ramon Llull in the Middle Ages, and of Petrus Ramus in the Renaissance. Llull introduced tree diagrams and forest encyclopedias organizing different disciplines (including law, medicine, theology, and logic). He also introduced abstract diagrams in an attempt to encode combinatorially the knowledge of God's creation in a universal language of basic symbols; their combination would then generate knowledge of the secrets of creation. Ramus introduced diagrams representing dichotomies and gave prominence to the view that the starting point of all philosophy is the classification of the arts and sciences. The search for a universal language would continue to be a driving force behind the project of unifying knowledge.

The emergence of a distinctive tradition of scientific thought addressed the question of unity through science's designation of a privileged method, set of concepts, and language. In the late sixteenth century Francis Bacon held that a unity of the sciences was a result of the organization of material facts in the form of a pyramid with different levels of generalities; these would be classified in turn according to disciplines linked to human faculties. In accordance with the Pythagorean tradition as well as with the Bible's dictum in the Book of Wisdom, Galileo proclaimed at the turn of the seventeenth century that the Book of Nature had been written by God in the language of mathematical symbols and geometrical truths; and in it the story of Nature's laws was told in terms of a reduced set of primary qualities: extension, quantity of matter, and motion. In the seventeenth century, mechanical philosophy and Newton's work became the most promising framework for the unification of natural philosophy. After the demise of Laplacian molecular physics in the first half of the nineteenth century, this role was taken over by energy physics.

Descartes and Leibniz gave this tradition a rationalist twist centered on the powers of human reason; it became the project of a universal framework of exact categories and ideas, a *mathesis universalis*. Like Llull's, their conception of unity was determined by rules of analysis of ideas into elements and their synthesis into combinations. According to Descartes, the science of geometry, with its demonstrative reasoning from the simplest and clearest thoughts, constitutes the paradigm for the goal of unifying all knowledge. Leibniz proposed a General Science in the form of a Demonstrative Encyclopaedia. This would be based on a "catalogue of simple thoughts" and an algebraic language of symbols, *characteristica universalis*, which would render all knowledge demonstrative and allow disputes to be resolved by precise calculation.

Belief in the unity of science, along with the universality of rationality, was at its strongest during the European Enlightenment. The most important expression of the encyclopedic tradition came in the mid-eighteenth century from Diderot and d'Alembert (1751–1772), editors of the *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers*. Following earlier classifications by Nichols and Bacon, their diagram presenting the classification of intellectual disciplines was organized in terms of a classification of human faculties. Diderot stressed in his own entry, "Encyclopaedia," that the word signifies the unification of the sciences. The function of the encyclopedia was to exhibit the unity of human knowledge. Diderot and d'Alembert, in contrast with Leibniz, made classification by subject primary, and introduced cross-references instead of logical connections.

For Kant the unity of science was not the reflection of a unity found in nature; rather, it had its foundations in the unifying nature or function of concepts and of reason itself. Kant saw one of the fuctions of philosophy as determining the precise unifying scope and value of each science. For instance, he contrasted the methods employed by the chemist, organized by empirical regularities, with those employed by the mathematician or physicist, organized by a priori laws, and held that biology is not reducible to mechanics (as the former involves explanatons in terms of final causes). A devoted follower of Newton's achievements and insights, he maintained through most of his life that mathematization and a priori universal laws were preconditions for genuine scientific character (like Galileo and Descartes earlier, and Carnap later, Kant believed that mathematical exactness constitutes the main condition of the possibility of objectivity). By the end of his life, after having become acquainted with Lavoisier's achievements in chemistry, Kant thought of the unification of physics and chemistry not so much in terms of mathematization but, rather, in terms of the a priori principles regarding the properties of a universal ether (Friedman 1992). With regard to biology—insufficiently grounded in the fundamental forces of matter—its inclusion requires the introduction of the idea of purposiveness. More generally, for Kant, unity was a regulative principle of reason, that is, an ideal guiding the process of inquiry toward a complete empirical science with its empirical concepts and principles grounded in the so-called concepts and principles of understanding that constitute and objectify empirical phenomena.

Kant's ideas set the frame of reference for discussions of the unification of the sciences in German thought throughout the nineteenth century. He gave philosophical currency to the notion of worldview (*Weltanschauung*) and, indirectly, worldpicture (*Weltbild*), thereby establishing among philosophers and scientists the unity of science as an intellectual ideal. (In the German intellectual tradition culminating in philosophers such as Windelband, Ricket, and Dilthey, a worldview often included elements of evaluation and life meaning.) This tradition influenced the physicists Max Planck and Ernest Mach, who engaged in a heated debate about the precise character of the unified scientific worldpicture, and culminated in the first two decades of the twentieth century with the work of Albert Einstein. Mach's (1897) view, which was more influential, was phenomenological and Darwinian: The unification of knowledge took the form of an analysis of ideas into elementary sensations (neutral monism) and was ultimately a matter of adaptive economy of thought (see Mach, Ernest). Planck adopted a view that took as fundamental the principles of energy and entropy. These worldpictures constituted some of the alternatives to a long-standing mechanistic view that since Newton had affected biology as well as most branches of physics. In the same German tradition, amidst the proliferation of books on the unity of science, the German energeticist Wilhelm Ostwald (1913) declared the twentieth century the "Monistic century."

In the twentieth century the unity of science is a distinctive theme of the scientific philosophy of logical positivism. The logical positivists, notably the members of the Vienna Circle, adopted the Machian banner of "unity of science without metaphysics," a model of unity based on a demarcation between science and metaphysics, consisting of a unity of method and language that included all the sciences, natural and social (see Vienna Circle). Notice that a common method does not imply a more substantive unity of content involving theories and their concepts. A stronger model, recommended by Rudolf Carnap (1934), was seen in Hilbert's axiomatic approach to formulating theories in the exact sciences and in Frege's and Russell's logical constructions in mathematics. This model was predicated on the formal values of simplicity, neutrality, and objectivity and was characterized by axiomatic structures and rigorous reductive logical connections between concepts and laws of the different sciences at different levels. Physics, with its genuine laws, was fundamental and lay at the base of the hierarchy. Because of the emphasis on the formal and structural properties of representations, their individuality, like that of nodes in a railway network, was determined by their place in the whole structure and, hence, presupposed connective unity. Alternatively, all scientific concepts could be constructed out of classes of elementary experiences—not atomic in the Machian sense—but derived from the field of experience as a complex whole in the manner proposed by Gestalt psychology. This construction of scientific knowledge took into account the possibility of empirical grounding of theoretical concepts and testability of theoretical claims. Carnap was influenced by the empiricist tradition (especially Russell and Mach) and the ideals of simplicity and reductive logical analysis in the early works of Russell and Wittgenstein. From the point of view of a formalistic neo-Kantian tradition, Carnap's models of unity express his concern with the possiblity of objectivity of scientific knowledge.

Otto Neurath, by contrast, favored a more realistic and less reductive model of unity predicated on the complexity of empirical reality. He spoke of an "encyclopedia model" instead of the classic ideal of the pyramidic "system model" (see Neurath, Otto). The encyclopedia model took into account the presence within science of ineliminable and imprecise terms from ordinary language and the social sciences and emphasized a unity of language and the local exchanges of scientific tools (specifically, Neurath stressed the material-thing-language called "physicalism," not to be confounded with the emphasis on the vocabulary of physics). Thus, it was not constrained by Carnap's ideals of conceptual precision, deductive systematicity, and logical rigor; it was meant as a tool for cooperation and was motivated by the need for successful treatment (prediction and control) of complex phenomena that involved properties studied by different theories or sciences, in a conception of unity of science at the point of action. Neurath spoke of a "boat," a "mosaic," an "orchestration," a "universal jargon" (see Unity of Science Movement). For both Carnap and Neurath, the ideal of unified science had deep social and political significance. It supported a kind of thinking and, in particular, a social and political ideology, that was free of the obscurantist metaphysics responsible for much social unhappiness. At the same time, Karl Popper was defending a demarcation criterion based on the falsifiability of all genuinely scientific propositions (see Popper, Karl Raimund).

After World War II, a discussion of unity engaged philosophers and scientists in the Inter-Scientific Discussion Group in Cambridge, Massachusetts (founded by Philip Frank, himself one of the founders of the Vienna Circle), which would later become the Unity of Science Institute. The group was both an extension of the Vienna Circle and a reflection of local concerns in a culture of computers and nuclear power (Galison 1998a). The characteristic feature of the new view of unity was the idea of cross-fertilization, instantiated in the creation of war-boosted interdisciplines such as cybernetics, computation, electroacoustics, psychoacoustics, neutronics, game theory, and biophysics.

Two new views developed by logical positivists in the United States again placed the question of unity of science at the core of philosophy of science: Carl Hempel's (1965) deductive-nomological model of explanation and Ernest Nagel's (1961) model of reduction (see Explanation; Reductionism). Hempel's model characterizes the scientific explanation of events as a subsumption under an empirically testable generalization. In the 1950s, when positivism was extending to the social sciences, the model was offered as a criterion of demarcation. Explanations in the historical sciences too had to fit the model if they were to count as scientific. The applicability of Hempel's model was soon contested, notably by William Dray, and this opened a debate about the nature of the historical sciences that remains unresolved. In the process, some have claimed as historical some of the natural sciences such as geology and biology. It has been argued that Hempel's model, especially the requirement of empirically testable strict universal laws, is satisfied neither in the physical sciences nor in the historical sciences, including biology. Nagel's model of reduction was a model of explanation as well as of scientific progress. It required extensional equivalence between descriptions; bridge principles between coextensive but distinct terms in different theories; and a deductive relation between the laws involved in the reduction. Feyerabend promptly rejected the demand of extensional equivalence as inadequate for "meaning invariance" and, moreover, believed such meaning invariance between theories to be unattainable.

Since Nagel's influential model of reduction by derivation, most dicussions of the unity of science have been cast in terms of reductions between concepts and between theories. The hierarchy of levels of reduction was set by the levels of aggregation of entities all the way down to atomic particles, thus rendering microphysics the fundamental science. Oppenheim and Putnam (1958) intended to articulate an ideal of science as a unity reductive to the most elementary concepts and laws. They also defended the empirical claim that the evolution of science manifested a trend in that direction. The rejection of such models and their emendations has occupied the last three decades of philosophic discussion about unity in physics, and especially in psychology and biology. Initially the availability of laws and theories and the feasibility of global reductions within the latter two was contested. The status, form, and interpretation of bridge principles connecting different levels was called into question. Arguments concerning new notions such as supervenience and multiple realizability by Putnam, Kim, Fodor (1974), and others led to a distinction between type-type and token-token reductions and the examination of its implications (see Reductionism). The possibility and the necessity of reductions in a number of forms have been widely discussed in connection with issues of explanation, realism, and scientific progress. A number of discussions of specific sciences have addressed the question of the inconsistencies between some of their theories or models—for instance, the problem of the relation between Newtonian mechanics and thermodynamics, or between relativity theory and quantum physics as the conceptual foundation of quantum gravity (Maudlin 1996). Other discussions have focused on the search for unification—for instance, of quantum field theory and cosmology in physics, or of theories of evolution and development in biology. A focus of some projects has been on the value and variety of kinds of approximations and the difficulties of extending to different theories specific methodological standards, models of explanation, and even abstract categories such as rationality, lawlikeness, determinism, causality, locality, or individual and separable particles (see Causality; Determinism; Laws of Nature).

Debates on unification and reduction have taken place as part of discussions of specific sciences and also within the sciences themselves. The success of the unified theories of fundamental forces in the 1970s revived the ideal of reductionism in physics. The view has found its most vocal proponents among elementary particle physicists such as Steven Weinberg (1993) and has been contested by condensed-matter physicists such as Philip Anderson (1972). What is fundamental? Can there be unity without fundamentality? The form that unity, especially in physics, takes and should take is a controversial matter that has led to pluralism within the physics community (Batterman 2002). At the same time, along with a common (unifying) concern, techniques and models are shared even if their significance is understood differently (Cat 1998; Galison 1998b; Maudlin 1996). It has also been argued that unity is not a form of explanatory power and cannot guarantee an ontological unity of nature (Morrison 2000).

Similar ideals and debates have also regained force in biology, especially in molecular genetics and sociobiology, in light of successes in evolutionary and genetic explanations and the interest in genetic mapping, exemplified by the Human Genome Project. Beyond the attempts earlier in the century to synthesize genetics and evolutionary biology in the so-called evolutionary synthesis, the renewed reductionist trend has been explained in part by the personal and intellectual connection between biology and physics (Keller 1990). In fact, the connection between the trends in both fields can be traced to the institutional links between the Manhattan Project and the Human Genome Project. Proponents of reductionist views such as E. O. Wilson (1999) have encountered the opposition of other biologists such as Ernest Mayr (1982) and R. C. Lewontin (1992). This atmosphere has stimulated debates and detailed work in philosophy of physics and of biology, and more recently the emergence of philosophy of chemistry.

In antireductionist quarters, models of unification without reduction have appeared. For instance, the notion of "interfield theories" (Darden and Maull 1977) is based on the idea that theories and disciplines do not match neat levels of organization within a hierachy; rather, many, in fact, cut across different such levels. Others have emphasized the idea of science as a process, manifesting a historical unity with a Darwinian-style pattern of evolution (Hull 1988). Concepts of unity have also been defended on the cognitive grounds that unification, measured as the number of independent laws conjoined in a theoretical structure, contributes understanding and confirmation (Friedman 1974) or explanatory power in terms of few derivation or argument patterns (Kitcher 1981). Distinctions have been introduced among different categories of reduction and reductionism, such as global/partial, constitutive/theoretical/explanatory, and ontological/epistemological (Mayr 1982; Schaffner 1993; Sarkar 1998). More recent proposals address concerns with complexity and emergence and locate unity in the common framework of abstract categories and idealizations for characterizing systems running through physics, biology, and economics (Auyang 1998).

A more radical departure is the recent criticism of the methodological values of reductionism and unification in science and its position in and effect upon society. This view argues for the replacement of the emphasis on global unity—including unity of method—by an emphasis on disunity and epistemological and ontological pluralism. Some suggestions link disunity with instrumentalism about higher-level sciences such as biology and sociology (Rosenberg 1994). Another picture comes from the members of the so-called Stanford School, such as John Dupré (1993), Ian Hacking (1996), Peter Galison (1998b), and Nancy Cartwright (1999). Dupré (1993) has argued that the disunity of science can be given adequate metaphysical foundations that make pluralism compatible with realism. He criticizes a mechanistic paradigm characterized by determinism, reductionism, and essentialism and defends the views that science depends on metaphysical assumptions and that scientific and nonscientific empirical inquiries suggest that science does not and cannot constitute a unified single project. This is supported, in turn, by three pluralistic theses: (1) Against essentialism, there is always a plurality of classifications of reality into kinds; (2) against reductionism, there exists equal reality and causal efficacy of systems at different levels of description; and (3) against epistemological monism, there is no single methodology that supports a single criterion of scientificity, nor a universal domain of its applicability, but only a plurality of epistemic and nonepistemic virtues. The concept of science should be understood, following the later Wittgenstein, as a family-resemblance concept.

Hacking distinguishes a plurality of scientific styles to argue for a disunity of science (Hacking 1996) (see Scientific Style). Maxwell had remarked in the nineteenth century that unity of science is the "cross-fertilization of the sciences." Galison has replaced traditional analyses that consider the observational domain as homogeneously underpinning theoretical structures and their developments and replacements by an anthropological analysis of the subcultures of experimentation. He explains the strength, coherence, and continuity of science in terms of local coordinations of intercalated symbolic procedures and meanings, instruments, and arguments, which he calls "trading zones" (Galison 1998b). Cartwright has argued that laws cannot be both universal and true; there exist only patchworks of laws and local cooperations. Like Dupré, Cartwright adopts a kind of scientific realism but denies that there is a universal order, whether represented by a theory of everything or a corresponding metaphysical principle. The empirical evidence, she argues, suggests far more strongly the idea of a patchwork of laws, often in local cooperation. Theories apply only where and to the extent that their interpretive models fit the phenomena studied (Cartwright 1999). She explains their more or less general domain of application in terms of causal capacities and arrangements she calls "nomological machines." On these grounds she rejects strong distinctions between natural and social sciences. Whether as a hypothesis or as an ideal, the debates continue over the form, scope, and significance of unification in the sciences.

Jordi Cat

References

Anderson, P. W. (1972), "More Is Different," *Science* 177: 393–396.

Auyang. S. (1998), *Foundations of Complex-System Theories*. New York: Cambridge University Press.

Batterman, R. (2002), *The Devil in the Details*. New York: Oxford University Press.

Benardete, S. (1986), Plato: Sophist—Translation and commentary. Chicago: University of Chicago Press.

Carnap, R. (1934), *The Unity of Science*. London: Kegan Paul, Trench, Trubner & Co.

Cartwright, N. (1999), *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Cat, J (1998), "The Physicists' Debates on Unification in Physics at the End of the 20th Century," *Historical Studies in the Physical and Biological Sciences* 28: 253–300.

d'Alembert, J., and D. Diderot (eds.) (1751–1772), *Encyclopédie, ou dictionnaire raisonné des sciences, des arts at des metiers*. Paris: Plon.

Darden, L., and N. Maull (1977), "Interfield Theories," *Philosophy of Science* 44: 43–64.

Dupré, J. (1993), *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Fodor, J. (1974), "Special Sciences, or the Disunity of Science as a Working Hypothesis," *Synthese* 28: 77–115.

Friedman, M. (1974), "Explanation and Scientific Understanding," *Journal of Philosophy* 71: 5–19.

——— (1992), *Kant and the Exact Sciences*. Cambridge, MA: Harvard University Press.

Galison, P. (1998a), "The Americanization of Unity," *Daedalus* 127: 45–71.

——— (1998b), *Image and Logic*. Chicago: University of Chicago Press.

Hacking, I. (1996), "The Disunities of Science," in P. Galison and D. Stump (eds.), *The Disunity of Science: Boundaries, Contexts, and Power*. Stanford, CA: Stanford University Press, 37–74.

Hempel, C. (1965), *Aspects of Scientific Explanation*. New York: Free Press.

Hull, D. (1988), *Science as Progress*. Chicago: University of Chicago Press.

Keller, E. F. (1990), "Physics and the Emergence of Molecular Biology: A History of Cognitive and Political Synergy," *Journal of the History of Biology* 23: 389–409.

Kitcher, P. (1981), "Explanatory Unification," *Philosophy of Science* 48: 507.

Lewontin, R. C. (1992), *Biology as Ideology*. Chicago: University of Chicago Press.

Mach, E. (1897), *Analysis of the Sensations*. Chicago: Open Court.

Maudlin, T. (1996), "On the Unification of Physics," *Journal of Philosophy* 93: 129–144.

Mayr, E. (1982), *The Growth of Biological Thought*. Cambridge, MA: Harvard University Press.

Morrison, M. (2000), *Unifying Physical Theories: Physical Concepts and Mathematical Structures*. New York: Cambridge University Press.

Nagel, E. (1961), *The Structure of Science*. New York: Harcourt, Brace and World.

Oppenheim, P., and H. Putnam (1958), "The Unity of Science as a Working Hypothesis," in H. Feigl, M. Scriven, and G. Maxwell (eds.), *Concepts, Theories, and the Mind–Body Problem: Minnesota Studies in the Philosophy of Science*, vol. 2. Minneapolis: University of Minnesota Press.

Ostwald, W. (1913), *Monism as the Goal of Civilization*. Hamburg: International Committee of Monism.

Rosenberg, A. (1994), *Instrumental Biology, or the Disunity of Science*. Chicago: University of Chicago Press.

Sarkar, S. (1998), *Genetics and Reductionism*. New York: Cambridge University Press.

Schaffner, K. (1993), *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.

Weinberg, S. (1993), *Dreams of a Final Theory*. New York: Viking Press.

Wilson, E. O. (1999), *Consilience*. Cambridge, MA: Harvard University Press.

See also **Logical Empiricism; Physicalism; Reductionism; Unity of Science Movement; Vienna Circle**

UNITY OF SCIENCE MOVEMENT

The ideas and projects that constituted the unity of science movement were launched in 1934 as part of the intellectual life of the Vienna Circle (see Vienna Circle). The movement was spearheaded by Otto Neurath, although the theme of unity of science had been central to the concerns of other members of the Circle since its inception. In the founding manifesto of the Vienna Circle, Neurath, Hans Hahn, and Rudolf Carnap (1929) laid out a scientific world-conception and a constructive agenda that would combat the obscurantist metaphysical and theologizing trends in German culture, proclaiming that "the goal ahead is unified science." They also stressed the importance of unity of science for social and political life. In particular, Neurath recurringly wrote on the connections between the project of unity of science and the movements for economic socialization, educational reform, and peaceful and cooperative internationalization and unification of mankind.

At the end of the Eighth International Congress of Philosophy held in Prague in September 1934, Neurath proposed a series of International Congresses for the Unity of Science. These took place in Paris in 1935 and 1937; Copenhagen in 1936; Cambridge, England, in 1938; Cambridge, Massachusetts, in 1939; and Chicago in 1941. For the organization of the congresses and related activities, Neurath founded in 1936 the Unity of Science Institute, renamed in 1937 the International Institute for the Unity of Science, a special department of his Mundaneum Institute at The Hague. Neurath had founded the Mundaneum in 1934, after fleeing Vienna, and it already included the International Foundation for Visual Education, founded in 1933. The Institute's executive committee was composed of Neurath, Philip Frank, and Charles Morris. The Organization Committee for the International Congresses for the Unity of Science was composed of Neurath, Carnap, Frank, Joergen Joergensen, Morris, Louis Rougier, and Susan Stebbing.

In the contexts of scientific philosophy, unity of science, and social and educational reform, Neurath's work on museums and exhibits was devoted to the divulgation of information—typically of a social and demographic nature—in statistical and pictorial languages, known as the Vienna Method. The pictorial language was renamed in 1937 ISOTYPE (International System of Typographic Picture Education) (Reich 1995).

The central and most emblematic project of the Unity of Science Movement was the publication of the *International Encyclopedia of Unified Science*. The Encyclopedia was an early idea of Neurath's, which he had discussed with Einstein, and was inspired by the French *Encyclopédie* of the eighteenth century. It was meant to promote international cooperation among scientists through the exchange of methods and concepts from different fields of inquiry and instantiate an ideal model of society "a republic of scientists." The editorial board was composed of Neurath as editor-in-chief and Carnap and Morris as assistant editors. Morris negotiated its publication by the University of Chicago Press. It would consist of twenty volumes each containing ten monographs of historical and foundational content (with a systematic logical analysis of the sciences and a survey of their development, stressing connections to other disciplines and suggesting new directions). Eventually, between 1937 and 1970 twenty monographs were published, including T. S. Kuhn's *The Structure of Scientific Revolutions*.

After World War II, the Unity of Science Movement continued and developed around Philip Frank's Inter-Scientific Discussion Group, created in 1944 in Cambridge, Massachusetts, and then his Institute for the Unity of Science, under the auspices of the American Academy of Arts and Sciences in Boston, with support from the Rockefeller Foundation and royalties from the Encyclopedia through Marie Neurath's transfer of her inheritance rights. The list of associates and participants in discussions included Birkhoff, Bridgman, Carnap, Conant, Feigl, Frank, Hempel, Holton, Jakobson, Leontieff, Morris, Nagel, Quine, Santillana, Skinner, and Wiener. The Institute held several conferences, planned a dictionary presenting the operational meaning of three hundred concepts, promoted the development of new interdisciplines such as cybernetics, neutronics, and electroacoustics, and introduced enhanced integrated models of university general education (adopted by Conant at Harvard). Gradually the movement, as well as the political engagement of logical empiricism, suffered in the political climate of the McCarthy era, losing funding while philosophy of science emphasized logical or formal dimension and value neutrality. In 1958 the Institute moved with Feigl to the Minnesota Center for Philosophy of Science, being replaced in Boston by what became the Boston Colloquium for Philosophy of Science, organized by Robert Cohen and Marx Wartofsky at Boston University. Finally, in 1973 what remained of the Institute was absorbed into the Philosophy of Science Association of the United States.

Jordi Cat

References

Neurath, O., H. Hahn, and R. Carnap (1929), *Wissenschaftliche Weltauffassung: Der Wiener Kreis*. Vienna: Vienna Circle.

Reich, G. (1995), *A History of the* International Encyclopedia of Unified Science. Ph.D dissertation, University of Chicago.

See also **Logical Empiricism; Neurath, Otto; Unity and Disunity of Science; Vienna Circle**

V

VERIFIABILITY

Verifiability is often linked to questions about empirically verifiable statements as a basis for claims to knowledge, and hence it is also linked to notions like verifiably true, confirmation, and inductive verification, as well as to disputes about "foundationalism" (Reichenbach 1976). But it is usually connected with the logical positivists' verifiability criterion (theory) of cognitive meaning, which was derived from the classical empiricists' analyses of traditional notions of material substance and causality and their accounts of the origin of ideas.

Berkeley argued for idealism by claiming that, as one does not and cannot have an idea of material substance, one cannot meaningfully assert that physical objects exist independently of an apprehending mind. Hume developed the empiricist theme by construing causation so as to dismiss mental substance as a causal agent while, as Berkeley had done earlier, rejecting causal accounts of perceptual experience in terms of material objects, as well as claims about causal powers linking physical events and objects. The arguments of the empiricists developed the theme, implicit in earlier philosophical works, that ideas are of two types: *simple ideas*, which are not composed of other ideas; and *complex ideas*, which are composed of other ideas. Complex ideas can be associated with defined terms of a language, while simple ideas can be seen as correlates of primitive or basic terms. Berkeley took simple ideas to be derived from experience. Thus, his dismissal of the term 'matter' as meaningless depended on denying the experience of material substance (as Descartes had done in his often cited discussion of a melting piece of wax) and on claiming that there was no tenable analysis of the concept. It was meaningless to take material substance as Locke's "I know not what" supporting the properties of an object.

For Hume, the linguistic base of undefined terms was more closely restricted to what was presented in experience, and simple ideas became copies of sensory and "inner" impressions. Since complex ideas still resulted from combinatorial and associative mental operations, one could have an idea of, say, a golden mountain or a centaur but could not have an idea of matter, or of causal powers linking material objects and events. Nor could one have a Berkeley-like "notion" of mental substance. The logical positivists (empiricists) of the Vienna Circle were influenced by Russell's (1956) "principle of acquaintance" (58), which required every understandable proposition to be composed of constituents that one was already acquainted with. Accordingly, they attempted to formulate a tenable verifiability criterion for meaning that would dismiss classical philosophical (metaphysical) statements as meaningless while retaining the statements of science. This conformed to an intellectual current of the time that held that "philosophy of science is philosophy enough" (Quine 1953, 446, cited out of context) and, like Hume, relegated metaphysical works to "the flames."

The focus on verifiability was not just a reaction against traditional philosophy. Einstein had brought on a conceptual shock, which some people attributed, in part, to the introduction of meaningless concepts into the conceptual framework of scientific theories. One example was the notion of "absolute (nonlocal) simultaneity"; no such conceptual shock would have resulted if temporal concepts had been tied to specified procedures of measurement. Construing simultaneity in suitable terms would ensure speaking of temporal events relative to a framework of "clocks." The verifiability criterion was designed to provide a conceptual cushioning against any further shocks. But it bred an extreme offspring, operationism. As the term implies, this required meaningful scientific terms to be defined operationally, that is, by conditional statements with antecedents specifying an operational procedure (e.g., using a thermometer) and consequents specifying a resulting observation (e.g., a reading of a thermometer). Operationism spread, and "positivists" in the human, or behavioral, sciences argued that if these fields were to rise to scientific respectability, they had to be purged of meaningless concepts (Bergmann 1954; Bridgman 1927; Feigl 1949). An empiricist criterion of meaning was seen as a key to the door of scientific development.

The verifiability criterion ran into obvious problems. No statement of general law, covering "all" cases of some kind, could be verified, even in principle. At best, one could talk in terms of obtaining stronger degrees of confirmation or, like Karl Popper, in terms of falsifiability.

An early reformulation of verifiability in a popular version is found in Ayer's (1946) *Language, Truth, and Logic*. Building on a familiar distinction, Ayer took all "cognitive" statements to be either synthetic or analytic. Analytic statements were those of mathematics and logic, along with stipulative "definitions." (Misconstruing Frege and Russell's diverse forms of logicism, Ayer took arithmetical truths to be true by definition.) With the addition of statements true by the "semantic" rules of a language, these would correspond to Carnap's (1956) *L*-truths. Synthetic statements were divided into two kinds. Observation statements, such as "This is a swan" (O_1) and "This is white" (O_2), could be "verified" as true or false by observation. Nonobservation statements, such as a generality like "All swans are white" (G), would qualify as empirically meaningful if they were appropriately linked to observation statements. Thus, a statement Q would be meaningful if the conjunction of Q and some observation statement O logically entailed an observation statement O^* that was not entailed by O. Statement G would then qualify, because the conjunction of G and O_1 entails O_2, which is not entailed by O_1 (Ayer 1946, 13).

Ayer's criterion did not succeed in ruling out a variety of unacceptable statements. (For example, a conditional with O_1 as antecedent and a conjunction of O_2 and any meaningless statement as consequent would qualify as empirically meaningful.) Numerous attempts to repair the criterion elicited counterexamples exposing deficiencies in the revisions (Church 1949). Another problem was that the statement of the criterion did not itself satisfy the criterion. One response to this objection, derived from considerations of "liar"-type paradoxes, was that no statement could be expected to apply to itself. Another response was that the criterion was a "proposal," not a claim, and therefore did not have a truth value.

Hempel's attempt to revitalize the criterion was more radical. Developing early ideas of Russell and Carnap, he took a cognitively significant statement as one that could be transcribed into an empiricist language grammatically modeled on Russell and Whitehead's *Principia Mathematica*. Hempel considered Russell's principle, the "requirement of definability," too restrictive; he required only some descriptive (nonlogical) primitive predicates of such a schema to be correlated with observable properties. Hempel sought to resolve two further problems facing the empiricists' verifiability criterion, posed by dispositional predicates (such as 'temperature' and 'soluble') and by theoretical terms (such as 'electron'). Both kinds of terms were empirically meaningful: The former could be embedded in an empiricist language as terms "implicitly defined" by means of reduction sentences like Carnap's, and the latter could be connected to an empiricist ("observation") language as "theoretical constructs" belonging to a "partially interpreted" calculus in which some statements were correlated with those of an empiricist language (Carnap 1936–1937, 1937, 1956, and 1967; Hempel 1951 and 1952). But Hempel did not resolve the philosophical issues posed by causation and dispositions, and questions arose about the justification of his criterion for cognitive significance, along with technical issues regarding the reduction of theoretical terms through explicit definitions or reduction sentences (Bergmann 1951).

Dispositions, Laws, and Reduction Sentences

Carnap's and Hempel's logical apparatus included the familiar truth-table conditional used in lawful generalities like $(\forall x)(Fx \to Gx)$ and in definitions of dispositional predicates. That gave rise to familiar problems derived from the "paradoxes" of material implication. Defining 'is soluble' *S* in terms of dissolving *D* if put into water *P* would force one to take a destroyed piece of wood, or any object never put into water, as satisfying the defining clause. Carnap proposed resolving the problem by introducing dispositional predicates into an empiricist language by means of pairs of reduction sentences. The simplest case is illustrated by the pair

$$Px \to (Dx \to Sx)$$

$$Px \to (\neg Dx \to \neg Sx)$$

which introduces but does not define *S*. A burned wooden match that was never put into water need not then be taken as soluble.

But Carnap's analysis raised new problems. As *S* is introduced by a reduction pair or a set of reduction pairs, it is supposedly implicitly defined by such "semantic rules." Yet implicit definitions amount to axioms using an uninterpreted term, of which they provide neither an analysis nor an explication, irrespective of the transcription of *S* by the ordinary-language expression 'is soluble.' Moreover, by using various reduction pairs for a predicate, Carnap (1956) was forced to take factual claims as "rules" for the use of the term (228). Reduction sentences, unlike designation rules or explicit definitions, are problematic candidates for semantic rules. Unlike a theoretical term such as 'electron,' 'soluble' is a philosophically problematic term linked to questions about dispositions, powers, and causality. Introducing it into an empiricist schema as a primitive term would seem to introduce such problems as well, the resolution of which, along empiricist lines, requires the elimination or analysis of such terms. This can be done partly by appealing to lawful generalizations. What problems remain are about the laws themselves, which must be faced in any case (see Laws of Nature).

When it is said of a piece of sugar that it is soluble, it is implicitly asserted that sugar is soluble: that the object is of a soluble kind, sugar. Dispositional predicates like 'soluble' are used in a context of such lawful generalities. Construing causal laws in terms of regularities, one will also use a standard conditional and reject a primitive causal conditional, expressing either a basic causal relation (nomic necessity, i.e., having the force of natural law) or a primitive counterfactual conditional (had *p* occurred, *q* would have occurred). Neither would fit the conditions for an empiricist language using a standard conditional and requiring "modal" concepts to be construed metalinguistically (Carnap 1956, 175, 243; Hochberg 1981; Reichenbach 1976).

Consider the statement that sugar *S** is soluble. The problematic ordinary-language statement can be unproblematically analyzed in terms of the claim that

$$(\forall x)(S * x \to (Px \to Dx)) \tag{1}$$

is a lawful generality. The assertion that wood *W* is not soluble can be taken, in one sense, as denying that

$$(\forall x)(Wx \to (Px \to Dx)) \tag{2}$$

is a statement of law. This is distinguished from the claim that wood is insoluble, analyzed as the claim that

$$(\forall x)(Wx \to (Px \to \neg Dx)) \tag{3}$$

is a lawful generality. Of course, (3) is what one might mean by saying that wood is not soluble, but this is irrelevant to the fact that the distinctions resolve the problems posed by dispositional predicates, without introducing the problems raised by Carnap's reduction sentences. Purported dispositional properties and problematic primitive predicates are avoided by acknowledging various lawful generalities and denials of such generalities.

'Soluble' and 'insoluble' can be said to constitute a disposition pair. Structurally, it is like a Carnapian reduction pair in that both $Px \to Dx$ and $Px \to \neg Dx$ are required, as in (1) and (3) above. The point can be emphasized when we note that the Carnapian reduction pair for *S*, used earlier, is logically equivalent to the pair $Sx \to (Px \to Dx)$ and $\neg Sx \to (Px \to \neg Dx)$. Thus a reduction pair can be considered as implicitly acknowledging the distinction between not-soluble and insoluble, and hence the disposition pair.

However, one need not take various pairs of Carnapian reduction sentences, connecting diverse "tests" for a single dispositional predicate, as semantic rules for such a term. One also avoids introducing terms like *S* as primitive terms, since one can introduce diverse explicit definitions of different predicates, such as *S* and *I*, or $Sx = df\ Px \to Dx$, or $Ix = df\ Px \to \neg Dx$. This is unproblematic if one does not take *S* and *I* as transcriptions of the problematic terms 'soluble' and 'insoluble' (as Carnap does with *S*) and does not take the definitions of *I* and *S* as providing analyses of the problematic terms (as Carnap's reduction sentences purport to do). The context of the relevant laws, like (1) and (3), and further laws connecting the various explicitly defined terms (specifying other procedures for determining solubility), will furnish the analyses. This makes explicit the point that the empirical connections between the various reduction pairs Carnap and Hempel use for a single primitive term are not semantic rules (Bergmann 1951; Hochberg 1967; Reichenbach 1976).

HERBERT HOCHBERG

References

Ayer, Alfred Jules (1946), *Language, Truth, and Logic*, 2nd ed. London: Golancz.

Bergmann, Gustav (1951), "Comments on Professor Hempel's 'The Concept of Cognitive Significance,'" *Proceedings of the American Academy of Arts and Sciences* 80: 78–86.

——— (1954), "Sense and Nonsense in Operationism," *Scientific Monthly* 79: 210–214.

Bridgman, Percy (1927), *The Logic of Modern Physics*. New York: Macmillan.

Carnap, Rudolf (1936–1937), "Testability and Meaning," *Philosophy of Science* 3 and 4: 419–471 and 1–40.

——— (1937), *The Logical Syntax of Language*. London. Routledge and Kegan Paul.

——— (1956), *Meaning and Necessity*, 2nd ed. Chicago, IL: University of Chicago Press.

——— (1967), *The Logical Construction of the World and Pseudoproblems in Philosophy*. Translated by Rolf George. Berkeley and Los Angeles, CA: University of California Press.

Church, Alonzo (1949), "Review of Ayer's *Language, Truth, and Logic*, " *Journal of Symbolic Logic* 14: 52–53.

Feigl, Herbert (1949), "Operationism and Scientific Method," in Herbert Feigl and Wilfrid Sellars (eds.), *Readings in Philosophical Analysis*. New York: Appleton-Century-Crofts, 498–509.

Hempel, Carl Gustav (1951), "The Concept of Cognitive Significance," *Proceedings of the American Academy of Arts and Sciences* 80: 61–77.

——— (1952), "Problems and Changes in the Empiricist Criterion of Meaning," in Leonard Linsky (ed.), *Semantics and the Philosophy of Language*. Urbana: University of Illinois Press, 163–185.

Hochberg, Herbert (1967), "Dispositional Properties," *Philosophy of Science* 2: 184–204.

——— (1981), "Natural Necessity and Laws of Nature," *Philosophy of Science* 4: 386–399.

Quine, Willard Van (1953), "Mr. Strawson on Logical Theory," *Mind* 62, 433–451.

Reichenbach, Hans (1976), *Laws, Modalities, and Counterfactuals*. Berkeley and Los Angeles, CA: University of California Press.

Russell, Bertrand Arthur William (1956), *The Problems of Philosophy*. London: Oxford University Press.

See also **Ayer, Alfred J.; Bridgman, Percy; Carnap, Rudolf; Cognitive Significance; Hempel, Carl Gustav; Logical Empiricism; Schlick, Moritz**

VERISIMILITUDE

The concept of verisimilitude, or truthlikeness, attempts to capture the idea that one scientific theory may be closer to the truth than a rival theory. Karl Popper thought that this idea was a key to the analysis of scientific progress and the growth of knowledge. In spite of the failure of Popper's own attempt, many philosophers of science share the view that the possibility of explicating the notion of verisimilitude is vital to the success of scientific realism (see Popper, Karl Raimund; Scientific Realism).

Historical Introduction

The concept of verisimilitude was introduced within the ancient debates about scepticism. The official doctrine of the Academic skeptics was that, in order to avoid error, the wise man should not assent to any propositions. However, in the first century B.C.E., Carneades argued that some impressions may be more or less "convincing" or "persuasive," and the wise man may follow such '*pithane*' beliefs in practical action. The Greek term '*pithanon*' was translated by Cicero into Latin by '*probabile* and *veri simile*' (i.e., "like the truth").

When the modern calculus of probability was developed in the seventeenth century, the Latin terms '*probabilitas* and *verisimilitudo*' were used as synonyms. Similar terminology can be found in many Western languages. For example, in German, '*Wahrscheinlich*' means literally "truth seeming," and in Swedish, '*sannolik*' means "truthlike." The English words '*likely* and *likelihood*' belong to the same etymological family. The term '*veri simile*' occurs also in Cicero's poetic and rhetoric, indicating a narrative that is "possibly true" or "plausible" in the sense that it conforms to the expectations of the public. This doctrine is still alive in the notion of verisimilitude ('*vraisemblance*' in French) of modern theories of literature.

In the philosophy of science, Karl Popper introduced the notions of verisimilitude and truthlikeness in 1960 (Popper 1963). In developing his falsificationism, he wanted to find a concept that differs from the notions of probability and confirmation used in Rudolf Carnap's inductive logic (see Carnap, Rudolf; Inductive Logic). According to Popper's fallibilism, one never has any degree of certainty or even probability that the best theories in science are true. As an admirer of Tarski's definition of truth, which he interpreted as a satisfactory explication of the correspondence theory of truth, Popper's (1963) notion of verisimilitude was intended to express "the idea of a degree of better (or worse) correspondence to truth" (233).

Some nineteenth-century coherence theories of truth confused the notions of truth and complete truth by claiming that every partial truth is also partly false—for instance, "This rose is red" does not tell everything about this rose. Popper (1963) made it clear that his concept "represents the idea of approaching comprehensive truth" (237). It differs from probability, since it combines truth and information content, while "probability combines truth with lack of content" (ibid). Verisimilitude is an objective or semantic notion, not an epistemological one, but sometimes there may be arguments to appraise that "we may have made progress towards the truth" (Popper 1972, 58). Like Carneades and Cicero, Popper further suggested that, in cases where one must make a pragmatic preference between theories, it is rational to act upon that theory that can be claimed to be most truthlike.

Popper's attempted definition was refuted by David Miller (1974) and Pavel Tichý (1974). A new approach, based upon the notion of similarity, was then initiated by Tichý, Risto Hilpinen, and Ilkka Niiniluoto. Debates about the details of the definition of truthlikeness, including the possibility or desirability of such a definition, have continued ever since. The first monographs on the subject were published by Graham Oddie (1986), Niiniluoto (1987), and Theo Kuipers (1987). Surveys of more recent work are given by Niiniluoto (1998), Kuipers (2000), and Sjoerd Zwart (2001).

Popper's Definition

Popper's qualitative criterion is applicable to theories as deductively closed sets of statements. Let T and F be the sets of the true and false statements, respectively, in some interpreted language L. Then, for a theory A in L, the *truth content* of A is the intersection $A \cap T$, and the *falsity content* of A is the intersection $A \cap F$. According to Popper, theory A is *more truthlike* than theory B if and only if $B \cap T \subseteq A \cap T$ and $A \cap F \subseteq B \cap F$, where one of the set-inclusions is strict. Intuitively, A should have larger truth content than B, but smaller falsity content than B, or A should have (set-theoretically) more true consequences and fewer false consequences than B. An equivalent formulation of this criterion states that the symmetric difference $A\Delta T = (A - T) \cup (T - A)$ should be a proper subset of $B\Delta T$.

Popper's definition has some nice properties. The complete truth T has the maximal truthlikeness among all theories. If A and B are true, and A is logically stronger than B (i.e., A logically entails B, but not vice versa), then A is more truthlike than B. If A is false, then its truth content $A \cap T$ is more truthlike than A itself. However, Miller (1974) and Tichý (1974) proved that this definition does not work in the intended way, since it cannot be used for comparing false theories: If A is more truthlike than B in Popper's sense, then A must be true. The problem with Popper's approach turns out to be the fact that, by increasing the truth content of a theory, one at the same time increases its falsity content.

A model-theoretic modification of Popper's approach has been proposed by Miller and by Kuipers (1987). Let $Mod(A)$ be the class of models of A, i.e., the L-structures in which all the sentences of A are true. Then define A to be at least as truthlike as B if and only if $Mod(A)\Delta Mod(T)$ is a subset of $Mod(B)\Delta Mod(T)$. This is equivalent to Popper's original definition with the following modification: Popper's requirement for truth content is preserved, but in the requirement for falsity content the class F of false sentences is replaced by the logically weakest false theory Φ in L. If the truth is finitely axiomatizable by a sentence τ in L, so that $T = \mathrm{Cn}(\tau)$, the set of sentences that are logically entailed by τ, then Φ can be defined as $\mathrm{Cn}(\neg\tau)$, i.e., Φ is the theory axiomatized by $\neg\tau$.

If T is a complete theory, then the model-theoretic definition has an implausible consequence: Among false theories, if theory A is logically stronger than B, then A is also more truthlike than B. It is thus vulnerable to the "child's-play objection":

It allows a theory to be improved simply by joining new falsities to it.

Kuipers applies this definition to what he calls "nomic truthlikeness": A theory A asserts the physical possibility of the structures in $Mod(A)$, so that in this context T is usually not a complete theory. However, he admits that the "naive definition" is simplified in the sense that it treats all mistaken applications of a theory as equally bad. For this reason, in his recent work Kuipers has developed a "refined definition," which, by using a qualitative treatment of similarity, allows that some mistakes are better than others (Kuipers 2000).

The Similarity Approach

Popper's approach to truthlikeness employs the concepts of truth and logical consequence, but not the notion of likeness or similarity. Hilpinen (1976), instead, represented theories by classes of possible worlds and assumed, as a primitive notion, the concept of similarity between possible worlds. If possible worlds are replaced by maximally informative descriptions of states of affairs in a given language L, and such descriptions are called *constituents* in L, then the basic problem of the similarity approach is to introduce a distance between the constituents of L (Niiniluoto 1987). For numerical statements (including singular sentences and quantitative laws), such a distance can be defined by means of the underlying metric. In a simple propositional language with three atomic sentences p, q, r, the constituents are conjunctions of the form $\binom{+}{\neg}p \wedge \binom{+}{\neg}q \wedge \binom{+}{\neg}r$, where the atomic sentences occur as negated or unnegated, and the distance between two constituents is the number of diverging claims about the atomic sentences. Jaakko Hintikka (1973) has shown that the notion of constituent can be generalized to full first-order logic. The technicalities in defining the distance between first-order constituents is discussed in Oddie (1986) and Niiniluoto (1987).

The next step is to extend the distance measure to arbitrary theories. Each theory H in a first-order language can be expressed as a disjunction of constituents. Then the distance of a theory H from the truth depends on the distances from the truth of those states allowed by H. Let C^* be the complete truth τ, i.e., the true constituent of L, and let theory H be the disjunction of the states $C_1, C_2, \ldots, C_n$. Let d_{i^*} be the distance of C_i from C^*. Then, in the approach of Tichý and Oddie, the distance of H from C^* is defined by the average function $\sum d_{i^*}/n$ (Oddie 1986). This definition does not satisfy Popper's requirement that, among true theories, truthlikeness covaries with logical strength. In Niiniluoto's approach, 'truthlikeness' is defined by the weighted average of the minimum distance min d_{i^*} and the (normalized) sum $\sum d_{i^*}$ of all distances ($i = 1, \ldots, n$) (Niiniluoto 1987; Kuipers, 1987). Here the minimum distance alone serves to define the notion of *approximate truth*, while the additional sum-factor defines penalties for all the mistakes allowed by the theory. Complications of this min-sum-definition, arising from cases involving continuous quantities, are discussed by Kieseppä (1996).

When the similarity approach is applied to cases where the true constituent is a modal statement with a nomic necessity operator, it gives an explication of what L. J. Cohen (1980) calls *legisimilitude*. This gives also the possibility of interpreting Kuipers' account of nomic truthlikeness within the similarity approach.

The min-sum-definition of degrees of truthlikeness satisfies Popper's basic conditions, but it is also stronger than Popper's approach in the sense that all rival theories (in language L) are comparable with respect to their verisimilitude. The Miller-Tichý refutation and the child's play objection are avoided. Moreover, by this definition it is possible that some false theories are so close to the truth that they are more truthlike than weak true theories.

Zwart (2001) notes that the followers of Popper's explication give "content definitions" in the sense that the least truthlike of all theories is the negation $\neg\tau$ of complete truth τ (i.e., the disjunction of all false constituents), while the "similarity definitions" imply that the worst theory is the "complete falsity" (i.e., the constituent at the largest distance from the truth). If $\tau = p \wedge q \wedge r$, then these two alternatives are represented by $\neg p \vee \neg q \vee \neg r$ and $\neg p \wedge \neg q \wedge \neg r$.

The most famous objection against the similarity account of truthlikeness is Miller's (1974) argument about *language-dependence*. Miller shows that, by a suitable translation between two languages, truthlikeness ordering may be reversed. This argument is related to the more general point that metric properties need not be preserved in one-to-one mappings between quantitative spaces. It has provoked a lot of discussion about the invariance properties of truthlikeness measures (Niiniluoto 1998; Zwart 2001).

Applications

The logical definitions of 'verisimilitude' attempt to tell how close alternative theories are to a given target. If the complete truth is unknown, as is always the case with scientific research problems, then a measure of truthlikeness does not yet indicate how one could know that one is close to the truth. Still, such measures of truthlikeness may be useful as tools in debates concerning scientific realism (Niiniluoto 1999; Kuipers 2000). In particular, they prove that it is meaningful to speak about *scientific progress* toward the truth (see Scientific Progress), and they can be used to define the notion of *convergence* to the truth for scientific theories. Truthlikeness serves to clarify the principle of charity that can be used to defend *reference* invariance in scientific change (see Scientific Change). These notions can be applied to sequences of theories from the history of science. Measures of verisimilitude allow the study of whether truthlikeness explains the success of science. They also help to make sense of the idea that *idealized theories*, which are known to be false, may still be close to the truth. The technical notion of similarity between statements provides a general framework for the study of *approximation* in science (see Approximation).

Epistemological and Methodological Approaches

To solve the epistemological problem of truthlikeness, Niiniluoto (1987) has proposed that the distance $Tr(H,C^*)$ of a hypothetical theory H from the unknown truth C^* can be estimated by the expected value of the degree of verisimilitude of a theory. For this purpose, it has to be presupposed that there is a probability measure P such that $P(C_i \mid e)$ is the epistemic probability (rational degree of belief) of constituent C_i given the available evidence e. Then the *expected verisimilitude* $ver(H \mid e)$ of theory H given evidence e is defined by the sum $\sum_i p(C_i \mid e)Tr(H,C_i)$, where i ranges over all constituents and $Tr(H,C_i)$ is what the degree of truthlikeness of H would be if C_i were the true constituent. If the evidence e entails that C_j is the true constituent, then the expected verisimilitude of H on e equals $Tr(H,C_j)$. Expected verisimilitude differs from the notions of posterior probability, confirmation, and corroboration, since $ver(H \mid e)$ may be high even when theory H is incompatible with evidence e.

The measure $ver(H \mid e)$ makes it possible to integrate the theory of truthlikeness to the Bayesian framework where scientific inference is analyzed as the maximization of expected epistemic utilities (see Niiniluoto 1987; Festa 1993).

A different methodological approach is proposed by Zamora Bonilla (1992), who defines "truthlikeness without truth" by measuring directly the distance of a theory from the available experimental laws.

Ilkka Niiniluoto

References

Bonilla, J. P. Zamora (1992), "Truthlikeness without Truth: A Methodological Approach," *Synthese* 93: 343–372.

Cohen, L. Jonathan (1980), "What Has Science to Do with Truth?" *Synthese* 45: 489–510.

Festa, Roberto (1993), *Optimum Inductive Methods*. Dordrecht, Holland: Kluwer.

Hilpinen, Risto (1976), "Approximate Truth and Truthlikeness," in Marian Przeecki, K. Szaniawski, and Ryszard Wojcicki (eds.), *Formal Methods in the Methodology of Empirical Sciences*. Dordrecht, Holland: D. Reidel, 19–42.

Hintikka, Jaakko (1973), *Logic, Language-Games and Information*. Oxford: Oxford University Press.

Kieseppä, Ilkka (1996), *Truthlikeness for Multidimensional, Quantitative Cognitive Problems*. Dordrecht, Holland: Kluwer.

Kuipers, Theo (ed.) (1987), *What Is Closer-to-the-Truth?* Amsterdam: Rodopi.

——— (2000), *From Instrumentalism to Constructive Realism*. Dordrecht, Holland: Kluwer.

Miller, David (1974), "Popper's Qualitative Definition of Verisimilitude," *British Journal for the Philosophy of Science* 25: 168–177.

Niiniluoto, Ilkka (1987), *Truthlikeness*. Dordrecht, Holland: D. Reidel.

——— (1998), "Verisimilitude: The Third Period," *British Journal for the Philosophy of Science* 49: 1–29.

——— (1999), *Critical Scientific Realism*. Oxford: Oxford University Press.

Oddie, Graham (1986), *Likeness to Truth*. Dordrecht, Holland: D. Reidel.

Popper, K. R. (1963), *Conjectures and Refutations*. London: Routledge and Kegan Paul.

——— (1972), *Objective Knowledge: An Evolutionary Approach*. Oxford: Oxford University Press.

Tichý, Pavel (1974), "On Popper's Definition of Verisimilitude," *British Journal for the Philosophy of Science* 25: 155–160.

Zwart, Sjoerd (2001), *Refined Verisimilitude*. Dordrecht, Holland: Kluwer.

See also **Bayesianism; Confirmation Theory; Popper, Karl Raimund; Realism**

VIENNA CIRCLE

The Vienna Circle consisted of a group of about three dozen researchers drawn from the natural and social sciences, logic, and mathematics, which met regularly in Vienna between the two world wars to discuss philosophy. The work of this group constitutes one of the most important and influential philosophical contributions of the twentieth century, in particular in the development of analytic philosophy and history and philosophy of science (Stadler 2001 and 2003a).

The Vienna Circle was first publicly announced in 1929 with the publication of what came to be called its manifesto, *Wissenschaftliche Weltauffassung. Der Wiener Kreis* [The Scientific Conception of the World: The Vienna Circle], edited by the Verein Ernest Mach (the Ernest Mach Society) and authored by Carnap, Hahn, and Neurath (1929). The Vienna Circle was essentially a modernist movement, at the center of which was the "Schlick Circle," a discussion group organized in 1924 by Moritz Schlick. Rudolf Carnap, Herbert Feigl, Philipp Frank, Kurt Gödel, Hans Hahn, Otto Neurath, Felix Kaufmann, Viktor Kraft, Karl Menger, Friedrich Waismann, and Edgar Zilsel belonged to its inner circle (see Carnap, Rudolf; Hahn, Hans; Neurath, Otto; Schlick, Moritz). Their meetings were also attended by Olga Taussky-Todd, Olga Hahn-Neurath, Rose Rand, Gustav Bergmann, and Richard von Mises, and on several occasions by visitors such as Hans Reichenbach, Alfred J. Ayer, Ernest Nagel, Quine Willard Van, and Alfred Tarski. Members of the periphery, most of them as participants, were Egon Brunswik, Karl Bühler, Josef Frank, Else Frenkel-Brunswik, Heinrich Gomperz, Carl Gustav Hempel, Eino Kaila, Hans Kelsen, Charles Morris, Arne Naess, Karl Popper, Frank P. Ramsey, Kurt Reidemeister, and the alleged "genius," Ludwig Wittgenstein, who had a special influence on some members of the group (see Ayer, Alfred Jules; Hempel, Carl Gustav; Nagel, Ernest; Popper, Karl Raimund; Quine, Willard Van; Ramsey, Frank Plumpton; Reichenbach, Hans). In addition, the mathematician Karl Menger organized in the years 1926–1936 an international Mathematical Colloquium, which was attended by Kurt Gödel, John von Neumann, and Alfred Tarski among many others (Menger 1994).

This international and interdisciplinary discussion circle was pluralistic and committed to the ideals of the Enlightenment. It was unified by the aim of making philosophy scientific with the help of modern logic on the basis of experimental and everyday experience. The general aims of the movement were expressed in its publications such as *Schriften zur Wissenschaftlichen Weltauffassung* [Publications on the Scientific Conception of the World], 1929–1937, in eleven volumes; *Einheitswissenschaft* [Unified Science], 1933–1938, in seven volumes; the journal *Erkenntnis,* 1930–1940 (the 1939 volume was called *Journal for Unified Science*); and the *International Encyclopedia of Unified Science*, 1938–1970 (Neurath, Carnap, and Morris 1971).

Given this story of scholarly success, the fate of the Vienna Circle was tragic. The Verein Ernest Mach was suspended in 1934 by the Austrian Nazis; Schlick was murdered in 1936; and, around this time, many members of the Circle were forced to leave Austria for racial and political reasons. Thus, soon after Schlick's death, the Circle disintegrated. As a result of the emigration of so many of its members and adherents, however, the Circle's ideas became more and more widely known, especially in Scandinavia, Britain, and North America, where they contributed hugely to the emergence of modern philosophy of science (Timms and Hughes 2003; Hardcastle and Richardson 2003). In Germany and Austria, however, the break that was caused by the forced emigration of the Vienna Circle's members was felt on the philosophical and mathematical scene for a long time (Heidelberger and Stadler 2003).

"Logical Positivism" and/or "Logical Empiricism"

The name 'Vienna Circle' was used for the first time in 1929 in the manifesto mentioned earlier. It was suggested by Neurath and was supposed to have a pleasant connotation similar to *Vienna Woods* or *Viennese Waltz.* At the same time, the term was to indicate the origin of this philosophical movement and its collective orientation (Frank 1949). In the programmatic essay of 1929 the position of the antimetaphysical "radical" left wing around Carnap, Neurath, Hahn, Frank, and others was especially prominent. This group supported the idea of a physicalist unity of science, most commonly referred to as "logical empiricism" as found later in the program of the *International Encyclopedia of Unified Science* (Neurath 1946). By contrast, the more moderate wing of the Vienna Circle, around Schlick, Waismann, Feigl, and others, emphasized their adherence to a dualism of science and philosophy with different names like "consistent empiricism" (Schlick 1950, 462f) or 'logical positivism' (Kraft 1950).

The widely used term "logical positivism" actually originated in Blumberg and Feigl's (1931) article of that title. Here both authors gave a concise description of the new anti-Kantian synthesis of logical and empirical factors proclaiming the impossibility of synthetic a priori truths. They went on to describe the philosophical transformation from the old to the new positivism with the adoption of symbolic logic, epistemology, and research into the foundations of science. Finally, they explained, following Wittgenstein, "the purpose of philosophy as the clarification of the meaning of propositions and the elimination of ... meaningless pseudo-propositions" (269).

Scientific Philosophy and Philosophy of Science

Proponents of "scientific philosophy" thought of philosophy not as an autonomous discipline existing prior to science but as a critical discipline dependent upon the results of the natural and social sciences, logic, and mathematics. Turning around Kant's dictum, they claimed that philosophy without science is empty, and science without philosophy is blind. Adoption of this scientific conception of philosophy does not, however, determine what epistemology, methodology, and ontology one is committed to. Nonetheless, all adherents of scientific philosophy demanded exact methods, a critical attitude, and a more or less empirical orientation. They opposed irrational and theological systems of philosophy (*Systemphilosophie*) and viewed science in general in a positive way.

Historically, Mach's philosophy provided the foundation for the development of the positions adopted within the Vienna Circle (see Mach, Ernest). The term 'logic of science' (Carnap's [1934] *Wissenschaftslogik*), known since the mid-1930s as "philosophy of science," was later used to describe these positions (see Carnap, Rudolf). This implied a general scientific conception of philosophy as well as an attempt to provide a philosophy for all sciences (including human sciences). In addition, within the Vienna Circle, philosophy was regarded both as a form of linguistic analysis and as a discipline drawing on the foundations of the natural and social sciences.

At the same time, there were divergences of philosophical approaches within the Vienna Circle. Those such as Schlick defended a methodological dualism of philosophy and science, and those such as Neurath sought to integrate philosophy altogether within a scientific conception of the world (see Neurath, Otto; Schlick, Moritz). In Schlick's view, the classical philosophical positions of empiricism and rationalism were integrated with the help of modern logic and mathematics, but a distinction between philosophy and science still remained. Neurath's more radical physicalism, or "encyclopedism," of logical empiricism aimed at overcoming philosophy itself within his collective project of an *International Encyclopedia of the Unity of Science* (Neurath 1946). This divergence in philosophical approaches left room for debates within the Circle on such topics as the merits of phenomenalist and physicalist languages, coherence and correspondence theories of truth, logical syntax and semantics, verification and confirmation, and ideal and natural languages. At the same time, there was a certain consensus on the merits of logical analysis of language, a fallibilist epistemology, a scientific attitude to the world, and the unity of scientific explanation and knowledge in general.

The rivalry between Schlick's "consistent empiricism" and Neurath's physicalist unified science is a complex matter. Certain views were shared by both, such as the view of philosophy as a critique of language in accordance with Wittgenstein's philosophy of the *Tractatus* of 1922. However, while the principle of verification (see Verifiability), logical atomism, and the picture theory of language are constitutive features of the entire movement, by themselves they do not characterize the Vienna Circle. Theoretical elements like logicism, verifiability, methodological phenomenalism and physicalism, a fallibilist theory of knowledge, conventionalism, and realism, together with an empiricist encyclopedism, were cornerstones of the internal pluralistic development of logical empiricism from the 1930s onward (see Logical Empiricism). This development also reflected the influence of Neurath's pragmatic point of view within the Circle. In particular, the objection toward any dualism of "language" and "world" (as *Wirklichkeitsphilosophie*), with the attendant denial of any absolute "foundation of knowledge" (Schlick 1934) is representative of this nonreductive naturalism and methodological holism in the spirit of Pierre Duhem's and Henri

Poincaré's philosophy of science (see Conventionalism; Duhem Thesis; Poincaré, Henri). This form of relativism and naturalism already anticipated the pragmatic and historical turn after World War II in the philosophy of science, which contributed to overcoming the linguistic turn and the so-called received view of philosophy of science.

The rejection of synthetic a priori judgments remained an important element of the logical empiricism of the Vienna Circle. According to Russell and Whitehead in the *Principia Mathematica* (see Russell, Bertrand), symbolic logic and mathematics were regarded as purely analytical and a priori (independent of any experience). Analytic truths of these kinds were contrasted with contingent statements of the natural sciences and ordinary everyday experience, as synthetic *a posteriori* judgments (see Analyticity). But there was no further class of synthetic a priori judgments; instead there was thought to be an important class of "meaningless" sentences, without any cognitive content. The elements of this class were seen as "metaphysical" in the sense that they are not part of knowledge at all, even though they may express some realm of commonsense experience (see Cognitive Significance).

This position of the classical Vienna Circle is most prominently represented by Carnap's (1931) "Elimination of Metaphysics through Logical Analysis of Language," which developed a program for a unified rational reconstruction of science (see Rational Reconstruction). But the question as to whether an empirical basis could serve as the foundation for all knowledge received strongly divergent answers from coherence theorists about truth (influenced by Neurath) and correspondence theorists (influenced by Schlick) (Hempel 1993). Also, the apparently strict distinction between analytic and synthetic sentences was questioned (Menger 1979, 1–60). The ideal of one language of science, logic, and mathematics was radically weakened within the Vienna Circle itself with Menger's and Carnap's principle of tolerance long before Quine (1953) put forward his critique of the "Two Dogmas of Empiricism" (see Quine, Willard Van). Thus, contrary to popular belief, a heterogeneous pluralism of views was in fact characteristic of the Vienna Circle—for example, regarding ethics (Schlick, Menger, Kraft), the alternatives of realism versus positivism (Schlick, Carnap, Feigl, Kraft, Kaufmann) and verificationism versus falsificationism (both positions criticized by Neurath), and last, but not least, matters of ideological and political preference, such as conservative liberalism versus leftist socialism. In the later period of the Vienna Circle, the contested verification principle was gradually abandoned and replaced by some form of a probabilistic confirmation methodology based on the principle of "connectibility" (von Mises 1951) (see Cognitive Significance; Verifiability).

Scientific Conception of the World and Scientific Humanism

The unity of science movement, with its six International Congresses for the Unity of Science, held from 1935 to 1941, and the ambitious publication project of the *International Encyclopedia of Unified Science,* 1938–1970, had a broader cultural meaning and goal, most notably the attempt to improve the human condition and to promote social reform and the intellectual struggle against irrationalism and the totalitarian worldview (see Unity and Disunity of Science; Unity of Science Movement). It was a manifestation of a late-Enlightenment conception of science with a socially inspired antimetaphysics. Between the two world wars, metaphysics was seen as a correlative feature of German idealism as well as of Austrian fascist "universalism," as represented by the economist Othmar Spann.

The practical impulse behind this therapeutic destruction of metaphysical systems, then, was the desire for a scientific attitude based on human experience, directed against the *Zeitgeist* of totalitarian universalism and cultural pessimism (as criticized by Neurath [1921] and [1931]). Therefore, traditional philosophy, first of all, had to be reduced to a critical analysis of language, because most proponents of logical empiricism thought that an exact and sober usage of the scientific language is a precondition for all problem-oriented philosophizing—and moreover a sort of moral obligation.

Social criticism and collective work in philosophy of science formed a programmatic unity striving for a sweeping improvement of the human condition. Whereas in the natural sciences considerable progress had already been made, the situation in the social and cultural sciences was not so transparent, and was influenced by the ongoing *Methodenstreit* since the turn of the century (Kaufmann [1936] 1999). Although some members of the Vienna Circle, like Kaufmann, Neurath, and Zilsel, contributed essentially to this neglected field, their contributions have been largely ignored in the historiography on the Circle for a long time. In this respect it is worth mentioning that after the disintegration of the Vienna Circle, its former members still occasionally made reference to the "scientific conception of the world" when speaking about general ideological questions. For example, Carnap spoke about "scientific humanism" as a view shared by the majority of the logical empiricists (Carnap 1963, 81ff). After the dissolution of the Vienna Circle, the forced migration of most of its members, and the dispersion of the logical empiricist movement from its centers in Central Europe, the twin aims of transforming philosophy and establishing a philosophy of science could be envisaged only once the ties to their previous cultural context and audience had been severed. But even in these difficult times the proponents of the exiled Vienna Circle organized six well-attended, prestigious International Congresses for the Unity of Science: Paris (1935 and 1937), Copenhagen (1936), Cambridge, UK (1938), Cambridge, Massachusetts (1939), and Chicago (1941). One can thus say that the demise of the Vienna Circle in the German-speaking world was accompanied by the transformation of Viennese *Wissenschaftslogik* into philosophy of science in the Anglo-Saxon scientific community.

Recent Reassessments

The new historiography on logical empiricism started with the rediscovery of Ernest Mach (1838–1916) as a precursor of Gestalt theory, evolutionary epistemology (possibly radical), constructivism, and the modern historically oriented philosophy of science. Already in Mach's reception in the Vienna Circle one can see not only a certain pluralism of views but also a polarization of the various positions (Mach's influence on Carnap's [1967] *Aufbau/Logical Structure*, the critical distancing to "psychologism" in the manifesto, the alternative to the principle of economy, etc.).

Even prior to World War I, the predecessor of the Vienna Circle—the "First Vienna Circle"—had begun to take shape both as an organization and as a philosophy (Uebel 2000). Within a discussion circle (*inter alia*, with Frank, Hahn, and Neurath) at a coffeehouse, traditional "academic philosophy" grew more scientific. This so-called First Vienna Circle met regularly as of 1907 to discuss the synthesis of empiricism and symbolic logic as modeled after Mach, Boltzmann, and the French conventionalists (Pierre Duhem and Henri Poincaré) (see Conventionalism; Duhem Thesis; Poincaré, Henri). This was also seen as an indirect answer to Lenin's polemical remarks against Mach in his book, *Materialism and Empirio-Criticism*, which remained very influential in eastern Europe from its publication in 1909 to the "Velvet Revolution" of 1989–1990.

This early phase in the development of logical empiricism can also be interpreted as an anti-Cartesian turn in epistemology and philosophy of science, which undermined both the synthetic a priori and the secure foundations of knowledge. In the middle of the permanent crisis of philosophy between reform and revolution in society and science, the further development of this "scientific philosophy" had, in any case, been initiated.

With the conflict-laden appointment of the physicist and philosopher Moritz Schlick (1882–1936) to Mach's chair for natural philosophy of the "inductive sciences" in Vienna in 1922, the heyday of scientific philosophizing in the post–World War I period was prolonged. Even though Schlick ([1918, 1925] 1974) felt committed to an epistemological realism in his main work, *General Theory of Knowledge*, he began his inaugural lecture with a programmatic allusion to Mach, under the sway of the Viennese tradition up to Wittgenstein, that almost all philosophy is natural philosophy.

In the phase during which the Schlick Circle became a veritable institution, there was already a pluralism of positions that had emerged in the field of tension between Wittgenstein's *Tractatus* and Carnap's *Logischer Aufbau der Welt* [Logical Construction of the World] (Carnap 1967). Yet notwithstanding all the discrepancies between Carnap's "rational reconstruction" and the philosophy of ideal language (Wittgenstein), all those involved came to identify with a philosophical reform movement as opposed to academic philosophy.

This radical program, in turn, left an indelible mark on avant-garde art (i.e., constructivism, associated with Gerd Arntz, the artist of Neurath's pictorial language) (see Neurath, Otto) and literature, as well as architecture (*Werkbund* and Bauhaus), centering around Ludwig Wittgenstein, Paul Engelmann, Adolf Loos, Josef Frank, and Neurath's efforts within the Congrés International d'Architecture Moderne (Nemeth and Stadler 1996). Clarity and precision as ends in themselves and features of scientific philosophy bridged both Wittgenstein's cultural pessimism and the socioculturally enlightened impetus of the modernist Vienna Circle.

With this convergence of various elements of philosophy of science, theoretical innovation was accelerated in the phase in which the Vienna Circle made public appearances and expanded its international contacts. The latter development was accompanied by the disintegration and uprooting of logical empiricism in the German-speaking world. In this sense, the phenomenon of the Vienna Circle is a prototypical case study of intellectual emigration (Stadler 2003b, c).

To all appearances, there seem to be two diametrically opposed trends. While the international influence of the Vienna Circle was steadily growing, the group had been systematically marginalized in Austria and Germany starting in the early 1930s. The murder of Schlick and the disgraceful, for the most part anti-Semitic, reactions to it brutally ushered in the process that can be described as the "demise of scientific reason" (Stadler and Weibel 1995). This took place in parallel with the general trend at universities, which at the time were increasingly coming under the influence of a growing anti-democratic and racist discourse dominated by clerical-fascist and National Socialist forces. This development led to the anschluss of Austria by the Third Reich, which culminated in systematic dismissals, banishment, and annihilation of many leading intellectuals and others (Stadler 2002).

Concluding Remarks

Looking at the current definitions of the Vienna Circle, one can quickly recognize the difficulty of providing a representative description of the Circle and of logical empiricism in its entirety. Even the autobiographical accounts of Vienna Circle members since the classical period of the Schlick Circle show a remarkable variance—depending on focus and underlying motivations.

What these texts have in common is the refutation of metaphysics as well as of philosophy as a discipline in its own right. As an alternative, one finds a tendency toward a (physicalist) unified science that later culminated in an empiricist encyclopedia project and includes the principle of tolerance as applied to logic and scientific languages. Here, the contours of epistemological options emerge. If one also takes into account that the manifesto represents only one variant of the Vienna Circle at the end of the 1920s, then it becomes amply clear that there existed only a limited consensus.

In addition, it is obvious that neither the autobiographical accounts of contemporaries nor the historical accounts originating shortly after 1945 were able to provide an adequate picture of the Vienna Circle. Moreover, there exists only a partial, albeit broad, overlap of the concept of the Vienna Circle with that of logical empiricism in general when one takes into account the related movements of the Berlin Circle around Hans Reichenbach and the Warsaw Group around Alfred Tarski (Danneberg, Kamlah, and Schäfer 1994).

Is it still possible to find a sort of basic agreement here to unite the members of the Vienna Circle, both the central figures and those on the periphery? First of all, the Vienna Circle method is a way of philosophizing based on linguistic analysis and a great amount of problem-oriented, open-ended discussion. This was experienced personally by Arne Naess, who focused several times on the Circle's "thought style," which, in (not only) his opinion, leads to an inherent "pluralism of tenable worldviews" (Naess 2003). Second, the use of an unambiguous language, together with exact methods, is certainly a main legacy of the Circle and those associated with it. It is only when this exact formal approach is adopted that the content and positions can be constructively criticized and refuted—a characteristic that most current modern and postmodern philosophies lack.

The explicit and hidden history of the Vienna Circle from *Wissenschaftslogik* to the recent philosophy of science documents the wide range, pluralism, and diversity of the Viennese heritage and message. Be it called "scientific philosophy" (as initiated by Schlick), "scientific humanism" (according to Carnap), or a "republic of scholars" (following Neurath), it is a guide to an intellectual journey that continues through the present day and probably on into the future.

FRIEDRICH K. STADLER

The author thanks Camilla Nielson for this translation.

References

Blumberg, Albert, and Herbert Feigl (1931), "Logical Positivism: A New Movement in European Philosophy," in *Journal of Philosophy* 28: 281–296.

Carnap, Rudolf, Hans Hahn, and Otto Neurath (1929), *Wissenschaftliche Weltauffassung. Der Wiener Kreis*. Vienna: Artur Wolf Verlag. Abridged English translation "The Scientific Conception of the World: The Vienna Circle," in O. Neurath (1973), *Empiricism and Sociology*. Edited by Marie Neurath and Robert S. Cohen. Dordrecht and Boston: Reidel, 299–318.

Carnap, Rudolf (1931) "Überwindung der Metaphysik durch logische Analyse der Sprache," *Erkenntnis* 2: 219–241.

——— (1934), *Logische Syntax der Sprache*. Vienna: Springer. English translation: *The Logical Syntax of Language*. London: Routledge, 1937.

——— (1963), "Intellectual Autobiography," in Paul A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*. LaSalle, IL: Open Court, 3–84.

——— (1967), *The Logical Structure of the World*. Berkeley and Los Angeles: University of California Press.

Danneberg, Lutz, Andreas Kamlah, and Lothar Schäfer (eds.) (1994), *Reichenbach und die Berliner Gruppe*. Braunschweig-Wiesbaden: Vieweg.

Frank, Philipp (1949), *Modern Science and its Philosophy*. Cambridge: Cambridge University Press.

Hardcastle, Gary, and Alan W. Richardson (2003) (eds.), *Logical Empiricism in North America*. Minneapolis: University of Minnesota Press.

Heidelberger, Michael, and Friedrich Stadler (eds.) (2003), *Wissenschaftsphilosophie und Politik/Philosophy of Science and Politics*. Vienna and New York: Springer.

Hempel, Carl G. (1993), "Empiricism in the Vienna Circle and in the Berlin Society for Scientific Philosophy: Recollections and Reflections," in Friedrich Stadler (ed.), *Scientific Philosophy: Origins and Developments*. Dordrecht, Netherlands: Kluwer, 1–10.

Kaufmann, Felix ([1936] 1999), *Methodenlehre der Sozialwissenschaften*. Vienna and New York: Springer.

Kraft, Viktor (1950), *Der Wiener Kreis. Der Ursprung des Neopositivismus*. Vienna and New York: Springer.

Menger, Karl (1979), *Selected Papers in Logic and Foundations, Didactics, Economics*. Dordrecht, Netherlands: Reidel.

——— (1994), *Reminiscences of the Vienna Circle and the Mathematical Colloquium*. Edited by Louise Galland, Brian McGuinness, and Abe Sklar. Dordrecht, Netherlands: Kluwer.

Naess, Arne (2003), "Pluralism of Tenable Worldviews," in F. Stadler (ed.), *The Vienna Circle and Logical Empiricism: Reevaluation and Future Perspectives*. Dordrecht, Netherlands: Kluwer.

Nemeth, Elisabeth, and Friedrich Stadler (eds.) (1996), *Encyclopedia and Utopia: The Life and Work of Otto Neurath (1882–1945)*. Dordrecht, Netherlands: Kluwer.

Neurath, Otto (1921), *Anti-Spengler*. München: Callwey. English translation in M. Neurath and R. Cohen (eds.) (1973), *Empiricism and Sociology*. Dordrecht, Netherlands: Reidel, 158–213.

——— (1931), *Empirische Soziologie. Der wissenschaftliche Gehalt der Geschichte und Nationalökonomie*. Vienna: Springer. English translation in M. Neurath and R. Cohen (eds.) (1973), *Empiricism and Sociology*. Dordrecht, Netherlands: Reidel, 319–421.

——— (1946) "The Orchestration of the Sciences by the Encyclopedia of Logical Empiricism," in *Philosophy and Phenomenological Research* VI/4: 496–508.

Neurath, Otto, Rudolf Carnap, and Charles Morris (eds.) (1971), *Foundations of the Unity of Science: Toward an International Encyclopedia of Unified Science* (Vol. I, nos. 1–10; Vol. II, nos. 1–9). Chicago and London: Chicago University Press.

Quine, Willard Van Orman (1953), *From a Logical Point of View: Nine Logico-Philosophical Essays*. Cambridge, MA: Harvard University Press.

Schlick, Moritz ([1918, 1925] 1974), *General Theory of Knowledge*. Translated by Albert E. Blumberg. Vienna and New York: Springer 1974.

——— (1934), "Über das Fundament der Erkenntnis," in *Erkenntnis* 4: 79–99.

——— (1950), "Moritz Schlick," in *Philosophen-Lexikon*. Edited by Werner Ziegenfuss and Gertrud Jung. Berlin: de Gruyter.

Stadler, Friedrich (2001), *The Vienna Circle. Studies in the Origins, Development, and Influence of Logical Empiricism*. With the first publication of the protocols (1930–31) of the Vienna Circle and an interview with Sir Karl Popper (1991). Vienna and New York: Springer.

——— (2002), "The Emigration and Exile of Austrian Intellectuals," in Rolf Steininger, Günter Bischof, and Michael Gehler (eds.), *Austria in the Twentieth Century*. New Brunswick, NJ, and London: Transaction Publishers, 116–136.

——— (ed.) (2003a), *The Vienna Circle and Logical Empiricism: Reevaluation and Future Perspectives*. Dordrecht, Netherlands: Kluwer.

——— (2003b), "The 'Wiener Kreis' in Great Britain: Emigration and Interaction in the Philosophy of Science," in E. Timms and J. Hughes (eds.) *Intellectual Migration and Cultural Transformation: Refugees from National Socialism in the English-Speaking World*. Vienna and New York: Springer, 155–180.

——— (2003c), "Transfer and Transformation of Logical Empiricism: Quantitative and Qualitative Aspects," in G. Hardcastle and A. Richardson (eds.) *Logical Empiricism in North America*. Minneapolis: University of Minnesota Press, 216–233.

Stadler, Friedrich, and Peter Weibel (eds.) (1995), *Vertreibung der Vernunft* [The Cultural Exodus from Austria]. Vienna and New York: Springer.

Timms, Edward, and Jon Hughes (eds.) (2003), *Intellectual Migration and Cultural Transformation: Refugees from National Socialism in the English-Speaking World*. Vienna and New York: Springer.

Uebel, Thomas E. (2000), *Vernunftkritik und Wissenschaft. Otto Neurath und der Erste Wiener Kreis*. Vienna and New York: Springer.

von Mises, Richard (1951), *Positivism: A Study in Human Understanding*. New York: Dover.

See also **Analyticity; Carnap, Rudolf; Conventionalism; Demarcation, Problem of; Empiricism; Hahn, Hans; Hempel, Carl Gustav; Logical Empiricism; Mach, Ernest; Neurath, Otto; Phenomenalism; Popper, Karl Raimund; Rational Reconstruction; Quine, Willard Van; Schlick, Moritz; Unity and Disunity of Science; Unity of Science Movement; Verifiability**

VISUAL REPRESENTATION

Graphs, diagrams, drawings, sonographs, and x-rays are commonly used in contemporary science in the process of research, in communicating results, and in education. In contrast to more familiar images—paintings, snapshots, children's drawings—visual representations in science often, though not always, depict phenomena that cannot be seen: structures too small to see with visible light (electron micrographs), relations among properties (graphs), steps in a mechanism (diagrams). Philosophers of science have studied linguistic and mathematical representations in order to understand science, but they have only recently begun to investigate what visual representations contribute, and how they do so. Visual representations appear to play philosophically significant roles in scientific reasoning. They are prevalent in journal articles, where scientists express and defend hypotheses. These papers are relatively formal communications, subject to disciplinary standards of clarity and objectivity, and reviewers scrutinize figures as well as text in evaluating the arguments presented. Some figures appear to express scientific hypotheses (e.g., the diagram of the double helix), while others are presented to provide support for a hypothesis. Can a visual representation express a hypothesis? What kind of inferences can a visual representation support? How does the visual format of a figure relate to its role? Scientists frequently express concern about the accuracy of figures, just as they do about the accuracy of linguistically or mathematically expressed claims. Can pictures be accurate—or true? In general, what advantages do visual representations offer over linguistic and mathematical representations?

In order to answer these questions, it is essential to understand the nature of visual representations as such, and to know in what ways they differ from other types of representations. An analysis that can apply to all symbols used in science is required to avoid dependence on assumptions about the kinds of representations involved. Thus, the first step in understanding visual representations in science is clarification of key representational features. Different kinds of visual representations can then be compared with each other, and with linguistic representations. This preliminary step is also required in order to understand the epistemic roles played by figures—fundamentally, to explain how visual representations express and support scientific claims.

Visual Symbols

Visual representations, like written or spoken sentences and numerical formulas, are external objects that function as symbols. No object conveys content on its own; symbols must be interpreted. Goodman (1976) shows that resemblance is neither necessary nor sufficient for representation: Just about any thing can be designated to refer to some other thing—Wellington does not represent his portrait even though they resemble one another. Comprehension of symbols, including pictures, involves interpretation, which is conventional in the sense that the relation between symbol form and content is not determined just by either intrinsic features of the symbol or a resemblance relation between the form of the symbol and its referent; something extrinsic to both objects (symbol and referent) is necessary to determine that one refers to another. This is obvious for linguistic representations, whose form does not bear a visible relation to its referent. Comprehension of pictures often feels so natural and automatic that viewers are not conscious of it, but application of the appropriate interpretive conventions is necessary to comprehend visual representations. For example, in order to understand a watercolor of a landscape, one has to apply the appropriate interpretive conventions to the colors and shapes in order to comprehend it as a representation of a landscape. These are *different* conventions from those used to interpret a different kind of picture, such as a black-and-white photograph. A gray tone in the watercolor must be interpreted differently from a gray tone in the photo. All forms of representation used in a scientific article share this most general feature of symbols: They require application of the appropriate interpretive conventions in order to know what the symbol refers to. The interpretive conventions govern relations between symbol and referent for particular systems of symbols, so that although it is tempting to try to analyze pictures by focusing on individual specimens, the full understanding (as with linguistic representations) requires knowing the systems of which they are components.

The content of visual representations is not *entirely* conventional in the above sense; relations between the visible form of visual symbols and their referents are also involved in determining the content of pictures. There is a fundamental difference between textual and visual symbol systems. Visual systems are based on a spatial format: Visual representations are symbols in which some spatial relations are interpreted to mean something about the referent. In some visual symbol systems, spatial features of the symbol refer to spatial features. For example, spatial relations among circles in a ball-and-stick diagram of a molecule refer to spatial relations among the atoms in the molecule. In other visual systems, however, spatial features of symbols refer to nonspatial features of the object represented. A spatial feature of a time line—length—represents a temporal feature: duration. Graphs are visual symbols in which spatial features represent relations among properties, like the relation between gas volume and pressure. Other visible features such as color may also contribute to the meaning of visual representations. However, the referential role of spatial relations is the fundamental feature of visual representations.

For this reason, the visible *forms* of visual representations are related to their referents. This relation varies among visual symbol systems, but each system is characterized by a relation between symbol form and content that holds for that system, and this relation is the basis for interpreting, and thus comprehending, the symbol as such. This relation holds between the *interpreted* visible features of the symbol, on the one hand, and the properties and relations represented, on the other (not between all features of the symbol and all those of its subject).

In contrast to the spatial format of visual representations, linguistic representations have a sequential format; the sequence of letters and spaces alone is sufficient to determine the meaning of text. Serial symbol systems typically comprise symbols whose spatial form is arbitrary with respect to their meaning. Though some serial systems, such as written Chinese, include pictographic characters, this is not necessary for serial representation, and the relationship between symbol form and referent in the pictographic characters serves a mnemonic purpose, but is not systematic (not all Chinese characters are pictographic). Chinese is a serial system because the meanings of statements are determined by the ordering of characters, rather than by spatial relations among them. For serial systems like alphabetical and numerical symbol systems, the shapes of letters are entirely unrelated to the referents of the words and sentences they compose. Relative spatial position of serial characters can contribute to meaning—poets might vary between-word spacing for emotional effects—but it is not necessary; the sequence of characters is sufficient to determine symbol meaning.

Types of Visual Symbols

This difference in format does not account for important differences *among* visual representations, so it cannot explain why scientists use different kinds of visual representations. Further analysis of the properties of visual symbol systems allows for categorizing two different kinds of visual representation and explaining what makes them differ from each other, as well as identifying an important feature that one of these visual types, but not the other, shares with serial representations. Diagram systems, such as those of electrical circuitry and chemical structures, *look* different from visual representations that appear much more like pictures, such as electron micrographs and satellite photos. This is because the former symbol systems share some features with linguistic representations—in addition to having the spatial format shared by all visual representations.

Some symbol systems consist only of markings that can each be identified as instances of a particular character—unless they are simply illegible (Goodman [1976] calls these syntactically articulate systems). Text, numerical formulas, and wiring diagrams have this type of syntax. It allows for compositionality: All these systems consist of unambiguously recognizable atomic characters (e.g., numerals, letters, Chinese characters) combined in rule-governed ways. In a visual symbol system with this kind of syntax, some spatial relations among the atomic characters will also be interpreted to refer to some relation among the referents of the atomic character. The meaning of a molecular diagram is a function of the reference of the atomic characters and how they are arranged (see Figure 1). These are compositional symbol systems, like written languages and numerical systems, all of whose symbols consist of atomic characters that are always identifiable as particular characters. Their meanings are a function of the identity and composition of atomic characters, whether the composition is sequential (for serially formatted systems) or

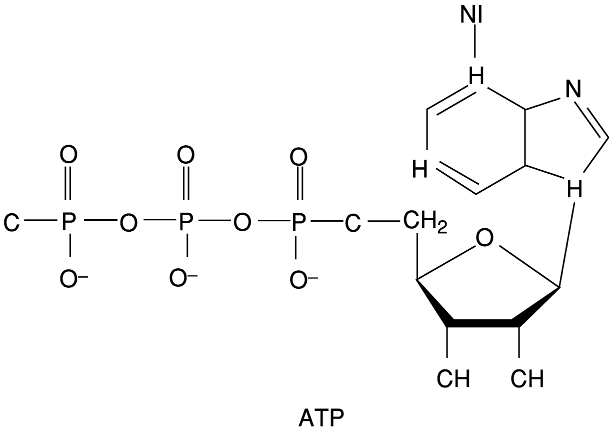


Fig. 1. Diagram of adenosine triphosphate (ATP). In this diagram, lines and letters serve as atomic characters that refer to bonds and different kinds of atoms, respectively. Spatial relations among those atomic characters are also interpreted: for example, contiguity of a P, a single line, followed by an O, to refer to the bond between a phosphorous and an oxygen atom. In this way, the atomic characters and spatial relations among them are used to represent the structure of the molecule.

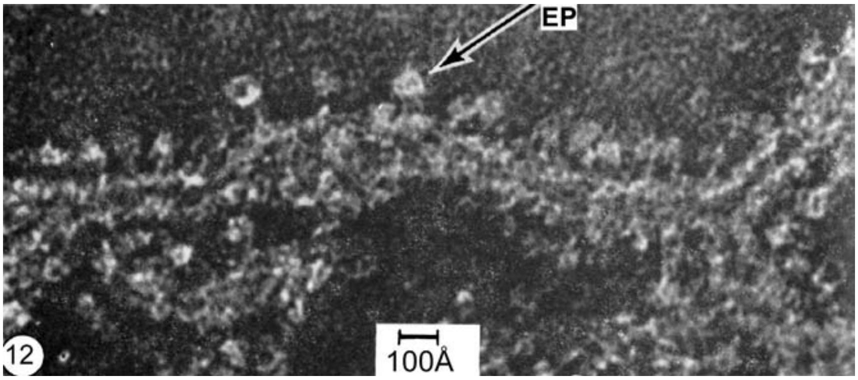


Fig. 2. Electron micrograph of inner mitochondrial membranes. The image is produced by a process in which a very thin sample is prepared with a stain that does not mix well with biological material, and also repels electrons. A beam of electrons is aimed at the sample, and electrons go through the areas where biological material is present but are deflected by the stain. This process produces a pattern of light areas on the image whose shape matches the way the electrons passed through the sample. From Fernández-Morán 1962.

spatial. Because of this compositionality, which is in part due to the fact that visible forms of atomic characters are arbitrary with respect to their meaning, diagrams have some of the convenience and flexibility of textual representations (whose spatial features are entirely arbitrary with respect to meaning). These are formats with which it is easy to express very abstract or general ideas. For this reason they are useful for representing mechanisms.

Photographs, natural history drawings, and electron micrographs have some important syntactic and semantic differences (Figure 2 is an example of this type of figure). These images are not composed of discrete atomic elements. They are characters from systems in which any difference—for *at least one* interpreted spatial feature of the figure, such as the shape of a curve on a graph—corresponds to a difference in the character that figure instantiates. And every different character has a different referent. So two electron micrographs with different two-dimensional arrays of light-to-dark scaling represent different structural features. Because the smallest spatial difference correlates with a different referent, visual representations from these systems can represent very complex properties (like the particular shape of a subcellular structure), and these systems comprise symbols that together represent a dense range of these properties.

Another prominent difference among visual representations is their degree of abstraction. It is not possible to explicate this difference in terms of discrete categories, but studies of the relation between abstractness of symbol form and symbol content in scientific contexts have presented some interesting results. Michael Lynch (1988) discusses figures in which an electron micrograph and a schematic drawing are paired together, showing that figures with different degrees of abstraction can play different epistemic roles. In these pairs, the micrographs serve as evidence for the more abstract visual representation. Lynch describes the visible differences between the two and relates those differences in symbol form to differences in the content each conveys, in order to support his claim that the inferential move from less to more abstract representation is not a matter of mere simplification of form in terms of a reduction of visual stimuli. The schematic drawings are idealizations, generalizations, or extrapolations from the pictures that support them. There is a difference in the kind of fact depicted by each of the two representations in the pair. The picture refers to a particular case, while the schematic drawing expresses a claim that is general (applying to multiple cases), in addition to being more abstract (leaving out features represented in the picture).

The use of figures that vary in abstraction also raises questions about the relationship between abstractness of representation and accuracy. Hall (1996) argues that the degree of naturalism of a pictorial style is distinct from its capacity for accurate representation. He supports his thesis with contrasts of realistic pictorial styles that misrepresent human anatomy with very schematic diagrams that convey accurate information. Given the role of interpretation in comprehension of visual representations, this phenomenon can now be explained. Accurate visual representations are interpreted to represent a state of affairs that obtains. There is no warrant for assuming that visual representations are meant to represent *all* properties of their subjects; the properties represented, as well as the visible features of the symbol used to convey them, vary among visual symbol systems. Very schematic figures, then, are not inaccurate just because they do not represent in detail the features of their referents. Thus, understanding figures as elements of visual symbol systems that vary according to system conventions clarifies the relation between representational styles and semantic value.

Scientific Reasoning and Visual Representations

Model-Based Reasoning

Most contemporary philosophers of science who have discussed the use of figures focus on the content of the figures, rather than presenting an analysis of how that content is conveyed—what the visual format contributes. Nersessian (1992) and Giere (1996) discuss scientific visual representations, and, while these papers also focus on content, their model-based accounts of scientific reasoning offer a promising route to understanding how figures might be involved in that reasoning. (see Scientific Models). Nersessian (1992) discusses the development of the theory of electromagnetism, arguing for an important role for analogical reasoning, which involves a systematic mapping of relations between a source and a target domain. Nersessian supports her claim with a discussion of Maxwell's figure, which she describes as "a *visual* representation of an *analogical* model" (Nersessian 1999, italics in original.). Nersessian suggests that analogical thinking involves mental models, which embody the relations holding among the events and entities involved in the object of thought (see Scientific Metaphors). The usefulness of a visual representation to this kind of analogical reasoning can be explained by the distinctive feature of visual representations: their spatial formatting. A visual representation of a model might be particularly efficient for mental modeling, since the content is conveyed through spatial and other visible relations.

Comprehending the representation requires mapping the perceived visible features to the appropriate relations and properties of the referent. An independent step of generating a mental image of the model is not required, as it would be if the model were conveyed through a serial representation. Consider the difference between a linguistic description of an analog clock face and a picture of one. The statement that the big hand is on the eight and the small hand is just below the ten provides information from which it is possible to infer the time. Most people would comprehend the description, then create a mental image of the clock, and from that image determine the time. Comprehending a picture of a clock seems to eliminate a step by combining the comprehension of the representation with the construction of the mental image.

Giere (1996) focuses on external, rather than mental, models: Scientific theories are abstract objects that need not have linguistic character. Judgments about models are made by comparing them with real objects, because a model serves as a prototype to assess similarity to putative instances of the model. Giere considers diagrams embodiments of completely abstract models; he claims that, as such, they can serve as prototypes for similarity judgments. The ubiquitous nature of resemblance relations makes it difficult to account for the *particular* similarities that are relevant for a particular judgment. This is another area in which further study of scientific visual representations as such could have an interesting payoff: understanding visual representations as isomorphic symbols provides some resources to explain how diagrams make certain features salient. The isomorphic relation stands between some, but not all, of the visible features of the figure and its content, as determined by the interpretive conventions for that symbol system. So the capacity to understand a figure as a representation explains the capacity to identify the features by which similarity should be assessed.

Evidential Roles

Visual representations often appear to be presented as support for scientific claims. Brown (1997) describes several diagrams from mathematics and argues that they function as proofs. Folina (1999) responds by pointing out that these diagrams are not proofs in the sense of deductive demonstrations, though she concedes that the diagrams may contribute some other form of support. Some diagram *systems* have been shown to support proofs. Hammer (1995) presents analyses of several logical diagram systems (Venn, Euler, and Peirce), including soundness and completeness proofs. However, these demonstrations do not shed light on how an individual visual diagram could function as a proof.

Giere (1996) presents examples of figures from research in geology, in which the patterns of two different types of data are placed in visible correspondence with one another, in accordance with the model in question. However, while in this case the *form* of the figures is recognized as important to confirmation, the relation between form and content that distinguishes visual representations is not discussed.

Wimsatt (1991) provides examples of figures from various disciplines that facilitate inferences, including cases in which the use of multiple kinds of figures are involved. Wimsatt does not give an explicit account of how figures support inferences, but the paper emphasizes the usefulness of visual representations in handling data that vary along more than one dimension. The analysis of visual representations can clarify how figures do this: For making inferences from a single figure, the two-dimensional format is an advantage, because the dimensions can be used to express relations that hold among the data. For example, by plotting experimental results on a graph, in which the position of each data point is determined by its value on each of the axes, relations among the data can be determined very efficiently. Thus graphs provide a way to represent the data themselves, by using spatial location along the axes to refer to values. In addition, graphs facilitate inferences of *higher-order* properties, because the spatial relations *among* the plotted points are meaningful: They are interpreted according to the conventions governing the interpretation of location with respect to the two axes. So spatial relations among the plotted points support inferences about relations among the properties represented by individual data points.

In addition to being spatially formatted, and thus representing relations among parts of their referents, some figures provide evidence for a hypothesis because the figure is causally related to what it represents. An electron micrograph, for example, is made by beaming electrons through a very thin biological sample stained with material that does not mix with biological material and that deflects electrons. The electrons can go through the slide only in areas without stain, so the array of detected electrons reflects the array of unstained area on the slide, thereby producing the light areas of the image. The process of producing the micrograph correlates the form of the image (light versus dark areas) with the shape of the sample. The micrograph thus represents the structure of the sample. Note that this does not guarantee that the sample has the structure the micrograph represents: The procedure might produce such an image, due to problems in staining or machine malfunction. Accuracy is not guaranteed.

But if researchers think their procedure is reliable, they will take the micrograph as an accurate representation of the sample structure. One could draw the same structure, but the fact that the form of this visual representation is *causally related* to the sample that the figure represents is an important source of the epistemic warrant of the micrograph. Pictorial representations generated by imaging techniques gain this status because of the causal relation between symbol and object. For this reason pictorial representations can play a role very similar to evidence claims: Their epistemic warrant derives primarily from causal interactions between the object studied and a detector, rather than by inference from other representations. This is similar to perceptually grounded linguistic observation statements. The veracity and the evidential relevance of such figures depend on the procedure by which they are produced. In order for a figure to serve as evidence, scientists need some assurance that the technique that produced the image generates accurate representations of the subject matter. This assurance can come in the form of understanding appropriate causal connections between the technique and the representations it generates, or through experience of the reliability of a method. The critical difference between a figure like a micrograph and a linguistically expressed observation claim is that the *form* of the figure is causally related to its content.

The micrograph also exemplifies another way in which the *kind* of visual symbol involved makes an important contribution to science. Imaging techniques produce visual symbols with distinctive, “pictorial” representational features described above. The characters are not composed of unambiguously identifiable atomic characters, and these systems have the capacity to represent a dense set of referents, due to the precise correlation between symbol form and referent.

This system-level feature grounds another important advantage of visual representations: Many imaging techniques provide an experimental method that results in a comprehensible representation of a state of affairs, even if the precise vocabulary needed to describe that state of affairs through linguistic representations does not yet exist. Their content is a function of the visible form of the symbol and does not depend on a vocabulary of arbitrarily associated symbols and meanings. So a person who knows how to interpret such a symbol can do so even if it represents a novel phenomenon. For example, a person who knows how to interpret micrographs like Figure 2, in which the light areas represent biological material, can comprehend a micrograph from the same system even if it has a shape the viewer has never seen or heard of before. That means that the person can comprehend the micrograph, as a representation of the form of the sample structure, even if it represents a structure that was completely unknown prior to the production of the micrograph. This is quite different from alphabetical systems; because the connection between symbol form and referent is arbitrary, one cannot comprehend an unfamiliar word simply by looking at it. Pictorial symbol systems offer an important advantage in science, allowing newly discovered and extremely complex phenomena to be represented and fully comprehended. Once a representation of the phenomenon is in hand, it is possible to assign a linguistic name to it, but having a *hypothesis* about this structure is not a prerequisite. Imaging techniques provide tools to comprehensibly represent phenomena that do not yet have a linguistic label.

Finally, the capacity of visual representation to support scientific knowledge requires that such representation be usable objectively. Like linguistic representations, pictures can be used for both rhetorical and nonrhetorical purposes. The capacity of figures to represent phenomena gives them the potential to contribute objective support for a scientific hypothesis. Identifying *how* to use visual representations objectively is not a trivial matter, and scientists have expressed active concern in this area. A historical study by Daston and Galison (1992) shows that choices about what images to publish can be influenced by scientists' beliefs about appropriate ways to avoid subjectivity in scientific communication.

Necessity

Are visual representations necessary for contemporary science? In some cases, human cognitive limitations make the use of visual representations necessary for *comprehension* of the content the author wishes to convey (Perini 2001). For example, diagrams of macromolecular structures are external representations whose contents can be expressed through serially formatted symbols: The coordinates for atomic locations can be printed out as a list as well as used to make a diagram. But, while humans readily understand the diagrammatic representation and can understand the individual items on the serially formatted list of atomic coordinates, they cannot use that list to generate a mental representation of the structure of the molecule. Comprehension of the structure requires a representation of the spatial relations among the parts of the molecule, and the computational task involved in calculating those relations from the list of individual atomic coordinates is simply too complicated.

A different set of circumstances that would make visual representations necessary is if they conveyed content that was both essential to science and not expressible with serial representations such as linguistic or numerical symbols. There are two different reasons why the content of visual representations might fail to be expressible with serial representations. First, pictorial content would not be linguistically expressible if pictures conveyed a different *kind* of content from linguistic representations. However, the analysis in terms of symbol system features provides no reason to think that there is such a difference. The difference between linguistic and visual representations concerns features of the symbol system (of its characters on the one hand and of the referent set on the other), and not the nature of the content conveyed. Second, pictorial content would not be linguistically expressible if the *amount* of content conveyed by the visual representation could not be symbolized with serial representations. If an infinite set of representations were the *only* possible linguistic translation of a figure, actual expression with physical symbols would be impossible. Kitcher and Varzi (2002) argue that a map of Manhattan is worth a nondenumerable set of linguistic representations. They reach this conclusion after claiming, without support, that the real map is not the physical symbol, but an abstract object. This abstract object amounts to a shape consisting of infinitely many contiguous line segments, each of which can be linguistically described. The question of necessity does not turn on whether the map is really the physical symbol or an abstract object, so this aspect of their paper can be taken as given. The paper does not show that the map is necessary due to amount of content, however, because Kitcher and Varzi's analysis does not provide any reason to think that the map could not also be expressed by a finite serial symbol. There is no reason why one complicated linear expression describing that shape would not serve equally well as a translation.

The question of translatability tends to draw philosophical attention because of the disciplinary focus on linguistic representations, but it is important to bear in mind that the philosophical significance of visual representations in science does not depend on whether or not they are necessary to convey the content they do. Diagrams and tables both can be easily translated into serially formatted linguistic or mathematical representations, yet they are the preferred expression of the data. This suggests that the visual format itself contributes to scientific reasoning. Perini (2004) offers a preliminary analysis of the role of two-dimensional formatting in the presentation of evidence, and concludes that tables—consisting entirely of numerical symbols, but formatted in two dimensions—serve as evidence in part because that way of presenting the data facilitates inferences of higher-order relations among the data. Furthermore, while the same conclusion can be drawn from the data set presented serially, the reasoning involved would be different. Thus those wishing to understand the reasoning actually presented in a scientific paper employing a table would need to consider the data in their two-dimensional presentation. The previous section showed that imaging techniques like electron microscopy have distinctive capacities for evidential support because of the pictorial nature of the representations they produce. Even if figures are fully translatable, they still make philosophically significant contributions to science *as visual representations*—and must be understood as such in order to understand the reasoning contemporary scientists actually use.

LAURA PERINI

References

Brown, James R. (1997), "Proofs and Pictures," *British Journal for the Philosophy of Science* 48: 161–180.

Daston, Lorraine, and Peter Galison (1992), "The Image of Objectivity," *Representations* 40: 81–128.

Fernández-Morán, Humberto (1962), "Cell-Membrane Ultrastructure: Low-Temperature Electron Microscopy and X-Ray Diffraction Studies of Lipoprotein Components in Lamellar Systems," *Circulation* 26: 1039–1065

Folina, Janet (1999), "Pictures, Proofs, and 'Mathematical Practice': Reply to James Robert Brown," *British Journal of Philosophy of Science* 50: 425–429.

Giere, Ronald (1996), "Visual Models and Scientific Judgment," in Brian Baigrie (ed.), *Picturing Knowledge: Historical and Philosophical Problems Concerning the Use of Art in Science*. Toronto: University of Toronto Press.

Goodman, Nelson (1976), *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis, IN: Hackett Publishing.

Hall, Bert (1996), "The Didactic and the Elegant: Some Thoughts on Scientific and Technological Illustrations in the Middle Ages and Renaissance," in Brian Baigrie (ed.), *Picturing Knowledge: Historical and Philosophical Problems Concerning the Use of Art in Science*. Toronto: University of Toronto Press.

Hammer, Eric (1995), *Logic and Visual Information*. Stanford, CA: CSLI Publications.

Kitcher, Philip, and Achille Varzi (2000), "Some Pictures are Worth $2^{\aleph_0}$ Sentences," *Philosophy* 75: 377–381.

Lynch, Michael (1988), "The Externalized Retina: Selection and Mathematization in the Visual Documentation of Objects in the Life Sciences," *Human Studies* 11: 201–234.

Nersessian, Nancy (1992), "How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science," in Ronald Giere (ed.), *Cognitive Models of Science: Minnesota Studies in the Philosophy of Science*, vol. 15. Minneapolis: University of Minnesota Press.

——— (1999), "Model-Based Reasoning in Conceptual Change," in Lorenzo Magnani, Nancy Nersessian, and Paul Thagard (eds.), *Model-Based Reasoning in Scientific Discovery*. New York: Academic/Plenum Publishers.

Perini, Laura (2001), "Explanation in Two Dimensions: Diagrams and Biological Explanation," presentation at the biannual meeting of the International for the History, Philosophy and Social Science of Biology, Hamden, CT, July 18–22.

——— (2004), "Visual Representations and Their Role in Confirmation," presentation at the annual meeting of the Philosophy of Science Association, Austin, TX, November 18–20.

Wimsatt, William (1991), "Taming the Dimensions—Visualizations in Science" in Arthur Fine, Mickey Forbes, and Linda Wessels (eds.), *PSA 1990*, vol. 2. Chicago: University of Chicago Press, 111–135.

See also **Confirmation Theory; Scientific Models**

INDEX

A

Abbot, Edwin, 829
Abduction, 1–3
  inference to the best explanation (IBE), 241
  realism, 689, 691–692, 694
  scientific change, 729
Abductive learning, generative grammar, 112–113
Abele, L.G., 222
Aboutness, intentionality as, 405–406
Abrahamsen, Adele, 153, 155
Absolute confirmation, 144–145
Absolute simutaneity, 787
Absolute space, 787
Absolute time, 205
Abstract calculus, Nagel on, 493
Abstract dynamical system, determinism, 203
Abstract interlevel reduction, 697
Abstraction
  scientific models, 741
  visual representation, 866
Abstract space, reductionism, 697–698
Abu Abdullah al Battanti, 34
Acceleration
  classical mechanics, 116, 117
  space-time, 786
Accelerators, particle, 538
Acceptability condition, cognitive significance, 139–140
Access consciousness, 159–160, 161, 162, 618
Accidental generalizations/regularities
  causality, 95, 99
  explanation, 278
  laws of nature versus, 439–443
Accuracy, model, 742
Achinstein, Peter
  scientific domains, 734
  theory structure, 827
    semantic view, 825
    syntactic (received) view, 823
Achronal hypersurface, relativistic space-time, 205
Ackermann, W., 675
Acquaintance, principle of, 851
Acquired immunity, 363
Acquisition of sense data, Hanson and, 345
Action
  experimental, 270–271
  Searle, 770
Action at a distance
  causality, 94
  locality, 455–456
  Newton's laws, 118
Action potential, mechanism of, 471–472
Action principles, classical mechanics, 120
Actual (versus potential) explanation, deductive-nomological (D-N) model, 276
Actualist ontology, experimental science and, 271
Actual models, theory structure, 826
Adams, Ernest, 679, 680
Adaptation and adaptationism, 3–7
  altruism, 8–9
  evolutionary biology, philosophy of, 70
  evolutionary psychology, 265, 266
  natural selection, 497–498
  species, 800
Adaptation and Natural Selection (Williams), 65
Adaptive economy of thought, 844
Adaptive evolution, 255, 264
Adaptive landscape model, 582
Addelson, Kathryn Pine, 296
Additive effects of alleles, 355
Additive genetic variance, 354, 355
Additivity, emergence, 231, 232
Ad hoc science, 192
Adler, Robert, 368
Advancement of Science, The (Kitcher), 784
Affirmations, protocol sentences, 612
Against Method (Feyerabend), 305, 306
Agamben, Giorgio, 364–365
Agency
  cognizers, 619
  experimental method, 274
  function, 315
Agent-centered cognition, 619
Aggregativity, reductionism, 698, 702
AGM (Alchourrón, Gärdenfors, and Makinson) model, 249–250
Agnosia, visual, 546
Agren, G., 313
Aim and Structure of Physical Theory (Duhem), 208
Air, G.M., 483
Ajdukiewicz, Kazimierz, 169
Akaike, H., 536–537
Akaike information criterion (AIC), 537
al Battanti, Abu Abdullah, 34
Albert, David, 204, 205, 596
  quantum mechanics, 656, 657
Alchemical theory, 102
Alchemy, 268
Alchourrón, C.E., AGM model, 249–250
Alchourrón, Gärdenfors, and Makinson (AGM) model, 249–250
Alcoff, Linda Martin, 297, 299, 301

INDEX

Alexander, R., 264
Alexander, Scott, 231, 232
Alexander, Stephen, 35
Algebraic calculus, 119
Algebraic quantum field theory (AQFT), 631–632
Algebras
  Boolean, *see* Boolean algebras
  Lindenbaum, 634, 635
  von Neumann type II, 506
Algorithm
  artificial intelligence, 28, 31
  computation, 29
  conservation biology, 165
  Marr's computer model of mind, 547
  scientific change, 731
  Turing and, 833
Alhazen, 545
Alkindi, 545
Allais, Maurice, 186, 188
Allais paradox, 186, 187
Allele frequency, population genetics, 583
Alleles, 330, 332, 355
  population genetics, 579, 583
  reductionism, 698
Allelic fitness, 311
Allen, C., 315
All Life is Problem Solving (Popper), 572
Allopatry, 253
Allostery model, molecular biology, 483
Almagest (Ptolemy), 33–34
Almeder, Robert F., 784
Alston, W., 245
Alternative (non-null) hypothesis, 804–805
Alternative RNA splicing, 482
Altruism, 8–11
  evolutionary, 8–9, 256
  fitness, 311
  natural selection, 498
  prisoner's dilemma, 10–11
  psychological, 9–10
  social sciences, 784
Ambler, E., 540
Amensalism, ecological communities, 217
American pragmatism, 726
American Structuralism (linguistics), 107
Amino acid sequencing, 485
Ampère, André Marie, 270
Ampliative inferences, 379, 380–381
Amsterdamski, Stefan, 193
Amundson, Ron, 72, 316, 320
Analog computation, 28
Analogical effects
  inductive logic, 390
  probability theory, 604
  scientific metaphors, 737
Analogical models, 742
Analog models, 827
Analysans/analysandum, explication, 287
Analysis
  chemistry, philosophy of, 103
  explication versus, 287, 288
  functional, 317, 318
  instrumentalism, 402
  modes of, Nagel on, 494
Analysis of invariance, 278
Analysis of Mind, The (Russell), 716
Analysis of Sensations (Mach), 467
Analysis of variance, 804
Analytical connection rules, Hanson, 344
Analytical Mechanics (Lagrange), 119
Analytical mechanics, defined, 116
Analytical sentences, 461
Analytic a priori propositions, Ayer and, 39
Analytic geometry, 357
Analyticity, 11–21
  Carnap and, 80, 82, 84–85, 86, 88
  cognitive significance, 132
  conventionalism, 175
  demarcation problem, 191
  empiricism, 238
  explication, 293
  Harman's synthesis of case against, 15–19
  historical background, 11–13
  logical empiricism, 461–463
  logical empiricism versus logical positivism, 458
  Popper and, 574–575
  Quine and, 659, 660–666
    two dogmas of empiricism, 13–14
    web of belief, 19–20
  realism, 688
  responses to Quine, Carnap, Grice, Strawson, and others, 14–15
  Vienna Circle and, 82, 860
Analytic philosophy, Russell and, 715, 717, 720
Analytic sentences, 15–17
Analytic statements, Duhem thesis, 209
Analytic/synthetic distinction
  Carnap, 86
  Poincaré and, 570
  Putnam and, 620
  Quine and, 662, 663
Andersen, Hanne, 762
Anderson, Carl, 539
Anderson, D., 536
Anderson, Elizabeth, 302
  decision theory, 185
  economics, 227
  feminist philosophy, 297
Anderson, James A.
  artificial intelligence, 27
  connectionism, 151, 153, 157
Anderson, J.R., 30–31
Anderson, Philip, 846
Angell, James, 61
Angermeier, P.L., 164
Anglo-Saxon philosophy of experiment, 269
Angstrom, A.J., 35
Angular momentum
  chemistry, philosophy of, 104
  conservation of, 118
  quantum conditions, 652
Anomalies
  causal, Reichenbach and, 710
  research programs, 714
Anomalistics, 194–195
Anscombe, G.E.M., 98
Anthropic principle, 21–23, 556
Anthropological mathematics, 673

Anthropology
  cognitive, 614
  cognitive science, multidisciplinary approach, 123
  Darwinian, 264
  feminist philosophy, 298
Anti-adaptationism, natural selection, 497–498
Anti-atomists, kinetic theory, 417
Anti-Cartesianism, Putnam, 620, 621, 625
Anti-de Sitter space-time, 206
Anti-foundationalism
  Lakatos and, 437
  Neurath and, 513
Antifoundationalism, epistemological, 83
Antigen-antibody reactions, 366
Anti-Kantianism, Vienna Circle, 859
Antimaterialism, Mach, 467
Antimetaphysics, *see also* Metaphysics
  Carnap and, 133
  logical empiricism, 459, 462, 463
  Neurath, 511
  Quine and, 664
Antipsychologism, 717
Anti-realism, 91; *see also* Instrumentalism; Realism; Social constructionism/constructivism
  Kuhn's scientific revolutions, 426
  Poincaré and, 570–571
  realism, 687, 689, 691–692
  scientific models, 740, 747
Antireductionism, *see also* reductionism
  explanation, 285
  molecular biology, 484
Anti-scientism, Putnam and, 620
Aplysia neurons, 516
Apoptosis, 338
Applications, domain of intended, 826
Applied artificial intelligence, 29
Appraised posits, 708
Approximate truth, 856; *see also* verisimilitude
Approximation, 24–27
  reductionism, 698
  unity and disunity of science, 845
  verisimilitude, 857
A priori
  chemistry, philosophy of, 103
  conventionalism, 169, 174
  empiricism, 235, 237, 243
  Euclidean space-time, 117
  geometry, 360
  Mach and, 468
  Putnam and, 620, 621
  Quine and, 662
  Reichenbach and, 705, 709
  scientific change, 730
A priori laws, unity and disunity of science, 844
A priori probability functions, 391
Arabidopsis thaliana, 486, 488
Arab science, 268
Arbitration scheme, game theory, 328
Archea, 338
Archimedes, 116, 272, 357
Arguments
  from the bad lot, 241
  explanation, 284, 285
  from illusion, 547
  inductive logic, 385
  realism, 689
  scientific style, 766
Ariew, Andre, 3
  biology, philosophy of, 69, 72
  innate/acquired distinction, 399
Aristotelian systems
  Kuhn's scientific revolutions, 421, 423, 428
  scientific models, 742
  scientific progress, 750
  scientific revolutions, 756
Aristotle, 307, 545, 780, 843
  astronomy, 33, 34
  demarcation problem, 189, 190
  explanation, 285
  heritability, 352
  inductive logic, 384
  Kuhn and, 421–422
  Kuhn on language, 428
  locality/local action, 451
  mechanics, 116
Arithmetic
  epistemology of, 569–570
  Peano, 357, 358
  Poincaré and, 569–570
  recursive, 360
  theories of, 823–824
Arithmetization of syntax, 84
Armendt, B.
  Dutch Book argument, 212, 213
  Ramsey, 677
Armitage, P., 808
Armstrong, David M., 90, 97
  Dretske-Tooley-Armstrong account of laws, 98, 440, 441–442
  intentionality, 409
  laws of nature, 440, 441, 443
  probability, 607
Armstrong, D., consciousness, 160
Arnauld, Antoine, 599, 605
Arnold, Vladimir I., 119, 122
Arntz, Gerd, 861
Arntzenius, Frank, 830
Arrow-Debreu-McKenzie model of pure exchange, 508
Artificial intelligence, 27–32
  and computation, 27–29
  connectionism, 150–158
  origins, 27
  philosophical issues, 29–31
    engineering versus science, 29–30
    hard versus soft, 30
    strong versus weak, 30
    weak versus strong equivalence, 30–31
  programs, models, theories, and levels, 31–32
  psychology, cognitive architecture, 615–616
  psychology, philosophy of, 614, 615, 619
  scientific change, 731
  Searle's critique of strong AI, 771–773
  Turing and, 834, 836
  von Neumann and, 503, 508–509
Artificial languages, Quine and, 664
Artificial life, 508
Artificial neural networks (ANN), 150–151, 155
Asexual reproduction, 798, 800

Aspray, William, 508
Association, Hume's empiricism, 523
Associationist account of laws, 95
Assortative mating, population genetics, 579
Assumptions
  Duhem thesis, 209
  Kuhn's disciplinary matrix, elements of, 421
  Kuhn's scientific revolutions, 426
  metaphysics and experimental approaches, 429
  parsimony, 531–538
Astrobiology, 37–38
Astrochemistry, 37, 38
Astronomia nova (Kepler), 34
Astronomy, philosophy of, 32–38
  astrobiology, 37–38
  dark matter, 38
  Kuhn's scientific revolutions, 422–424
  naked eye observations, 33–34
  physical sciences, philosophy of, 557
  radio, UV, X-ray, and gamma ray instruments, 36–38
  spectrometer, 35–36
  telescope, 34–35
Asymmetric probability distribution, 602
Asymmetry, of time, 831–832
Asymptotic rules of induction, 382
Atlan, Henri, 367
Atlas, celestial, 35
Atom and Cosmos (Reichenbach), 704
Atomes, Les (Perrin), 417
Atomic objects, complementarity, 140–143
Atomic orbitals
  approximation, 25
  chemistry, philosophy of, 103
Atomic phenomena
  observation, 525–526, 527
  particle physics, 538–544
  visual representation, 866, 868
  von Neumann and, 505
Atomic propositions, Ramsey on, 672
Atomic theory, quantum mechanics, 651–652
Atomism
  classical mechanics, 121
  conservation of energy, 119
  corporeal, 364
  kinetic theory, 415, 417–418
  Kuhn's scientific revolutions, 426
  logical, 722, 728
  methodological individualism, 478
  Millikan's oil drop experiment, 429
  protocol sentences, 610
  Vienna Circle, 859
Atomistic view of mental phenomena, 770
Atoms
  kinetic theory, 416
  models of, 740
Attention
  consciousness as, 160
  degrees of, 771
Attractors, dynamical systems theory, 122, 154
Attitudes toward risk, decision theory, 187–188
Atwood machine, 272
Auditory perception, 550
Aufbau (Carnap), *see Logische Aufbau der Welt, Der* (Carnap)
Aufbau principle, chemistry, 104
Aumann, R., 327, 507
Austin, J.L., 40, 768, 780
Auto-associative memories, recursive (RAAMs), 154
Automata
  cognitive science, 130
  von Neumann and, 508–509
Automated reasoning, 731
Automatic Computing Engine (ACE), Turing and, 835
Autopsychological beliefs, protocol sentences, 611
Autopsychological objects, 80, 81
  Carnap, 523
  protocol sentences, 610
Auuyang, Sunny F., 632
Auxiliary hypotheses, 209, 714, 839, 841
Averages, fitness, 312
Avery, Oswald T., 335
Avogadro's number, 417, 418
Awodey, S., 290
Axelrod, Robert M., 227
Axiomatic-deductive method, 268
Axiomatic method
  Hilbert, 357–358
  multiplicity of geometries, 360–361
  scientific style, 766
Axiomatic preference theory, Ramsey and, 677–678
Axiomatics/axioms
  conventionalism, 169
  explication, 288, 289–290, 293
Axiomatic set theory, 570
Axiomatization
  classical mechanics, 115
  explication, 289
  fitness, evolutionary theory, 314
  Hempel, 351
  inductive logic, 386
  instrumentalism, 403
  probability, 599–601
  Ramsey and, 680
  Reichenbach and, 705–706
  Russell and, 718
  set theory, 504–505
  theory structure, syntactic (received) view, 824
  unity and disunity of science, 844
  von Neumann and, 504–505, 509
Axiomatization of the Theory of Relativity, The (Reichenbach), 704, 705, 706
Axiom of choice
  Hahn and, 342, 343
  Ramsey on, 672
Axiom of infinity, 672, 674
Axiom of reducibility, 719
Axioms
  Ramsey, theory structure, 680
  theory structure, syntactic (received) view, 823
Axiom-to-theorem flow, Lakatos and, 438
Aydede, M., 618
Ayer, Alfred J., 38–40
  analyticity, 12, 13
  cognitive significance, 132, 133, 134, 135
  conventionalism, 169, 175
  empiricism, 238–239
  induction, problem of, 380
  logical positivism, 62, 459–460
  Quine and, 659
  verifiability, 852
  Vienna Circle, 82, 858

B

Bacciagaluppi, Guido, 648
Bachelard, Gaston, 757, 758
  chemistry, philosophy of, 101
  experimental method, 269, 270–271
Background information/knowledge
  confirmation theory, Bayesian, 148
  inductive logic, epistemic interpretations of probability, 386–387
  observation, 526–527, 528
Backpropagating rule, defined, 152
Backpropagation-trained networks, 156
Backtrackers, causality, 99
Bacon, Francis, 117
  crucial experiment (*experimentum crucis*) idea, 209
  demarcation problem, 189–190
  experimental method, 268, 269
  inductive logic, 384
  propositions and their testing, 524
  scientific style, 766
  unity and disunity of science, 843
Baconian method, scientific revolutions, 759
BACON system, 731
Bader, R., 103, 104
Bagheri-Chaichian, H., 355
Bailer-Jones, Daniela, 746
Baird, Davis
  chemistry, philosophy of, 102
  experimental method, 271
Balance of nature, 73
  ecology, 217–220
Baldwin, James Mark, 257
Ballot, Buys, 416
Ballungen (cluster concepts), 513, 611
Balmer series, 651
Balzer, W.
  approximation, 25
  theory structure, 824, 826
Banks, Erik, 468
Barad, Karen, 296, 302
Barbour, Julian, 117
Bare particulars, 374
Barinaga, M., 516
Barker, Peter, 762
Barkow, Jerome H., 264
Barnard, G., 808
Barnes, Barry, 776
Barnes, Eric, 1
Barnett, Leslie, 336
Barrell, B.G., 483
Barrett, Jeffrey
  determinism, 205
  quantum mechanics, 655, 658
  von Neumann, 506
Barrow, J.D., 21
Bartlett, M., 600
Barton, C., 582
Barton, N.H., 584
Base pairs, molecular biology, 486
Basic sentences, *see* Popper, Karl R.; Protocol sentences
Basic System of Inductive Logic, A (Carnap), 87
Bateson, Patrick, 267
Bateson, William, 330
Batson, C. Daniel, 8, 10
Batterman, Robert W.
  determinism, 204
  emergence, 234
  reductionism, 698, 699, 700, 701
  unity and disunity of science, 846
Bauch, Bruno, 79
Bauer, Henry, 194
Bauhaus School, 513
Bayes, Thomas, 41; *see also* Bayesianism
  Bayesian confirmation theory, 148
  inductive logic, 384
Bayesian information criterion (BIC), 535
Bayesianism, 41–60
  abduction, 2
  anthropic principle, 23
  Carnap, Rudolf, 88
  causality, 96–97
  confirmation theory, 148–150, 348–349
  decision theory, 181
  Dutch Book argument, 210–213
  epistemology, 250
  induction, problem of, 45–52, 382
  Kuhn's scientific revolutions, 425
  logic and methodology, 42–43
  minimal belief change, 52–55
  parsimony, 534–535
  physical sciences, philosophy of, 555
  prediction, 589–590, 591, 593–594
  probability, 605–606
  Ramsey and, 675, 677
  representing prior knowledge, 55–59
  scientific change, 730, 731
  statistics, 807, 814
    advances and controversy, 810–811
    inductive behavior philosophy, 805
    p-values and Bayesian posteriors, 812–813
    roles for probability in inference, 803
    significance testing controversy, 809
  testing approach, 43–45
Bayesian networks and causal models, 96–97, 616
Bayesian updating, 593–594
Bayes's rule, 42, 56
Bayes's theorem, 534, 593, 803, 804
  corroboration, 179
  probability theory, 600
  statistics, in meta-methodology, 810
Bayne, T., 160
Beadle, George, 334
Beanbag genetics, 583
Beatty, John
  biology, philosophy of, 69
  evolution, 253
  fitness, 312, 313, 314
  laws of nature, 443
  natural selection, 500, 501
  theory structure, 825
Bechtel, William
  connectionism, 153, 155
  mechanism, 469, 472, 473, 475, 477
  neurobiology, 514, 517
  reductionism, 699
Bedau, M., 321
Beeckman, Isaac, 346

Behavior
  adaptation and adaptationism, 6
  artificial intelligence, 27, 29–31
  decision theory, 184, 185
  evolutionary psychology, 263–267
  feminist philosophy, 298
  innate/acquired distinction, 395, 398
  linguistic, 111
  social sciences
    cultural materialism, 782
    functional analysis, 781–782
    new directions, 783–784
    rational choice program, 780–781
  teleological, 90
  weak versus strong equivalence, 30–31
Behavioral ecology, 263
Behavioral science
  connectionism, 151
  feminist philosophy, 296
  statistics, 808–809
Behaviorism, 61–63
  biological information, 64
  Chomsky on, 114
  demarcation problem, 191
  neurobiology, 521
  perception, problem of consciousness, 549
  psychology, philosophy of, 614
Behman, H., 704
Behrens, W.V., 57
Beisinger, S.R., 167
Bejerano, Gill, 7
Bekenstein, Jacob D., 413
Bekoff, C., 315
Belief
  confirmation theory, Bayesian, 149–150
  Duhem thesis, 208–210
  Dutch Book argument, 210–213
  empiricism, 525
  epistemology, justified true belief (JTB) analysis, 244–245
  feminist philosophy, 299
  function, 319
  intentionality, 408
  Kepler, 117
  protocol sentences, 613
  Quine's web of, 19–20
  Ramsey
    degrees of, 676–677
    degrees of, logic of consistency for, 677–678
    logic of truth, 678–679
    partial, probability and, 675–676
  Searle, 769, 770
  social constructionism, 776
  social sciences, 783
Belief change, minimal, 52–55
Belief-reality relationship, protocol sentences, 612–613
Bell, John S.
  quantum mechanics, 654, 655–656
  time, 831
  von Neumann and, 505, 506
Bellarmino, Roberto, 400
Bell-Kochen-Specter theorem, 656
Bell's theorem
  action at a distance, 455
  locality, 453–454
  nonlocality, 456
  quantum field theory, 632
  quantum mechanics, 454–455
Belnap, Nuel, 197
Belot, Gordon, 203, 204, 555
Beltrametti, E., 633
Benford's law, 58
Bennett, Jonathan
  causality, 99, 100
  Ramsey, 680
Bensaude-Vincent, B., 102
Benson, W.W., 254
Bentley, Richard, 118, 119
Benzer, Seymour, 334
Berger, J., 808
Berger, J.O., 812, 813
Berger, Peter, 774–775
Berget, S., 483
Bergmann, Gustav
  verifiability, 852, 853, 854
  Vienna Circle, 858
Berk, A., 483
Berkeley, George, 39, 513
  empiricism, 236
  instrumentalism, 400
  perception, 545
  verifiability, 851
Berlin, I., 134–135
Berlin Circle
  Hempel, 347
  Reichenbach and, 704, 709
  Schlick and, 727
Bernard, Claude
  experimental method, 269
  immunology, 364–365, 366, 369
Bernardo, Jose, 59
Bernoulli, Daniel, 416
Bernoulli, Jacob, 119, 607
Bernoulli, James, 45–46
  weak law of large numbers, 46, 47
Bernoulli, Johann, 119
Bernoulli, Nicholas, 46
Bernoulli's formula, kinetic theory, 416
Bernoulli systems, determinism, 204
Bernstein, David, 195
Berry, M., 234
Berry, Michael, 635
Bertrand Russell and Trinity (Hardy), 716
Bertrand's paradoxes, 58, 603
Best systems account, laws of nature, 442
Beta decay model, 538, 543
Bethe, Hans, 525, 526
Bethe-von Weizsäcker nuclear processes, 526
Betting/wager model, 384
  confirmation theory, Bayesian, 148
  Dutch Book argument, 210–213
  probability, 602
  Ramsey and, 677–678, 680
Beurton, Peter J., 72, 337
Bhaskar, Roy, 270, 271
Bias
  Akaike information criterion (AIC), 537
  experimental method, 274
  feminist philosophy, 298

Bickle, J., 826
Bifurcations, dynamical systems theory, 122
Big bang cosmologies, 831
Big Bang theory, 36, 37
Bigelow, John
  function, 318
  laws of nature, 440
Big systems, battle of, 759
Bijker, Wiebe, 776
Billiard ball model of gas, 740
Billing, J.K., 11
Billingsley, Patrick, 600
Binding, symbolic models, 153
Binmore, K., 327
Biochemistry
  experimental method, 270
  molecular biology, 481
Biodiversity, 73
Biodiversity conservation planning, 165, 801
Biogeography, 164–165, 253, 254
Bioinformatics, 485
Biological individuals, *see also* individuality
  biology, philosophy of, 71
  defining, 374–375
  kinds of, 378
Biological information, 64–68; *see also* Genetics
  evolution and development, 64–65
  informational gene concept
    history and current practice, 65
    problem of, 65–68
  innate/acquired distinction, 399
  molecular biology, 484–485
    central dogma, 335
    functional role of DNA, 489
  molecular genetics, 335–337
  philosophy of biology, 72
  terminology, 64
Biological notion of innateness, 398
Biological remnant models, 827
Biological role, function, 319
Biology
  astrobiology, 37–38
  cognitive science multidisciplinary approach, 123
  connectionism, 151
  conservation, 163–168
  emergence, 232–233
  and emergentism, 231
  evolutionary epistemology, 258
  experimental method, 270
  feminist philosophy, 298
  fitness, 310–315
  immunology, 363–369
  language/linguistics as, 109, 445, 446, 448
  Mach, 467–468
  molecular, 480–490; *see also* Molecular biology
  Nagel, structure of science, 493–494
  philosophy of, 68–75
    approaches to, 68–70
    determinism, 197
    developmental biology, 72
    ecological communities, 215–216
    ecology and conservation biology, 72–73
    evolution, 70–71, 251–257
    feminist philosophy, 296
    innate/acquired distinction, 395
    molecular biology, 71–72
    physicalism, *see* Physicalism
    systematic biology, 71
  and physicalism, 559
  prediction, 596–598
  psychology, cognitive architecture, 617
  realism, 689
  reductionism, 699, 701
  scientific metaphors, 739
  scientific progress, 750
  scientific revolutions, 755, 760
  social sciences, 783–784
  species, 795–802
  unity and disunity of science, 844, 846
Biology and Knowledge (Piaget), 261
Biomass, ecological communities, 219, 220
Biophysics, 845
Biorthogonal decomposition theorem, 646, 657
Birkhoff, Garrett, 848
  quantum logic, 633
  von Neumann, 506, 507
Birnbaum, A., 808, 811
Bishop, R., 793
Bjorklund, D.F., 267
Black, M., 381, 737–738
Black body radiation, 417, 650, 651
Black-body Theory and the Quantum Discontinuity (Kuhn), 419
Blackburn, Simon, 404
Black holes, 413
Blackmore, John T., 467
Blackmore, Susan, 262
Blank slate, 236
Blaug, M., 478
Bleier, Ruth, 298, 299
Blindness, inattentional, 548
Blind posits, 708
Blindsight, 546, 618
Blind spot, 548
Blind variation and selective retention (BVSR), 260
Block, N.
  cognitive science, 129
  consciousness, 159–160, 161, 162
  psychology, philosophy of, 618
Bloom, Paul, 4
Bloomfield, L, 107
Bloor, David, 776
Blumberg, Albert E., 458, 459, 460, 859
Boghossian, Paul
  analyticity, 14, 19
  Quine, 660, 661
Bohm, David, 656
  causality, 98
  determinism in quantum physics, 205
  hidden variable theory, 453
  quantum mechanics, 656
Bohm's theory
  determinism in quantum physics, 205
  locality, 455
  nonlocality, 453
  physical sciences, philosophy of, 555
  quantum mechanics, 656, 657
  realism, 692–693

Bohr, Niels, 356
  atom, model of, 747
  complementarity, 141–143
  quantum logic, 633, 636, 641
  quantum mechanics, 655, 657
  rational reconstruction, 684
  scientific metaphors, 737
Bohr objectivity, defined, 143
Bohr's atomic model, 747
Bohr's atomic theory, quantum mechanics, 651–652
Bohr-Sommerfield quantum theory, 120
Boltzmann, Ludwig, 342
  energy conservation, 119
  Hahn and, 341
  and Hertz, 121
  irreversibility, 411, 412
  kinetic theory, 415, 416, 417, 418
  reductionism, 700
  Schlick and, 725–726
  scientific revolutions, 755
Boltzmann constant, 650
Boltzmann entropy, 412–413
Boltzmann equation, 416, 417
Boltzmann machines, 616
Bolzano, Bernhard, 12–13
Bondi, Hermann, 36
Bonilla, Zamora, 856
BonJour, Laurence, 19, 248
Bonnet, Charles, 352
Boolean algebra logic, 639–640
Boolean algebras
  inductive logic, 386
  quantum logic, 633, 635
    classical mechanics and, 634
    partial, 633, 638, 639–640
Boolean lattices, 633–634
Boole. George, 46, 384, 570
Boorse, C., 317
Bootstrap account, Glymour's, 147–148
Borel, E., 323
Born, Max, 356, 629
Born-Oppenheimer approximation, 25, 105
Born rule
  complementarity, 141
  determinism, 204
  physical sciences, philosophy of, 556
Boscovich, Ruder J., 118
Bosons, 540
  effective field theory, 543–544
  gluons, 542
Bottom-up arguments, underdetermination of theories, 841–842
Bouchard, Frederic, 69
Boundaries
  classical mechanics, 120
  reductionism, 700
Boundary layer theory, 118–119
Boundedness
  individuality, criteria for, 377
  probability theory, 600
Boveri, Theodor, 353, 699
Boveri-Sutton hypothesis, 699
Bowler, P., 578
Box, G.E.P., 59
Boyce, M., 164, 166, 221
Boyd, Richard
  empiricism, 241
  realism, 688
  scientific metaphors, 738
Boyle, Robert, 118
  experiment, 268
  feminist philosophy, 299
  kinetic theory, 416
  perception, 546
  rational reconstruction, 684
Boyle's law, kinetic theory, 416, 418
Boys, S.F., 103
Bradie, Michael, 257
Bradley, F.H., 717
Bradshaw, G.L., 731
Brahe, Tycho, 34, 116–117
Brain, *see also* Neuroscience
  cognitive science, 123
  mechanism, 474, 477
  neurobiology
    localization and reduction, 516–518
    representation in, 520–521
  perception, 548–549
  psychology
    cognitive architecture, 616
    consciousness, 617
  Turing and, 836–837
  von Neumann and, 508–509
Brain imaging, 616, 617
Brain-mind relationship, functionalist view, 548
Braithwaite, R.B.
  game theory, 328
  Lakatos, 434
  logical empiricism, 463
  Ramsey, 674, 675, 679
  statistics, error probability philosophy, 803
Branch jumping, 761
Brandon, Robert N.
  biology, philosophy of, 69, 70, 71, 72
  determinism, 197
  fitness, 312
  individuality, 378
  intentionality, 409
  natural selection, 497, 500
  species, 799, 800
Brecht, Berthold, 304
Breeding experiments, 355
Brennan, Andrew, 215
Brenner, S., 336, 487
Brentano, Franz
  intentionality
    reintroduction of term, 406
    two theses of, 407–408, 409
  phenomenalism, 551, 552
Bresnan, J., 449
Bretthorst. G.L., 57
Brewer, W.F., 430
Bridge laws/principles, 233, 285
  confirmation theory, 147
  kinetic theory, 418
  reductionism, 699
  unity and disunity of science, 845
Bridges, Calvin B., 332

Bridgman, Percy W., 62, 75–77, 848
  demarcation problem, 191
  experimental method, 269
  instrumentalism, 402
  scientific revolutions, 758–759
  verifiability, 852
Brinkers, M., 155
Brinton, Mary C., 228
British emergentist tradition, 231
Britten, Roy J., 336
Broad, C.D.
  emergence, 232
  phenomenalism, 551, 552
Broad-sense heritability, 354
Brock, W., 102
Brodmann, Korbinian, 474
Broker, T., 483
Bromberger, Sylvain, 277
  causality, 92
  explanation, 282, 283
Brooks, R.A., 29, 619
Broome, John, 186
Brouwer, L.E.J., 357
  Hahn and, 343
  Hilbert's formalistic approach versus, 357
  Ramsey and, 674
  Russell and, 717
Brouwer's fixed point theorem, 508
Brower, L.P., 164
Brown, Harvey, 117
  quantum field theory, 632
  quantum measurement problem, 647
Brown, James R., 271, 867
Brown, K.S., 254
Brown, L.M., 538, 539, 541, 542
Brown, Robert, 417
Brownian movement/particles, 417, 507, 700, 701
Brun, Todd, 454
Brunswik, Egon, 858
Brush, S.G., 359, 417
Brush, Stephen
  Bayesianism, 45
  prediction, 590
Bub, Jeffrey
  determinism, 205
  quantum measurement problem, 648
  quantum mechanics, 656, 657
Buchanan, Allen E., 224
Buchanan, B.G., 29, 731
Buchwald, Jed Z., 269
Bueno, O., 825
Buffering, genetics, 338
Buhler, Karl, 858
Buller, David J., 3, 318
Bunsen, Robert, 35
Burali's greatest ordinal, 673
Burckhardt, R.W., 263
Burge, T., 111
Burian, Richard M.
  biology, philosophy of, 70, 73
  individuality, 378
Burks, A., 128
Burnet, Frank Mcfarlane, 366, 367, 368
Burnham, K., 536
Burtt, E.A., 754, 756
Busemeyer, Jerome R., 155
Buss, Leo, 364, 365, 377
Butler, Samuel, 257
Butterfield, Herbert, 754, 756
Butterfield, Jeremy, 120

C

Cabeza, R., 517
Calculability, effective, 833
Calculus
  algebraic, 119
  Cartan, 121
  Newton and, 118
  predicate first-order, 13
  variational, 119
Calculus of probability, 382, 854
Calculus ratiocinator, 722
Callender, C., 544
Campbell, Donald T.
  emergence, 231
  evolutionary epistemology, 257, 259, 260, 261, 262
  experiment, 269
  scientific progress, 753
Campbell, Richmond, 301, 591
Canalization, developmental, 266, 399
Canguilhem, Georges, 269, 758
Canonical commutation relations (CCR), 630
Canonical quantization, 629–630
Canonical transformations
  classical mechanics, 121
  defined, 120
  quantum logic, 634
Cantor, G.N., 739
Cantor, G.W., 504
Cantor set theory, 360, 504
Cao, Tian Yu, 543, 630, 632
Capacities
  causal, 97–98
  laws of nature, 443
Capella, Martianus, 33
Carnap, Rudolf, 79–89, 134
  analyticity, 13, 14, 16, 20
  Ayer and, 39–40
  Bayesianism, 51, 52
  cognitive significance, 132–133
    objections to, 138–140
    significance criteria, 135–138
  conventionalism, 169, 175–176
  demarcation problem, 191
  Duhem thesis, 209
  empiricism, 238, 239
  epistemology, 246
  ethics, 711
  explanation, 279
  explication, 287–294
  inductive logic, 384, 386, 387, 388–391
    desideratum 3, satisfaction of, 391–392
    historical epilogue, 393
  instrumentalism, 402
  Kuhn and, 430
  Kuhn's scientific revolutions, 424, 425
  life and work

Carnap, Rudolf (*Continued*)
    constructionist phase, 80–82
    legacy, 88
    probability and inductive logic, 86–87
    semantics, 85–86
    syntactic phase, 83–85
    Viennese positivism, 82–83
  logical empiricism, 459, 461, 463, 464
  logical positivism, 460–461
  Mach and, 467
  Nagel and, 492
  Neurath and, 513
  observation, 523–530
  phenomenalism, 551, 552
  prediction, 587
  probability theory, 600, 603, 604
  protocol sentences, 610–613
  Quine and, 659, 662–666
  Ramsey and, 679
  rational reconstruction, 681–685
  Reichenbach and, 704
  rules of succession, 50
  Russell and, 720
  Schlick and, 726
  scientific change, 730
  scientific progress, 750
  space, 80
  statistics, 809
  theory structure, 822–824
  unity and disunity of science, 844, 845
  unity of science movement, 848
  verifiability, 852, 853, 854
  verisimilitude, 855
  Vienna Circle, 858, 859, 860, 861, 862
Carnap and Logical Truth (Quine), 663, 664
Carnot, Lazare, 119, 755
Carroll, C.R., 164
Carroll, John, 443
Carroll, Lewis, 19
Carroll, Sean B., 254
Carruthers, P., 162
Carson, Scott
  biology, philosophy of, 71
  determinism, 197
  fitness, 312
Cartan calculus, 121
Carter, Brandon, 21, 23
Cartesian coordinates, 119
Cartesian dualism
  physicalism, 566
  Putnam and, 620–621
  Searle and, 770
Cartesian physics
  locality/local action, 451
  mechanics, 116
Cartesian theory of mind
  Chomsky and, 113–114
  consciousness, 160
Cartwright, Nancy, 511
  approximation, 25–26
  causality, 96–97, 99
  Duhem thesis, 210
  explanation, 279, 281
  Kuhn's scientific revolutions, 428
  laws of nature, 443
  prediction, 597
  reductionism, 700
  scientific models, 744, 746, 747, 748
  theory structure, 827
  unity and disunity of science, 846, 847
Carus, A.W., 290
Casella, G., 813
Case studies, ecological communities, 220
Casinelli, Gl, 633
Caspi, A., 399
Cassegrain telescope, 34
Cassirer, Ernest, 704, 766
Castle, D.G.A., 825
Caswell, Hal, 221
Cat, J., 511
  scientific metaphors, 739
  unity and disunity of science, 846
Catastrophic interference, defined, 156
Categoricity, 358
Category, species, 795, 796
Cattaneo, R., 483
Cauchy distribution, 57
Cauchy surface, 206, 830
Caughley, G., 164, 166
Causal anomalies, Reichenbach and, 710
Causal aspects of mechanisms, 470, 477
Causal asymmetry, time, 831
Causal chains, 282
Causal complexity, prediction, 596, 598
Causal decision theory, 183
Causal impact argument, physicalism, 566–567
Causal information concept, biological information, 67
Causal interactions, intentionality, 409
Causality, 90–101
  Aristotle and, 352
  challenges to, Galileo, Newton, and Maxwell, 92–93
  classical mechanics, Newton's laws, 117
  counterfactual accounts, 99–100
  emergence, 231; *see also* Emergence
  empiricism, 237
  explanation, 283
    deductive-nomothetic model, 277
    explanatory unification approach, 284–285
    mechanistic, 281–282
  features of, 92
  fitness, evolutionary theory, 311, 314
  generalist accounts, 95–98
    Bayesian networks and causal models, 96–97
    causal powers, capacities, universals, forces, 97–98
    probabilistic relevance accounts, 95–96
  Hanson and, 344, 346
  Hempel and, 95
  heritability, 355
  Hume and, 93–95, 237
  induction, problem of, 379
  intentionality, 409
  laws of nature, 441
  locality/nonlocality, 451, 456–457
  microscopic, 540
  Nagel on, 493
  Pearson and, 93–95
  phenomenalism, 553
  physicalism, 566–567
  physicalism, causal impact argument, 566–567

  physical sciences, philosophy of, 554, 555–556
  prediction, 586–587
  Ramsey and, 679–680
  Reichenbach and, 708–709
  Russell and, 720
  Schlick and, 728
  Searle and, 769–770
  singularist accounts, 98–99
  social sciences, 780
  space-time, 786, 790–791
  supervenience, 816–817
  time, asymmetry, 831–832
  unity and disunity of science, 846
  varieties of causation, 90–92
    reduction versus nonreduction, 91–92
    singular versus general, 90–91
  verifiability, 851, 852, 853
  visual representation, 868
Causal laws, Ramsey and, 680
Causal Markov conditions, 96, 97
Causal mechanical model, explanation, 281–282
Causal networks, scientific change, 731
Causal powers, 97–98
Causal theory of perception, 547
Causal theory of reference, 372–373
Cavalier-Smith, T., 483
Cavalli-Sforza, L.L., 267
Celestial mechanics
  classical mechanics, 118–119, 121
  scientific domains, 736
Cellular transport, transmembrane, 486
Central dogma of molecular biology, 335, 337, 483
Certainty, induction, problem of, 380
Ceteris absentibus conditions, 210; *see also* Provisos
Ceteris paribus clauses, *see also* provisos
  Duhem thesis, 210
  fitness, variances, 313
  laws of nature, 443
  Popper and, 577
  prediction, 596
  reductionism, 697
  theory structure, 825
C-functions, 392
Chakravarty, S., 508
Chalmers, D.
  cognitive science, 130
  consciousness, 160, 161
  psychology, philosophy of, 617, 618
Chance in evolution, natural selection, 500–501
Change, scientific, *see* Scientific change
Changeaux, Jean-Pierre, 259
Chaos/chaos theory/chaotic systems,
    see also Prediction
  classical mechanics, 116, 121
  determinism, 200, 203–204
  kinetic theory, 418–419
  molecular, 418
  prediction, 594–595
  scientific computing and, 508
  theory structure, 825
  Von Neumann's operator theoretic methods, 507
Chapman, S., 417
Characteristica universalis, 722
Charlesworth, Brian, 254, 584
Chemical basis of immune specificity, 366
Chemical mode of causation, 231
Chemical phylogenetics, 253
Chemical theory, 102
Chemistry
  approximation, 25
  astrochemistry, 35, 38
  emergence, 230, 231, 233
  experimental method, 270
  philosophy of, 101–106
    physical sciences, philosophy of, 555
    realism, 688
  physicalism, 560
  physical sciences, philosophy of, 557
  reductionism, 696, 700
  scientific domains, 736
  scientific metaphors, 739
  scientific models, 740
  scientific revolutions, 754, 760
  unity and disunity of science, 844
Chen, Xiang, 762
Chernyak, Leon, 364, 365
Cheverud, J.M., 355
Chew, Geoffrey, 541
Chinn, C.A., 430
Chisholm, Roderick, 239
  epistemology, 246
  intentionality, 407–408
  phenomenalism, 551
Chi-square goodness-of-fit test, 804
Choice
  axiom of, 342, 343, 672
  underdetermination of theories, 839
Chomsky, Noam, 106–115
  artificial intelligence, 31
  and behaviorism, 62–63
  cognitive science multidisciplinary approach, 124, 126
  consciousness, 161
  generative grammar
    conceptual issues in, 109–114
    development and evolution of, 107–109
  innate/acquired distinction, nativism, 396
  linguistics, philosophy of, 444–450
  psychology, cognitive architecture, 616
  scientific revolutions, 755
  Searle and, 770
Chomsky hierarchy, 114
Chord paradox of Bertrand, 58
Chow, L., 483
Christensen, Clayton, 761–762
Christensen, D., 212
Christie, J.R.R., 739
Chromosomal theory of inheritance, 353
Chromosomes
  genetics, 331, 332, 333, 334–335
  molecular biology, 486
  reductionism, 699
  Schrödinger on, 65
Chung-Ki Min, 59
Church, A.
  Carnap and, 85
  cognitive significance, 135
  Ramsey and, 675
  Turing and, 833, 834
  verifiability, 852
Churchland, Patricia S., 615

Churchland, Paul
  realism, 687
  reductionism, 700
  underdetermination of theories, 840, 841
Church-Turing thesis, 127, 833, 834, 835
Circular inertia, law of, 307
Circularity
  Harman's case against analyticity, 15
  premise, 381
  rule, 381
Circularity objection
  epistemology, 667, 668
  induction, problem of, 803
  Poincaré and, 570
  realism, 692
Cistron, 334
Cladistics, 71, 798–799
Clairaut, Alexis, 118
Claridge, M.F., 796
Clarification, explication, 292
Clarity, Hempel and, 347, 351
Clark, A., 619
Clark, J.J., 548
Clarke, Samuel, 34, 119
Classes
  classes of, 719
  versus individuals, 374
  no class theory of, 718
  Ramsey on, 672
Classical antiquity
  experimental tradition, 268
  scientific domains, 735
  scientific style, 766
  unity and disunity of science, 843
  verifiability, 852
Classical conditioning, 61
  psychology, philosophy of, 614
Classical dynamical systems, 116, 121–122
Classical empiricism, observation, 523–530
Classical ensembles, quantum logic, 635
Classical ethology, 263, 264–265
Classical field theory, quantum field theory, 630
Classical genetics, 330–333
  biology, philosophy of, 69
  molecular biology, philosophy of, 72
  reductionism, 701
Classical interpretation, probability, 602–603
Classical logic, quantum logic and, 637–638
  classical mechanics, 633–634
  probability amplitudes and Feynman paths, 636–637
Classical machines, cognitive science modeling, 127
Classical mathematics, von Neumann and, 505
Classical mechanics, 115–123
  ancient Greek and early modern mechanics, 116–117
  Bohr and, 143
  celestial mechanics and determinism, 118–119
  chaos and unpredictability in, 203–204
  complementarity, 140–143
  conventionalism, 171
  determinism, 200–204
  emergence, 234
  Feyerabend and, 306
  forces and energy, from conserved quantities to invariances, 119
  kinetic theory, 415
  mathematical mechanics and classical dynamical systems, 121–122
  Newton, 34
  Newton's laws, status of, 117–118
  particle physics, quantum field theory, 630
  prediction, 594–595
  quantum field theory, 630
  quantum logic and, 633–634, 639
  quantum theory, 650
  underdetermination of theories, 841
  variational principles and Hamiltonian mechanics, 119–121
  Von Neumann's operator theoretic methods and, 507
Classical molecular biology, 481, 482–483
Classical Greek philosophers, instrumentalism, 400
Classical physics
  Hilbert and, 359
  Nagel on, 493
Classical probability calculus, 636–637
Classical propositional calculus, 634
Classical systematics, 253
Classical thermodynamics, 554
Classification
  function and, 316
  instrumentalism, 400, 401
  scientific domains, 735
  scientific style, 766
  species, 795–802
Clatterbaugh, Kenneth, 92
Clausius, Rudolf, 416, 754, 755
Clausius' theory, 416
Clavius, C., 589
Cleanthes, 252
Clements, Frederic, 216
Clementsian community, 217
Clifton, Robert
  locality/nonlocality, 457
  quantum field theory, 632
  quantum mechanics, 656, 657
  von Neumann, 506
Climate, and evolution, 254
Climax stage of ecological succession, 215, 216
Clonal populations, as individuals, 375
Clonal selection theory (CST), 366
Cloning, experimental method, 272
Closed timelike curves, 830
Cloud chambers, particle physics, 538, 539
Cluster concepts (*Ballungen*), 513, 611
  species, phenetic, 797
  taxa as, 796
Coalescence theory, population genetics, 583, 584
Code, biological information, 64–65
Code, computer programming, 509
Codebreaking activity, Turing and, 835
Codons, 336
Coexistence laws, theory structure, 824
Coextension, lack of, 562
Cognition, 369
  analyticity, 12
  artificial intelligence simulation, 29
  cognitive science domain, 124
  empiricism, 236, 239
  instrumentalism, 400, 401
  psychology, philosophy of, 613–619

  rational reconstruction, 682
  scientific style, 765
Cognitive anthropology, 614
Cognitive architecture, 615–617
Cognitive capacities, evolution of, 260
Cognitive meaning, logical empiricism versus logical positivism, 458
Cognitive processes
  scientific change, 731
  thought style (*Denkstil*), 765
Cognitive processing
  mechanism, discovering organization of, 474
  psychology, philosophy of, 615–617
Cognitive psychology
  Chomsky, 107, 109, 110, 111
  scientific metaphors, 739
Cognitive representations, scientific revolutions and, 761
Cognitive science, 123–131
  artificial intelligence, 30–31, 771–773
  Chomsky and, 107, 109, 114
  connectionism, 151–152
    symbolic models, 153
  consciousness, 158–163
  definitions and assumptions, 123–124, 125–126
    computational assumptions, 124, 125, 127–128
    representational assumptions, 125–126, 128–129
  evolutionary psychology and, 265–266
  innate/acquired distinction, nativism, 396
  interfield relations within, 124, 126–127
  language and, 445
  neurobiology, 514, 520–521; *see also* Neuroscience
  perception, Marr's computer model of mind, 547–550
  psychology, philosophy of, 613–619
  Searle, 767–774
    artificial intelligence, 771–773
    consciousness, structure of, 770–771
  social sciences, 783
  Turing and, 833–834
  validity of, 129–131
Cognitive significance, 131–140
  Ayer and, 38, 39, 40
  beyond verifiability, 135–138
  Carnap and, 83, 86, 135–138
  criticisms/defenses, 138–140
  demarcation problem, 189, 191
  Hempel and, 83, 135–138, 350
  logical empiricism, 461–463
  prediction, 587
  protocol sentences, 610
  Quine and, 659
  scientific metaphors, 737–740
  verifiability requirement, 132–133, 852–853
  verifiability requirement, early criticisms of, 133–135
  Vienna Circle and, 860
Cognitive significance criterion, 83, 133, 134–138
Cognitive simulations, artificial intelligence, 30–31
Cognitive status of theories, Nagel on, 493
Cognitive systems, linguistics, 449
Cognitivism, and behaviorism, 63
Cognizers, as agents, 619
Cohen, C., 386
Cohen, Edward, 364
Cohen, I. B.
  classical mechanics, 118
  scientific revolutions, 754, 757–758
Cohen, Irun R., 367, 368
Cohen, J., 809
Cohen, L.J., 856
Cohen, Morris R., 491, 492
Cohen, Nicholas, 368
Cohen, Robert, 849
Cohen, S., 244, 249
Coherence theories
  epistemology, 246–247
  protocol sentences, 612
  Schlick and, 728
Cohesion species concept, 799–800
Cohn, Jonas, 682
Collapse theories of quantum mechanics, 607, 648
Collective intentionality, 783
Collins, C.B., 21
Collins, H.M., 269, 777, 778
Collins, Patricia Hill, 300
Collisions
  classical mechanics, 121, 122
  determinism, 200–203
  mechanics, 116
Color
  psychology, philosophy of, 618
  scientific domains, 734
Color force, 542
Color for Philosophers (Hardin), 618
Colyvan, Mark, 72, 73
Commensurability, Neurath and, 512; *see also* Incommensurability
Common cause principle (Reichenbach), 555–556, 709
Common knowledge, game theory, 326
Common sense, perception, 546–547
Commonsense psychology, 615
Commonsense worldview
  empiricism, 237
  phenomenalism, 552, 553
  Quine, 660
Communicability of meaning, physicalism, 558
Communities
  ecological, 215–227
  scientific, Kuhn's scientific revolutions, 422, 426–427
Commutation relations
  quantum field theory, 630
  quantum mechanics, 653
Comparative approaches
  adaptation and adaptationism, 7
  Bayesianism, 42
  confirmation theory, 145
  inductive logic, 385
  perception, 550
Comparative economic analysis, 228–229
Comparative probability, decision theory, 182
Compatibility, quantum field theory, 457
Competence level
  artificial intelligence, 31
  linguistics, 110
Competence theories, cognitive science, 127
Competing paradigms, 752
Competition
  ecological communities, 217
  evolutionary psychology, 263–264
Complementarity, 140–143
  quantum logic, 633
  quantum mechanics, 657

Complementarity logic, 633
Complementarity projection operator, 640
Complementary distribution, ecological communities, 222
Completeness, decision theory, 185
Complexity
  computational, 28
  conditional, 203
  ecological communities, 218–219
  explanation, 286
  prediction, 596, 598
  scientific models, 746–747
  unity and disunity of science, 846
  von Neumann and, 508
Complexity/stability hypothesis, ecological communities, 218–219
Componential aspects of mechanisms, 469–470, 477
Compositionality, symbolic models, 153
Compound propositions, quantum logic, 634, 637, 638
Compound systems, quantum measurement problem, 647
Comprehension, visual representation and, 869
Computation
  artificial intelligence, 27–29, 31
  cognitive science assumptions, 124, 125, 127–128
  connectionism, 150–158
  evolutionary psychology, neurological processes, 265
  experimental method, 271
  neurobiology, 509
  psychology, cognitive architecture, 615–616
  Turing and, 833–837
  unity and disunity of science, 845
  von Neumann and, 508–509
Computationalism/computational theory
  artificial intelligence, 27, 31
  Gibson's attack on, 549–550
Computational models, 740
Computational theories of mind, social sciences, 783
Computer models of mind
  artificial intelligence, 30, 31
  Marr, 547–550
    consciousness, problem of, 549
    criticism from neuroscience, 548–549
    criticism from perceptual psychology, 548
    Gibson's attack on computationalism, 549–550
Computers
  artificial intelligence, 29
  and classical mechanics, 121
  cognitive science modeling, 123, 127
  connectionist models, 150–158
  derived intentionality, 769
  experimental methods, 271
  Turing and, 835
  von Neumann and, 503, 508–509
Comte, Auguste, 35, 190, 513, 729
Comtian sociology, atomistic body, 364–365
Conant, James, 421, 848, 849
Concentration theorem of Jaynes, 55
Concepts
  complementarity, 143
  empiricism, 524
  field of experience, 844
  Hempel, 524
  Nagel on, 493
  operationalism, 76–77
  perceptual experiences as, 550
  phenomenalism, 553
  reductionism, 696
  scientific domains, 733
  scientific revolutions, 761
  scientific style, 766
  scientific style and, 765
  theory of (Kant), 12
  unity and disunity of science, 844
  verifiability, 852
Concepts of Supervenience (Kim), 818
Conceptual analysis
  Hanson, 344
  Russell and, 720
Conceptual change/evolution
  chemistry, philosophy of, 102
  Lakatos and, 435–436
Conceptual devices, instrumentalism, 400–401
Conceptual/intentional system, linguistics, 449
Conceptualization, cognitive processing, 615
Conceptual relativity, Putnam and, 620
Conclusions
  explanation, 284
  inductive logic, 385
Conditional complexity, 203
Conditional degrees of belief, 676–677
Conditionalization
  Dutch Book argument, 213
  probability, 606
  Ramsey and, 680
Conditionally deterministic theory, 199
Conditional probability
  formal theory of, 386, 534
  inductive logic, 387–388
  Kolmogorov axiomatization, 600
  quantum logic, 635, 636–637
  quantum mechanics, 654
  Ramsey and, 680
  statistics
    confirmation theory, 803
    probability assignment notation and, 804
Conditionals
  Ramsey and, 680
  verifiability, 853
Conditionals, material
  corroboration, 177
  inductive logic, 385
  quantum logic, 639
Conditionals, subjunctive, 177, 178, 607
Conditioning
  classical and operant, 61
  psychology, philosophy of, 614
Confidence, Dutch Book argument, 210–213
Confidence curves, 811
Confidence intervals, 806, 814
Configuration space
  classical mechanics, 121
  quantum field theory, 630
Confinement problem, 542
Confirmation
  cognitive significance, 136
  explicanda/explicata pairs, 288
  inductive logic, 393
  logical empiricism versus logical positivism, 458
  methodological individualism, 479

  verifiability, 851
  verisimilitude, 857
Confirmation function (Carnap), 87
Confirmation theory, 144–150
  abduction, 2
  absolute versus incremental confirmation, 144–145
  cognitive significance, 135
  comparative versus noncomparative, 145
  corroboration versus confirmation, 177
  demarcation problem, 192
  Dutch Book argument, 210–213
  Hempel, 348–349
  induction, problem of, 382–383
  inductive logic, 384, 385, 390, 393
  logical empiricism, 463
  methodological individualism, 478
  nonprobabilistic approaches, 145–148
  physical sciences, philosophy of, 555
  prediction, 586, 589–590
  probabilistic approaches (Bayesian), 148–150
  realism, 695
  scientific change, 730
  statistics, 803
  Vienna Circle, 860
Conflict, intragenomic, 498
Conformation, molecular biology, 483
Congruence, 357
Conjecture
  demarcation problem, 189
  Lakatos and, 435, 437
Conjunction, irrelevant, 147
Conjunction objection, empiricism, 241–242
Connectance, ecological communities, 218
Connectibility, Vienna Circle, 860
Connectionism, 150–158
  cognitive science, 127–128
  consciousness, 158–163
  and dynamical systems theory, 154–155
  molecular biology, 487–488
  and neuroscience, 155–157
  properties of models, 151–153
  psychology, cognitive architecture, 615–616
  psychology, philosophy of,
      consciousness, 617–618
  and symbolic models, 153–154
  Thorndike, Edward, 61
  Turing and, 836
Connectionist machines, 127, 128
Connection rules, Hanson, 344
Connectivity
  brain, 474
  molecular biology, 487–488
Connor, Edward, 222
Connors, B.W., 515
Conquest of Abundance, The (Feyerabend), 309
Conscious Mind, The (Chalmers), 617
Consciousness, 158–163; *see also* Mind/body problem
  behaviorism, 61–62
  biological significance of conscious processes, 61
  cognitive science, 130
  perception, 546, 547
    Marr's computer model of mind, 549
    neural correlates of consciousness, 550
    problem of, 549
  and physicalism, 559
  problems of, 160–162
  prospects, 162
  psychology, philosophy of, 617–618
  Searle on
    structure of, 770–771
    thought experiments, 772–773
  social sciences, 783
  Turing and, 836
  varieties of, 158–160
    creature versus state, 158
    essential versus nonessential, 158–159
    intentional verus nonintentional, 159
    phenomenal versus access, 159–160
    self-consciousness, 160
    transitive versus intransitive, 159
Consensus framework, conservation biology, 164
Consequence etiology
  causality, 90
  function, 317
Conservation area networks (CANs), 163, 167
Conservation biology, 163–168
  biodiversity concept, 163–164
  biology, philosophy of, 72–73
  consensus framework, 164
  ecological communities, 220
  multiple criterion synchronization, 167
  perspectives, 164
  place prioritization, 164–165
  surrogacy, 165–166
  viability analysis, 166–167
Conservation laws (physics)
  classical mechanics, 115, 118, 119
  determinism, 201, 202
  dynamical systems theory, 122
Conservatism, Quine and, 663
Conserved genes, 481
Conserved quantities, classical mechanics, 119
Consistency, logic of (Ramsey), 675, 677–678, 680
Consistency principle, 59
  geometrical axiom system, 358
  probability, 605
Consistency problems, quantum field theory, 629
Consistent histories approach
  quantum logic, 642–643
  quantum mechanics, 658
Constant of proportionality (Hubble's constant), 36
Constant of proportionality (likelihood principle), 44
Constitution, social constructionism, 774
Constitutive principles
  Duhem thesis, 209–210
  function, 319–320
Constraints
  developmental biology, philosophy of, 72
  feminist philosophy, 300, 302
  Gauss's principle of least constraints, 120
  methodological individualism, 479
Constructing Quarks (Pickering), 777
Constructionism/constructivism
  basis of logical empiricism, 523
  Carnap, 80–82
  constructionism versus constructivism, 775
  diffusion of, 775–776
  experimental method, 273
  reconstruction and, 682
  Vienna Circle, 861

Construction of scientific concepts
  Nagel on, 493
  social constructionism, *see* Social constructionism/constructivism
Construction of Scientific Facts, The, 777
Construction of Social Reality, The (Searle), 775
Constructive empiricism, 240–242, 528–529
  instrumentalism, 403–404
  prediction, 588
Contact action, 451
Content
  mental representations, 615
  naturalistic theory of, 131
  scientific style, 766
  visual representation, 869
Content determination, theory of, 129, 131
Contexts, Kuhn's disciplinary matrices and scientific revolution, 427
Contextual features, quantum mechanics, 656
Contextualist analysis of knowledge, epistemology, 244
Contiguity, locality/local action, 451
Contingency
  feminist philosophy, 302
  inductive logic, 390
  prediction, 598
  probability, 606
Contingency hypothesis, ecological communities, 220
Continuity
  axiom of, 357
  individuality, criteria for, 377
Continuum mechanics, 118–119
Continuum of Inductive Methods, The (Carnap), 87
Conditional probability function, 387–388
  inductive logic, 389, 390
Contradictions
  heterological, 673
  Lakatos and, 436
Control, experiment, 270
Controversies/disputes, scientific
  Hanson on, 344
  Millikan-Ehrenhaft controversy, 428
  Nagel and, 494
  Quine and, 660
  science wars, 764
  scientific revolutions, 756
  social constructionism, 777
  studies of, 269
  underdetermination of theories, 841–842
Conventionalism, 168–177
  analyticity, 13
  Carnap and, 80, 82, 84
  classical mechanics, 115, 117
  Duhem thesis and, 209
  explication, 290
  geometric, *see* Geometric conventionalism
  Grunbaum and, 170–174
  Hilbert on, 360
  instrumentalism, 401
  linguistics, philosophy of, 445
  logical positivism and, 459–460
  mathematical and logical, 174–176
  Neurath and, 512
  nontrivial, 706
  Poincaré and, 169–170, 568–569, 571
  protocol sentences, 612
  Reichenbach and, 705, 708, 711
  Schlick and, 728
  underdetermination of theories, 841
  Vienna Circle and, 859–860
Convergence, verisimilitude, 857
Convergence theorems, confirmation theory, Bayesian, 149–150
Cook-Deegan, R., 485
Cooper, Gregory, 72, 73, 223
Cooperative behavior
  altruism, 10–11
  game theory, 327
  molecular biology, 483
Cooperative games, 328
Coordinates
  classical mechanics, 121
  curve fitting, 533
Coordinating principles (Reichenbach), 705
Coordinative definitions, theory of, 705–706, 707
Copeland, B.J., 127, 834, 837
Copernican-Newtonian revolution, 421, 422
Copernican Revolution, The (Kuhn), 419, 422, 756
Copernican systems
  Kuhn's scientific revolutions, 421, 422, 423, 428
  parsimony, 532
Copernicus, N., 715, 754
  anthropic principle, 23
  astronomy, 34
  Bayesian rationale, 52
  Feyerabend on, 307
  instrumentalism, 400
  Kuhn's scientific revolutions, 427
  parsimony, 532
  scientific revolutions, 761
Copi, I., 386
Copy theory of knowledge, 262
Coreference, incommensurability, 372
Corey, R.B., 482
Corporeal atomism, 364
Corpuscularian theory, 102
Correlated equilibrium, 327–328
Correlation between relatives, population genetics, 580–581
Correspondence rules
  Ramsey and, 679
  theory structure, syntactic (received) view, 823
Correspondence theory of truth, 576, 612, 622
Corroboration
  Bayesianism and, 45
  demarcation problem, 192
  induction, problem of, 382
  prediction, 586
  scientific change, 730
  scientific progress, 751
  verisimilitude, 857
Cosmic ray physics, 538, 539
Cosmides, Leda
  evolutionary psychology, 264, 265, 266
  function, 319–320
Cosmology, *see also* Anthropic principle
  astronomy and, 32, 33, 36
  Newton on, 118
  physical sciences, philosophy of, 556

  quantum, 658
  time, beginning of, 831
  unity and disunity of science, 843
Cotterill, R., 617
Coulomb, Charles Augustin de, 119
Coulter, Jeff, 776
Counterexamples
  covering law models, 284
  explanation, 348
  Gettier problem, 244
  Lakatos and, 437–438
Counterfactual accounts, *see also* Causality
  causality, 91, 99–100
  cognitive science, 129
  Duhem thesis, 210
  history, Nagel on, 495
  physicalism, 559
  probability theory, 600, 607
  reductionism, 698
  verifiability, 853
Counterinduction, 381
Cournot, A.A., 607
Couturat, L., 570
Couvalis, George, 784
Covariance terms, heritability, 355
Covariation, phenetic, 797
Covering-law models
  causality
    causal powers, capacities, universals, forces, 97–98
    explanation versus, 95
    singular versus general, 90–91
  explanation, 276–280
    deductive-nomological (D-N), 276–279
    deductive-statistical (D-S), 276, 278–279
    inductive-statistical (I-S), 276, 279–280
  scientific progress, 750
Cowie, Fiona, 396
Cox, D.R., 803, 804, 808, 811
Cox, R.T., 42, 600
Coyne, Jerry, 254, 582
Cracraft, J., 799
Craig, I.W., 399
Craig, William, 402–403
Craig's theorem, 402, 403
Crane, Diana, 430
Crane, T.
  Chomsky, 113
  intentionality, 406
  physicalism, 560
Crasnow, S.L., 825
Craver, Carl
  biology, philosophy of, 72
  explanation, 282, 286
  mechanism, 469, 472, 473, 474, 475
Crawford, C., 264
Crease, Robert, 762
Creath, R., 138–139, 293
Creationism/creation science, 194–195
Creature consciousness, defined, 158
Crick, Francis H.C., 335, 336, 337, 366
  biological information, 65
  consciousness, 160, 161
  experimental method, 271
  genetics, 332
  molecular biology, 480–481, 485, 487
  parsimony, 532–533
  perception, 549
  psychology, philosophy of, 617
  realism, 689
  theory structure, models and, 827
Crisis science, 752
Crist, Eileen, 365
Criterion of ignorance, Bayes's, 46
Criterion problem, 193
Criticism and the Growth of Knowledge (Lakatos and Musgrave), 421
Critique of Pure Reason (Kant), 11, 360
Critique of the Power of Judgement (Kant), 232
Crombie, A.A., 766
Cross entropy, 54
Cross-reference, symbolic models, 153
Crow, J., 581, 583
Crusoe example, 513
Cruz, J., 245, 246, 247
Cryptanalysis, 835
Csiszar divergence, 55
Culp, S., 484
Cultural domains of objects, 80
Cultural historians, scientific revolutions, 758
Cultural materialism, social sciences, 782
Cultural objects, 80
Cultural philosophy, Schlick, 726–729
Cultural relativism, Feyerabend and, 305
Culture
  evolutionary epistemology, 262
  evolutionary psychology, 264
  feminist philosophy, 299, 302
  Neurath and, 511
  scientific style, 765
  and self-consciousness, 160
  social sciences, 783
Culture, scientific disciplines, *see* Disciplinary matrices and scientific revolution
Cummins, Robert, 3
  biology, philosophy of, 69
  connectionism, 153
  explanation, 286
  function, 318
Cummins function, 318
Current-algebra approach to strong interactions, 541
Curved space-time, 793–794, 830
Curve fitting, 743
Cushing, James
  locality, 452
  realism, 693
Cushing, J.T.
  complementarity, 141
  particle physics, 541
C-value paradox, molecular genetics, 336, 483
Cybernetic models, Darwinian and Lamarckian elements, 262
Cybernetics
  unity and disunity of science, 845
  unity of science movement, 848–849
Cyclic organization, mechanisms, 475
Cytoarchitecture, brain, 474
Cytology, 480
Cziko, Gary, 259, 262

D

da Costa, N.C.A., 826
da Costa, Newton, 741
Dale, A., 384
D'Alembert, Jean le Rond, 118, 843
D'Alembert's principle, 120
Dalton, John
  chemistry, philosophy of, 102
  scientific domains, 733
  scientific revolutions, 759
Danneberg, Lutz, 862
Darden, Lindley
  biology, philosophy of, 71–72
  explanation, 282
  mechanism, 469, 473, 474
  molecular biology, 488
  reductionism, 699
  unity and disunity of science, 846
Dark matter, 38
Darwin, Charles, 52, 112
  abduction, 1
  adaptation and adaptationism, 3, 4, 7
  altruism, 8
  evolution, 252
  evolutionary epistemology, 257–262
  fitness, 310–315
  heritability, 353
  and Mach, 468
  methodological individualism, 479
  natural selection, 497–502
  scientific metaphors, 737
  scientific revolutions, 754, 756, 760
  scientific style, 767
Darwin, Erasmus, 252
Darwinian anthropology, 264
Darwinian fitness, 311
Darwinian model
  and immunology, 365
  scientific theories, Popper and, 575–576
  unity and disunity of science, 844
Darwinism
  adaptation and adaptationism, 3–7
  developmental biology, philosophy of, 72
  epistemological, 260, 261
  evolutionary epistemology, 260
  heritability, 353
  immunology, 365
  population genetics, 578–585
  scientific revolutions, 754, 760
  unity and disunity of science, 846
  universal, 258
Daston, Lorraine, 869
Data
  neurobiology, 515
  scientific change, 731
  scientific models, 743
  underdetermination of theories, 839
  visual representation, 868
Davidson, Donald
  causality, 90, 91
  empiricism, 239
  incommensurability, 372
  intentionality, 408
  laws of nature, 443
  physicalism, 559, 566
  protocol sentences, 612–613
  scientific revolutions, 764
  scientific style, 767
  supervenience, 816
Davies, M., 160, 161
Davies, P.S., 320
Dawah, H.A., 796
Dawkins, M.S., 264
Dawkins, Richard, 4
  altruism, 8, 9
  evolutionary biology, philosophy of, 70
  evolutionary epistemology, 261
  evolutionary psychology, 264
  individuality, 376, 378
  natural selection, 498, 499
Day, Timothy, 1
D-consistent sets of histories, quantum logic, 642–643
Deacon, Richard, 716
Dear, Peter, 779
de Broglie, Louis, 656
de Broglie-Bohm theory, 656
de Broglie pilot wave theory, 120
Debye, P.J.W.
  Reichenbach and, 704
  Zeeman effect, 652
De caelo (Aristotle), 33
Decidability of mathematical propositions, Turing and, 833
Decision matrix, 181–182
Decision theory, 181–188, 507
  Bayesianism, 41
  completeness, 185
  confirmation theory, 148
  decision problems, 181–183
  Dutch Book argument, 212
  economic rationality, 227
  expected utility representation of preference, 183–185
  game theory, 323–329
  methodological individualism, 479
  Neyman-Pearson (NP) tests, 806
  probability, 605
  Ramsey and, 679, 680
  sure-thing principle, 185–188
Declining population paradigm, conservation biology, 164
Decoherence, quantum measurement problem, 648–649
Decomposition
  mechanism, identification of components, 473–474
  neurobiology, 515
  space-time, 790
Decomposition rules, generative grammar, 107
Deduction
  explanation, 348
  geometry, Euclidean, 116
  induction, problem of, 380
  Nagel on, 493
  Ramsey and, 680
  reductionism, 698
Deductive logic
  Carnap, Rudolf, 88
  relation to inductive logic, 384–385, 391–392
Deductive-nomological (D-N) models
  causality, 95
  explanation, 276–279
  functional explanations and, 286

  kinetic theory, 418
  mechanism, 473
  reductionism, 697
  unity and disunity of science, 845
Deductive-nomothetic model of probabilistic (DN-P) explanation, 281–282
Deductive-statistical (D-S) models, 276, 278–279, 348
Deep structure, transformational grammar, 107
D'Espagnat, Bernard, 647
Deficits, linguistic, *see* Normative concepts/normativity
de Finetti, Bruno
  Bayesianism, 50, 51, 148, 149–150
  Carnap and, 87
  confirmation theory, 148
  Dutch Book argument, 211
  probability, 600, 606, 608
  Ramsey and, 675, 677, 680
de Finetti's representation theorem, 50, 51
Definitions
  confirmation theory, 145
  explication, 287, 292
  impredicative, 570
  Lakatos and, 435
  Quine, 661
  Reichenbach and, 705, 706
  of science, research programs, 712–715
Deflationary metaphysics, 664, 665
De Gelder, B., 516
Degrees of belief
  confirmation theory, Bayesian, 149–150
  Ramsey, 676–677
Degrees of freedom
  quantum field theory, 629, 632
  quantum logic, classical mechanics and, 633–634
Degrees of probability, 675
De Groot, Morris, 57
DeJong, Willem R., 12
Delbrück, Max, 335
Demarcation, problem of, 188–197
  cognitive significance, 132
  difficulties and reasons for demise of philosophical problem, 193–194
  historical background, 189–193
  Popper and, 572–573, 574
  as social problem, 194–196
Demarcation criterion (Popper), 535–536, 573, 574
Democracy, particle, 541–542
Democratic participation in science, Feyerabend on, 308
Democratic relativism, 308
Democritus, 843
Demography, 480
De Moivre, Abraham, 46
De Morgan, Augustus, 52
De Muynck, Willem, 454
DENDRAL system, 731
Denkkollectiv (thought collective), 766
Denkstil (thought style), 765
Dennett, Daniel C.
  adaptation and adaptationism, 7
  artificial intelligence, 29
  biological information, 66
  evolutionary epistemology, 259
  fitness, 311
  intentionality, 409
  natural selection, 498
  psychology, philosophy of, 618, 619
  reductionism, 702
Denotation, scientific models, 744–745
Denotation/extension, explicanda/explicata pairs, 288
Deoccamization, cognitive significance, 138
de Queiroz, K., 799–800
De Regt, Henk W., 417
De revolutionibus orbium coelestium (Copernicus), 754
Derivative intentionality, 408
Derived intentionality, 408, 769, 772
DeRose, K., 244
Der Raum (Carnap), 80
Descartes, Rene, 589
  demarcation problem, 189
  evolution, 252
  explanation, 276–277
  Hanson and, 346
  innate/acquired distinction, 395, 397
  locality/local action, 451
  mind-body problem, 61–62
  perception, 545
  phenomenalism, 551
  scientific revolutions, 754, 757
  unity and disunity of science, 843, 844
Descartes-Euler formula, 434, 435, 437
Descent, individuality, 376–377
Descent of Man, The (Darwin), 257
Description
  approximation, 24
  prediction, 597
  Reichenbach and, 706
  scientific domains, 735
  scientific models, 744
Descriptive adequacy versus explanatory adequacy, 108, 109
Descriptive grammar, 108
Descriptive simplicity, Reichenbach and, 707
Descriptivism
  causality, 93
  classical mechanics, 121
  function, 316
Desiderata, inductive logic, 384, 391–393
Design
  function, 315
  scientific computing and, 508
Design problem, fitness, 311–312
Desirability, confirmation theory, 148
de Sitter space-time, 206
Destouches-Fevrier, Paulette, 633
Determinate-property state, prediction, 596
Determination, supervenience, 816–817
Determinism, 63, 197–208
  causality, 90
  chaos and unpredictability in classical physics, 203–204
  classical mechanics, 115, 118–119
  in classical physics, 200–203
  emergence, 232
  empirical evidence, 207
  feminist philosophy, 298
  formulations of, 198–200
  historical, Nagel and, 492
  Laplace on, 501
  Nagel and, 493, 495
  natural selection, 500–501
  physical sciences, philosophy of, 555
  prediction, 594, 596

Determinism (*Continued*)
  quantum logic, 634
  quantum mechanics, 654
  in quantum physics, 204–205
  in relativistic physics, 205–207
  scientific metaphors, 738
  unity and disunity of science, 846
Deus ex machina explanations, 757
Developmental biology, 480
  biological information, 64–65
  biology, philosophy of, 72
  evolutionary, 254
  evolutionary psychology, 266
  heritability, 352–353
  and immunology, 365
  innate/acquired distinction, 395, 398, 399
  molecular biology, 481, 486
Developmental canalization, 266, 399
Developmental genetics, 481
Developmental individuals, 378
Developmental information, molecular biology, 72
Developmental models, 740
Developmental psychology, 616–617, 757
Devitt, M., 129
Dewdney, C., 600
Dewey, John, 257
  evolutionary epistemology, 262
  instrumentalism, 400
  Nagel and, 491, 495
  Nagel on, 493
DeYoe, E.A., 516
Diagnostic species, 799
Diagnostic traits, 395
Dialectical approaches
  genetics, 331
  immune self, 365
  Lakatos and, 434, 435, 436
  prediction, 590–591
Dialects, generative grammar, 110
Dialogic approaches, quantum logic, 642
Dialogues Concerning Natural Religion (Hume), 252
Diamond, C., 835
Diamond, Jared, 221–222
Dicke, R.H., 23
Dickson, Michael, 648, 657
Diderot, Denis, 843
Dieks modal interpretation, quantum mechanics, 657
Dietrich, Michael, 253
Diffeomorphism, 205
Difference measure, Bayesian confirmation theory, 148
Differential equations
  classical mechanics, 115, 120
  von Neumann's numerical analysis of partial differential equations, 503
Differential forces, Reichenbach and, 707
Differential geometry
  and classical mechanics, 121
  projective geometry versus, 357
Dilemma, theoretician's, 62
Dilthey, W., 511, 765, 844
Dilworth, Craig, 752
DiMaggio, Paul M., 228
Dimensionality, Von Neumann's operator theoretic methods, 507
Dingler, Hugo, 269, 272
Dirac, P.A.M., 23
  chemistry, philosophy of, 103
  particle physics, 539
  quantum field theory, 629, 630
  quantum mechanics, 654
  Von Neumann-Dirac collapse formulation of quantum mechanics, 595
Dirac-von Neumann model, 654
Directedness, function, 316, 320–321
Direction, kinetic theory, 416
Direction of Time (Reichenbach), 709
Dirichlet priors, 51
Disagreement point, game theory, 328
Disciplinary matrices and scientific revolution, 421–427
  contexts: discovery, justification, development, 427
  elements of, 421
  normal science, 423–424
  progress, 422–423
  rationality, 424–425
  rationality and the social, 425–427
  scientific progress, 751
  Scientific Revolution, The, 427
Discontinuity, quantum, 651
Discontinuous change of state, 648
Discourse
  American Structuralism, 107
  scientific versus nonscientific, 86
Discovery, *see also* Scientific change
  American Structuralism, 107
  Hanson, on observation, 345
  Hanson on, 344, 346–347
  Kuhn's disciplinary matrices and scientific revolution, 427
  Lakatos and, 437
  linguistics, American Structuralism, 107
  mechanism, explanatory role of, 473
  of mechanisms, 473–474
    experiments in, 475–477
    organization of, 474–475
  rational reconstruction and, 684
  scientific change, 730–731
  scientific domains, 736
Discrete state machine, Turing and, 833
Discursive domains, 80
Disjunction problem
  function, 319
  physicalism, 562–563
Disposition, verifiability, 852, 853–854
Dispositional concepts, empiricism, 238
Dispositional predicates, 177, 853
Dispositional realism, chemistry, 103
Disputes, scientific, *see* Controversies/disputes, scientific
Disruptive technologies, 762
Distributed connectionism, psychology, 615
Distribution function, kinetic theory, 416
Distribution law, Maxwell's, 416
Distributive justice, 228
Distributivity, law of, 648
Diversity
  ecological communities, 219–220
  innate/acquired distinction, 395
DNA
  adaptation and adaptationism, 7
  biological information, 64–65, 66

  evolutionary biology, philosophy of, 70
  experimental method, 271
  genetics, 332, 339
    genetic code, 336
    introns and exons, 337
    molecular, 335–337
    Watson-Crick model, 333, 335, 336, 337
  heritability, 353
  molecular biology, 480–483
    functional role, 489
    G-value paradox, 486
    information, 484–485
    proteomics and, 487
    sequestered modular templates (SMT) model of cell, 488
  realism, 689
  reductionism, 698
DNA models, 827
DNA replication, 335
DNA sequences, 480, 481, 483
Dobzhansky, T., 253, 254
Dogma, neurobiology, 517
Dogmatism, evolution, 254
Domain of intended applications, theory structure, 826
Domain problems, QM, von Neumann and, 506
Domains, cognitive, 616–617
Domains of inquiry, Kuhn's disciplinary matrix, elements of, 421
Domains of objects, 80
Domains of science, *see* Scientific domains
Domain-specific effects in cognition, 616
  evolutionary psychology, 266
  function, 319–320
Domain-specifying assumptions, cognitive science, 125
Dominance
  decision theory, 182–183
  game theory, 326
Dominance, gene/allele, 330, 354
Donati, Giovan Battista, 35
Donoghue, M.J., 799
Donovan, A., 731
Doppelt, Gerald, 840
Doppler shift, spectral lines, 35
Dore, Mohammed, 508
Double slit thought experiment, Reichenbach, 710
Dowe, Phil, 99
Downes, S.M., 827
Doxastic theories, 245–246, 247
Draper, Henry, 35
Dray, William, 845
Dresden, M., 541, 542
Dretske, D., 158
Dretske, Fred
  causality, 97, 98
  intentionality, 408, 409
  laws of nature, 439, 440, 441
Dretske-Tooley-Armstrong account, 98; *see also* Laws of nature
Drèze, Jean, 229
Driesch, Hans, 132
Drift, evolutionary, 254–255, 497
  and fitness, 313–315
  natural selection, 500–501
  population genetics, 579, 582–583, 584
Drosophila melanogaster (fruit fly), 331, 332, 333, 355, 486
Dualism
  behaviorism, 61, 63
  methodological, 114
Dubislav, Walter, 704
Ducasse, C., 98
Duhem, Émile, 341
Duhem, Pierre, 147, 589
  chemistry, philosophy of, 101
  conventionalism, 168, 169
  demarcation problem, 191
  Duhem thesis, 208–210
  experimental method, 269
  feminist philosophy, 301
  instrumentalism, 400, 401, 402, 403
  Popper and, 577
  scientific revolutions, 757
  underdetermination of theories, 841
  Vienna Circle, 859–860
Duhem-Quine thesis, 841
Duhem thesis
  confirmation theory, 147
  conventionalism, 169, 170
  demarcation problem, 191, 192
  empiricism, 239
  feminist philosophy, 301
  induction, problem of, 382
  instrumentalism, 401
  logical empiricism, 462
  Neurath and, 512, 514
  Poincaré and, 569
  protocol sentences, 611
  Quine and, 662
  statistics, 810
  underdetermination of theories, 840
  Vienna Circle and, 859–860
Dumbell model, kinetic theory, 417
Dupré, John
  adaptation and adaptationism, 5
  evolutionary biology, philosophy of, 70
  explanation, 279
  systematic biology, philosophy of, 71
  unity and disunity of science, 846, 847
Durkheim, Emile, 478, 784
Dürr, Detlef, 656
Duschl, Richard A., 430
Dutch Book argument, 210–213
  confirmation theory, Bayesian, 148–150
  converse, 677
  prediction, 593
  probability, 605
  Ramsey and, 677
Dvurecenskij, Anatolij, 640, 643
Dynamical autonomy, 701
Dynamical law of quantum mechanics, 655
Dynamical structure, Minkowski space-time, 791
Dynamical systems
  chaos and unpredictability in classical physics, 203–204
  classical mechanics, 116, 121–122
    algebraic calculus, 119
    reduction of dynamics to statics, 120
  complementarity, 143
  determinism, 203
  physical sciences, philosophy of, 555
  Von Neumann's operator theoretic methods, 507

Dynamical systems theory
  classical mechanics, 122
  connectionism, 154–155
Dynamical theory, Maxwell, 452
Dynamic approaches to cognition, 615, 616
Dynamics
  causality, 93
  language, Lakatos and, 436

E

Earman, John, 204
  anthropic principle, 22
  complementarity, 141
  confirmation theory, 150
  determinism, 198, 199, 200, 201, 202, 203, 205, 206, 207
  fitness, 312–313
  induction, problem of, 382
  laws of nature, 443
  physical sciences, philosophy of, 555
  prediction, 595
  probability, 606
  quantum measurement problem, 647
  Ramsey and, 680
  realism, 692, 693
  time, 830
  underdetermination of theories, 841
Eccles, John
  emergence, 231
  evolutionary epistemology, 259
  Popper, 572
Ecological communities, 215–217
Ecological fitness, 311, 313–315
Ecology, 215–224
  balance of nature, 217–220
  biology, philosophy of, 72–73
  conservation biology, 163–168
  metaphysics and ecological communities, 215–217
  molecular methods, 480
  species, 800
  theories: contingency, predictive accuracy, explanation, 220–223
    evaluation of models, 221–222
    explanatory value of models, 222–223
    feasibility of models, 220–221
  theory structure, 825
Economic models, 741
Economic rationality, 227
Economics
  altruism, 10
  experimental method, 274
  game theory, 323–329
  Nagel and, 495–496
  philosophy of, 224–230
    ethical values in role of, 227–228
    feminist issues, 298
    intellectual role, 225
    laws of economics, 225
    Malthusian, 364
    methodological individualism, 226, 478
    Neurath and, 511–512
    rational debate, 228–229
    rationality, economic, 227
    realism, 226
    testability/falsifiability, 226–227
  Ramsey and, 671
  social sciences, 781
  theory structure, 825
  unity and disunity of science, 846
  von Neumann and, 507–508
Ecosystem descriptions, approximation, 24
Ecosystem ecology, 215, 801
Ecotone, 217
Eddington, Arthur S.
  conventionalism, 170–171
  space-time, 793–794
  Turing and, 833, 837
Edelman, Gerald M., 259
EDVAC, Turing and, 835
Edwards, A.W.F.
  Bayesian confirmation theory, 149
  parsimony, 533
  population genetics, 581
Edwards, W., 55, 808
Eels, Ellery, 149
Effect
  causality, 90
  law of (Thorndike), 61
Effective calculability, 833
Effective field theory, 543
Efficient causes (Aristotle), 352, 780
Efron, Bradley, 55
Egerton, Frank, 218
Egoism, 9, 11
Ehlers, J., 787
Ehrenhaft, Felix, 428, 429
Eibl-Eibesfeldt, Irenaus, 263
Eigenstate-eigenvalue link, 645, 646–647
Einheitswissenshaft, *see* Unity of science movement
Einstein, Albert, 356, 392
  demarcation problem, 191
  Einstein-Podolsky-Rosen relation, 452–453
  and empiricism, 237
  Feyerabend and, 306
  general relativity, *see* General relativity
  on geometry, 360
  Hilbert and, 359
  instrumentalism, 403
  kinetic theory, 417
  locality/local action, 451
  logical positivism, 460
  Mach and, 467, 468
  particle physics, 538–539
  quantization, 651
  quantum mechanics, 655
  realism, 692–693
  reductionism, 700
  Reichenbach and, 704, 705, 706, 709, 711
  Russell and, 716
  Schlick and, 727–728
  scientific domains, 733
  scientific metaphors, 737
  scientific revolutions, 756, 757, 759, 760
  scientific style, 765
  simultaneity, definition of, 705
  space-time
    general relativity, 791–794
    special relativity, 786, 788–791, 792

special relativity, *see* Special relativity
  unity and disunity of science, 844
  verifiability, 852
Einstein-Podolsky-Rosen relation
  locality, 454
  quantum field theory, 632
  quantum mechanics, 655
Einstein's field equations, 205, 792
  approximate solution to, 24
  Hilbert, 359
  physical sciences, philosophy of, 555
  time, 830, 831
Ekbohm, G., 313
E-language, 445
Elastic collisions, determinism, 200–203
Eldredge, N., 799
Electrodynamics, 788, 794
  Hilbert and, 359
  particle physics, 538
Electromagnetic force, unified description, 542
Electromagnetic theory, observation, 525–526, 527
Electromagnetism
  Ampère's experiment, 270
  and locality/local action, 452
  particle physics, 539
  quantization, 629, 651
  scientific revolutions, 755
Electronic Discrete Variable Automatic Computer (EDVAC), 835
Electrons
  discovery of, Kuhn on, 428–429
  leptons, 539
  particle physics, 539
Electro-weak theory, 543–544
Elementary experiences (elex), 81, 82, 131–132, 523
Elementary particle physics, *see* Particle physics
Elements (Euclid), 360
Eliminative empiricism, 237
Eliminative instrumentalism, 402–403
Eliminativism
  causality, 93
  psychology, philosophy of, 615, 618
Ellner, S.P., 167
Ellsberg, Daniel, 186, 188
Ellsberg paradox, 186, 187
Elman, Jeffrey L., 152–153, 155
Elster, Jon
  economics, theory of, 228
  methodological individualism, 479
  social sciences, 781
Elton, Charles, 221
Elucidation explicandum, 293
Embedded cognition, 616, 618–619
Embodied cognition, 616, 618–619
Embryology, *see also* Development biology
  heritability, historical development of ideas, 353
  and immunology, 365
Emergence, 230–235; *see also* Causality
  in contemporary philosophy of science, 234
  criticisms, 233
  German organicism, 232–233
  history of concept, 231–232
  locality/local action, 455
  molecular biology, 484
  Nagel on, 493
  physicalism, 560
  physical sciences, philosophy of, 557
  reductionism, 698, 702
  unity and disunity of science, 846
Emergentism, 92, 230
Emotion, scientific style, 765
Empirical beliefs, protocol sentences, 612–613
Empirical content, research programs, 713
Empirical equivalence
  realism, 692–693
  underdetermination of theories, 840–841
Empirical evidence, determinism, 207
Empirical knowledge, protocol sentences, 82–83
Empirical linguistic theory, Chomsky and, 108
Empirical prediction, models of, 587–589
Empirical progress, 714–715
Empirical sciences
  demarcation problem, 572–573
  geometry as, 360
  prediction, 586, 594–598
    biological and social sciences, 596–598
    classical mechanics and chaotic systems, 594–595
    quantum mechanics, 595–596
Empirical tests
  ecological diversity/stability hypothesis, 219
  economic theories, 226
Empiricism, 235–243; *see also* Logical empiricism
  analyticity
    Harman's case against, 17–18
    Quine and, 13–14
  axiomatics and, 289–290
  Ayer and, 39
  Bayesianism, 47
  behaviorism, 61–63
  Carnap and, 82
  causality, 93–94
  classical mechanics, 121
    Galileo, 116
    Newton, 117
  cognitive significance, 134–138
  consciousness, 160
  constructive, 240–242, 528, 588
  demarcation problem, 189, 190
  Duhem thesis, 208–209
  explanation, deductive-nomological (D-N) model, 276
  feminist philosophy, 297, 300, 301, 302
  Feyerabend and, 305–306
  function, 316
  genetics, 331
  and geometry, 169
  Hempel and, 350
  historical background
    early modern period, 236–237
    early 20th century, 237–239
  induction, problem of, 379
  innate/acquired distinction, 394
  instrumentalism, 400, 402
  kinetic theory, 417
  Lakatos and, 438
  liberalization of, 83
  logical empiricism, 458–465
  methodological individualism, 479
  Neurath and, 510–512

Empiricism (*Continued*)
  observation, 523–530
  perception, 545–546, 547
  physical sciences, philosophy of, 555
  Poincaré and, 568
  after positivism, 239–241
  prediction, 586–587, 588
  Quine and, 13–14, 660–666
  realism, 688
  Reichenbach and, 705
  research programs, 713–714
  Schlick and, 726
  scientific change, 729–732
  scientific domains, 733, 734, 735, 736
  scientific progress, 750
  scientific revolutions, 755, 756, 764
  skepticism about, 242–243
  social constructionism, 776
  unity and disunity of science, 844, 845
  verifiability, 851, 853
  Vienna Circle, 860
Empirical-analytical science, and technology, 269
Empty extension, Harman's case against analyticity, 15
Encapsulated capacities, cognitive architecture, 616
Encyclopedia model
  Neurath, 513–514
  unity and disunity of science, 844–845
  Vienna Circle, 859, 860
Encyclopedists, French, 843, 848
Endler, John A.
  adaptation and adaptationism, 7
  species, 796
Endscheidungsproblem, Turing and, 833
Energeticists, classical mechanics, 120
Energy
  classical mechanics, 119
  mechanics, 116
  phase space, 411–412
  unity and disunity of science, 844
Energy conservation, 116, 120, 122
Energy levels of hydrogen atom, 651–652
Engelmann, Paul, 861
Engineering, artificial intelligence as, 29–30
Enigma Code, 835
Enlightenment
  scientific revolutions, 755
  unity and disunity of science, 843
Enquiry Concerning Human Understanding (Hume), 47
Ensemble properties, fitness, 310–311
Ensembles, quantum logic, 635, 637
Enskog, D., 417
Entailment
  deductive and partial, 384
  inductive logic, 387, 392–393
Entanglement
  locality, 452–453
  nonlocality, 456
  quantum mechanics, 654
Entelechies, 90, 132
Enthymeme, 277, 379
Entity realism, 271
Entropy
  causality, 92
  chaos and unpredictability in classical physics, 203
  classical mechanics, 122
  irreversibility, 411, 412–413
  kinetic theory, 418
  MAXENT and MINENT, 54, 55, 58, 59
  physical sciences, philosophy of, 554, 555
  quantum mechanics, 651
  statistical-mechanical (Boltzmann), 412–413
  unity and disunity of science, 844
  Von Neumann's operator theoretic methods, 507
Enumerative induction/inference, 2, 380
Environment
  adaptation and adaptationism, 4–5
  behaviorism, 62
  body, atomistic, 364
  classical ethology, 264–265
  consciousness, 160
  evolutionary biology, philosophy of, 70
  evolutionary psychology, monomorphic mind thesis, 266
  experimental system interactions with, 270, 273
  innate/acquired distinction, 398
  intentionality, 409
  linguistics, Chomsky, 111
  natural selection, 500
  perception, 548
  and phenotypic variation, 331
  psychology, philosophy of, 619
  scientific models, 740
Environmentalism, ecology and, 215
Environmental sciences, scientific change, 731; *see also* Ecology
Epicurean socialism, Neurath and, 511
Epidemiology, 480
Epigenesis, 339, 352, 353
Epistatic components, genetic variance, 354
Epistatic interactions, heritability, 355
Episteme, scientific style, 765
Epistemically privileged language, protocol sentences, 82–83, 610
Epistemic anthropic principle (EAP), 23
Epistemic approaches, inductive logic, desideratum 3, satisfaction of, 392
Epistemic authority of science, demarcation problem, 195–196
Epistemic foundationalism, of Ayer, 40
Epistemic instrumentalism, 402, 403–405
Epistemic interpretations of probability, in inductive logic, 386–387
Epistemic justification, *see* Justification, theories of
Epistemic order of objects, 80
Epistemic seeing, (see that) Hanson and, 345
Epistemic significance, prediction, 588, 589–592
Epistemography, 779
Epistemological antifoundationalism, 83
Epistemological Darwinism, 260, 261
Epistemological theory of relativity, 706
Epistemology, 244–251
  approximation, 26
  of arithmetic, 569–570
  Bayesianism, 59
  biology, philosophy of, 69
  Carnap and, 82, 88
  causality, 90, 91
  chemistry, philosophy of, 105
  classical mechanics, Newton's laws, 117

  coherence theories, 246–247
  consciousness, 160, 161
  constructionism, 80–81
  demarcation problem, 189–190
  determinism, 197
  Duhem thesis, 208–210
  emergence, 230, 231
  empiricism, 240, 242, 243
    Ayer and, 238–239
    constructive, 240–241
  epistemic change, 249–250
  evolutionary, *see* Evolutionary epistemology
  experimental method, 268, 270, 271, 272–273, 274
  explanation, 284–285
  Feyerabend and, 305
  foundationalism, 245–246
  Gettier problem, 244–245
  Hanson, on facts, 345–346
  Hilbert and, 358, 360–361
  immune self, 364
  immunology, biological identity, 365
  innate/acquired distinction, 395–396
  instrumentalism, 400, 401
  justification, theories of
    doxastic/nondoxastic and internalist/externalist, 245
    externalist, 248–249
  Kuhn's scientific revolutions, 427
  Lakatos and, 437
  laws of nature, 442
  linguistics, philosophy of, 448
  logical empiricism, 464
  mechanism discovery, 477
  molecular biology, 489
  Neurath and, 512, 513
  observation sentences, 666–668
  parsimony, 531, 532
  physical sciences, philosophy of, 555, 556, 557
  Popper and, 572
  protocol sentences, 611
  Putnam and, 622–623, 626
  Quine and, 660, 666–668
  Ramsey, logic of truth, 680
  rational reconstruction, 681–685
  realism, 686, 688
  recursive, 360–361
  Reichenbach and, 709, 711
  Russell and, 720
  Schlick and, 727
  scientific change, 729–732
  scientific metaphors, 738, 739
  scientific models, 740, 744–746
  scientific revolutions, 422, 755, 757
  social constructionism, 778, 779
  space-time, 791
  theory structure, semantic view, 825
  Vienna Circle and, 859
  visual representation, 866, 868
Equations
  Kuhn's disciplinary matrix, elements of, 421
  scientific models, 744
Equations of motion
  classical mechanics, 121
  conservation of energy, 119
  particle physics, 543
Equilibrium analysis, game theory, 326
Equilibrium
  classical mechanics, Lagrange, 119
  game theory, evolutionary, 329
  Von Neumann's operator theoretic methods, 507
Equivalence
  artificial intelligence, weak versus strong, 30–31
  confirmation theory, 145
  Jaynes's principle of, 56
Equivalence principle, 794
Equivalent descriptions, theory of (Reichenbach), 706, 710
Eratosthenes, 33
Érdi, P., 615
Ereshefsky, Marc, 71, 800
Ergodicity, 203
  classical mechanics, 122
  von Neumann and, 507
Erlich, Paul, 366
Error
  artificial intelligence/computation, 28
  Bayesianism, 41, 56
  experimental method, 272–273
  induction, problem of, 382
  Kuhn's scientific revolutions, 426
  normal/Gaussian law of, 56
  parsimony, 535
  statistics
    Bayesian advances and controversy, 810
    confidence interval estimation procedures, 806
    error probability philosophy, 803
    error probability principle versus likelihood principles, 807–809
    Neyman-Pearson (NP) tests, 804–805
    reforms within, 811
    severity assessments, 811–812
Essay (Bayes), 41, 46, 52
Essay Concerning Human Understanding (Locke), 236
Essays in the Logic of Mathematical Discovery (Lakatos), 434
Essential consciousness, 158–159, 160
Essentialism (typology), species, 795
Essential Tension, The (Kuhn), 419, 426, 429
Estimation/degree of confirmation, explicanda/explicata pairs, 288
Estimator surrogates, conservation biology, 166
Ethics
  biological science, 69
  chemistry, philosophy of, 106
  demarcation problem, 189
  determinism, 197
  economics, 227–228
  experimental method, 274
  game theory, 507
  genetics/eugenics, 334
  neurobiology, 521
  physicalism, 558
  Reichenbach and, 704, 711
  Russell and, 720
  Russell-Einstein Manifesto, 716
  Schlick and, 726, 727, 728
  social sciences, 782, 784
  supervenience, 816
  Vienna Circle and, 860
Ethology
  classical, 263, 264–265
  innate/acquired distinction, 395, 398

Ethos, logical empiricism versus logical positivism, 458
Etiological function
  biology, philosophy of, 69
  function, 317–320
Etymologies (Isidore of Seville), 843
Euclid, 117, 357, 360, 766
Euclidean geometry, 116
  axiomatic method, 358
  axiomatization of, 357
  classical mechanics, 117, 121
  conventionalism, 169–170
  Hertz and, 120–121
  Hilbert, 360, 361
  instrumentalism, 401
  non-Euclidean geometry versus, 357
  Poincaré and, 568, 569, 570
  pseudo-Euclidean space, 790
  quantum mechanics, 653
  Reichenbach and, 705
  space-time, 788, 790
  time, 831
  underdetermination of theories, 841
Euclidean theory, Lakatos and, 438
Eudaimonist ethics, Schlick, 727
Eugenics, 334
Eukaryotic organisms, 335
Euler, Leonhard, 119
Euler diagrams, 867
Euler-Lagrange equation, second-order, 120
Euler solutions, three-body problem, 121–122
Evaluation
  cognitive science domain-specifying assumptions, 125
  ecological models, 221–222
Evarts, E.V., 515
Events, causality
  counterfactual accounts, 99
  ontological categories, 90
Everett, Hugh M., 205
Everett's relative state interpretation, quantum mechanics, 658
Evidence
  abduction, 1
  chemistry, philosophy of, 103
  confirmation theory, Bayesian, 148–150
  empiricism, 240–241, 242
  explanation, 281
  generative grammar, conceptual issues in, 110–111
  Kuhn and, 430
  likelihood approach to evaluation of, 149
  observation, 526–527
  problem of old evidence, 150, 590
  realism, 688
  Reichenbach and, 705
  research programs, 712
  verisimilitude, 857
  visual representation, 867–869, 870
Evidential decision theory, 183
Evolution, 251–257, 480
  adaptation and adaptationism, 3–7
  biological information, 64
  biology, philosophy of, 69, 70–71
  Chomsky and, 112
  current status, 254
  feminist philosophy, 298
  genetics, 332, 333, 335, 336, 338, 339
  heritability, 352–355
  history of evolutionary theory, 251–254
  individuality
    species as individuals, 376–377
    units of evolution, 378
  innate/acquired distinction, 395
  Mach, 468
  methodological individualism, 479
  molecular biology, 480, 481, 487
  natural selection, 497
  philosophical issues, 254–256
  population genetics, 578–585
  prediction, 596
  psychology, cognitive architecture, 617
  scientific revolutions and, 763
  species, 801–802
    reproductive isolation, 797–798
    units of, 796
  theory structure, 825
  units of
    individuals, 378
    species, 796
  unity and disunity of science, 846
Evolution, adaptive, 264
Evolution, dynamical (classical mechanics), 120
  determinism, 200
  dynamical systems theory, 122
Evolutionary altruism, 8–9
Evolutionary biology
  adaptation and adaptationism, 5
  altruism, 8–9
  approximation, 24
  biological information, 64
  molecular biology, 481
  molecular methods, 480
  parsimony, 532
  philosophy of biology, 70–71
Evolutionary drift, *see* Drift, evolutionary
Evolutionary epistemology, 256, 257–263
  biological science, 69
  ontogenetic projects
    evolution of epistemic mechanisms (EEM), 259
    evolution of epistemic theories (EET), 261–262
  phylogenetic projects
    evolution of epistemic mechanisms (EEM), 258–259
    evolution of epistemic theories (EET), 259–261
  Popper and, 575–576
  scientific change, 730
  scientific progress, 753
  scientific revolutions, 757, 763
  and tradition, 262
  Vienna Circle, 861
Evolutionary game theory, 328–329
Evolutionary models of scientific change, 750–751, 753
Evolutionary psychology, 256, 263–268
  alternatives to, 267
  biology, philosophy of, 69
  and cognitive science, 265–266
  function, 319–320
  language acquisition, 112
  massive modularity thesis, 266
  monomorphic mind thesis, 266–267
  psychology, cognitive architecture, 617
  social sciences, 784
Evolutionary species concept, 798

Evolutionary synthesis, 253
Evolutionary theory
  determinism, 197
  fitness, 310–315
  language as biological function, Chomsky and, 109
  natural selection, 497
Evolution of epistemic mechanisms (EEM), 258–259
Evolution of epistemic theories (EET), 258, 259–262
Exactness condition, explication, 289
Exadaptation, 4
Examples, Kuhn's disciplinary matrix, elements of, 421
Exchange models, economics, 508
Excluded middle, 174, 721
Exemplars, Kuhn's disciplinary matrix, elements of, 421
Existence
  intentionality and, 406
  phenomenalism, 553
Exobiology, 38
Exons, 337, 486
Expanding economy model (von Neumann), 507–508
Expected utility
  confirmation theory, decision theoretical approach, 148
  decision theory, 181–188
  probability, 605
  Ramsey and, 679
Experience
  constructionism, 80–81
  empiricism, 235, 523
    Hume and, 236
    Locke and, 236
  Hanson, on observation, 344–345
  induction, problem of, 379
  mathematical analysis versus, 116
  perceptual, 545
  protocol sentences, 610, 611, 612–613
  unity and disunity of science, 844
Experience and Prediction (Reichenbach), 427, 683
Experiment, 268–275
  action and production, philosophical implications, 270–271
  behaviorism, 61
  chemistry, philosophy of, 102, 103, 105–106
  crucial (*experimentum crucis*), 209
  curve fitting, 533–534
  ecological communities, 220
  heritability, 355
  history, 268–269
  Kuhn on language, 428
  Kuhn's disciplinary matrix, elements of, 421
  Lakatos and, 436
  mechanisms, discovery of, 475–477
  metaphysics and, 429
  Michelson-Morley, 789
  Nagel on, 493
  neurobiology, *see* Neuroscience, neurobiology
  observation, 526–527
  particle physics, 538
  philosophy of scientific experimentation, 269–270
  Poincaré and, 570
  research programs, 712–715
  scientific computing and, 508
  scientific domains, 733
  scientific experience, role of, 274
  scientific models, 745
  in social and human sciences, 274–275
  social constructionism, 777
  statistics, 804
  technology-experimental science relationship, 271–272
  theory and, 272–273
  underdetermination of theories, 841
Experimental error, 272–273
Experimental-mathematical thermodynamics, 268
Expert systems, scientific change, 731
Explaining Science (Giere), 430
Explanandum/explanans, 276, 277
  prediction, 590
  realism, 691
Explanandum partition, 281
Explanation, 275–287
  abduction, 1, 2
  alternatives to covering-law models, 280–285
    causal mechanical model, 281–282
    explanatory unification approach, 283–285
    pragmatic accounts, 282–283
    statistical relevance model, 280–281
  analyticity, Harman's case against, 15
  astronomy, 36
  causality, 95
  chemistry, philosophy of, 102, 103, 105
  classical mechanics, 115
  covering-law models, 276–280
    deductive-nomological (D-N), 276–279
    deductive-statistical (D-S), 276, 278–279
    inductive-statistical (I-S), 276, 279–280
  ecological models, 222–223
  economics, 225
  Feyerabend, 305
  function, 316
  functional and teleological, 285–286
  Hempel, 347–348, 351
  inference to the best explanation (IBE), 240, 441
  kinetic theory, 418–419
  Kuhn, paradigms, 420–421
  of laws, 276, 278–279
  linguistics
    American Structuralism, 107
    descriptive adequacy versus explanatory adequacy, 108, 109
  logical empiricism, 463
  logical empiricism versus logical positivism, 458
  mechanisms, 469–477
  mechanistic, 473
  methodological individualism, 478, 479
  molecular biology, 484
  Nagel on, 493
  natural selection, 501–502
  neurobiology, 514
  physical sciences, philosophy of, 556
  Popper and, 577
  prediction, 587, 588, 592, 597
  realism, 687, 688, 689, 691–692
  reductionism, 697, 698
  reductive, 285
  research programs, 715
  scientific domains, 733, 735, 736
  scientific metaphors, 738
  scientific models, 740, 748
  scientific progress, 750

Explanation (*Continued*)
  social, varieties of, 780–783
    cultural materialism, 782
    functional analysis, 781–782
    interpretive methodologies, 782–783
    rational choice program, 780–781
  social sciences, 780
  supervenience, 816–817
  unity and disunity of science, 845, 846
Explanation, Reduction, and Empiricism (Feyerabend), 304–305
Explanatory gap, 549
Explanatory store, 284
Explanatory unification approach, 283–285
Explicanda/explicata pairs, 288
Explicandum, explication, 287–288
Explication, 287–294
  analyticity, 13
  and axiomatics, 289–290
  Carnap on, 287–289
  Quine, objections of, 293
  rational reconstruction, 683–684
  Strawson, objections of, 290–292
Explicit approximation, 26
Explicit definition method, 679
Exponential decay law, approximation, 25–26
Extension, explicanda/explicata pairs, 288
Extension rule, Reichenbach, 708
Extensions, set theory predicate-extensions, 504–505
Extensive form, game theory, 324
Externalist theories
  Chomsky and, 111–112
  epistemology, 245, 248–249
External world
  language faculty and, 446–448
  Russell, 82
Extra-logical knowledge, analyticity, 12

F

Facts
  causality, ontological categories, 90
  Hanson on, 345–346
  linguistic, 111
  Reichenbach and, 705–706
  universals of, 680
Factual content criterion, 132
Factual (synthetic) statements
  Duhem thesis, 209
  Reichenbach and, 706
Fact/value dichotomy, Putnam and, 620
Fagerstrom, T., 313
Fair, David, 91, 99
Faith, D.P., 164
Falconer, D.S., 354, 355
Falk, Raphael
  biology, philosophy of, 72
  genetics, 331, 337
Fallacies of rejection, large N problem, 812
Fallacy of non-statistically significant results, 809
Fallibilism, 83
  demarcation problem, 190, 196
  empiricism, 238
  Lakatos and, 437
  scientific progress, 750
Falsehood, Lakatos and, 438
False models
  ecology, 221
  population genetics, 584
False predictions, 221
Falsification/falsifiability/falsificationism, *see also* Cognitive significance
  corroboration, 177–179
  demarcation problem, 191–192; *see also* Demarcation, problem of
  Duhem thesis and, 208–209
  economics, 226–227
  evolutionary epistemology, 260
  induction, problem of, 382
  Kuhn's scientific revolutions, 424–425
  parsimony, 535–536
  Poincaré and, 570
  Popper and, 574–575, 577; *see also* Popper, Karl R.
  prediction, 591
  Quine and, 668
  rational reconstruction, 684–685
  unity and disunity of science, 845
  verifiability, 852
  verisimilitude, 855–856
Falsity, confirmation theory, Bayesian, 148
Faraday, Michael
  and locality/local action, 452
  scientific domains, 736
  scientific revolutions, 755, 756
Farewell to Reason (Feyerabend), 305
Fausto-Sterling, Anne, 298, 299
Faye, J., 141, 142
Fayyad, U.M., 731
Feedback mechanisms, 471
  ecological communities, 216–217
  molecular biology, 483
  psychology, cognitive architecture, 616
Feedforward mechanisms, 470–471
Feedforward networks, 151–152
Feenberg, Andrew, 269
Feigenbaum, E.A., 27, 731
Feigl, Herbert, 82
  cognitive significance, 132
  induction, problem of, 381
  logical empiricism, 458, 459, 461
  logical positivism, 458, 459, 460
  physicalism, 558–559
  Schlick and, 726
  unity of science movement, 848, 849
  verifiability, 852
  Vienna Circle, 858, 859, 860
Feldman, J., 27
Feldman, M.W., 355
Felleman, D.J., 517
Feller, William, 609
Felton, David L., 368
Feltzer, James H., 177
Feminist philosophy of science, 295–303
  directions, 303
  immune self, 365
  internal critiques, 297–299
  neurobiology, 521
  research questions, 296–297
  scientific change, 730
  scientific metaphors, 739

  social constructionism, 778
  themes, 299–303
Fenner, Frank, 366
Fermat, P., 599, 602
Fermi, Enrico, 538
Fermilab, 538
Fermions, 540
Fermi theory of radioactive decay, 538, 543, 544
Fernández-Morán, Humberto, 866
Ferrier, S., 166
Festa, R., 387, 389, 856
Fetzer, James H.
  corroboration, 178, 179
  probability theory, 607
Février, Paulette, 633
Feyerabend, Paul K., 304–310
  Against Method, 306–307
  biography, 304–305
  demarcation problem, 190, 191
  empiricism, 305–306
  Hanson and, 344–345
  incommensurability, 370–373
  Kuhn and, 429, 760
  Lakatos and, 434, 438
  observation, 527
  Popper and, 572
  posthumous works, 309
  protocol sentences, 612
  relativism, 308–309
  research programs, 714
  scientific change, 730
  scientific domains, 734
  scientific metaphors, 738
  scientific revolutions, 759, 760, 763
  theoretical pluralism, 307–308
  unity and disunity of science, 845
Feynman, Richard, 541
  probability theory, 600
  quantum field theory, 631
Feynman diagrams, quantum field theory, 631
Feynman path integral approach, 120
Feynman paths, quantum logic, 636–637, 642
Fictional objects, scientific models, 743–744
Fiducial intervals, 806–807
Fieberg, J., 167
Field, Hartry, 565
Field aspect, quantum field theory, 631
Field equations, 24, 359; *see also* Einstein's field equations
Field of experience, scientific concepts, 844
Field operators, quantum field theory, 630
Field quanta, quantum field theory, 631
Fields
  probability theory, 600
  quantum field theory, 631
Field theory
  and classical mechanics, 115
  generalized, 359
  locality/nonlocality, 457
  quantum field theory, 629–630
  universal, 359
Field theory model, 544
  particle physics, 541
  space-time, 794
File-drawer problem, statistics, 809
Filling instructions, explanation, 284
Final causes (Aristotle), 352, 780; *see also* Teleology
Fine, Arthur, 454
  empiricism, 242–243
  instrumentalism, 400, 404
  physical sciences, philosophy of, 555
  quantum measurement problem, 647
  realism, 692
Fine, Terrence, 602
Fine-tuning, anthropic principle, 21–23
Finite point of view, Hilbert, 360–361
Finitism, Ramsey and, 674
Finsen, S., 313
First causes, 189
First digits, law of, 58
First-order equations, second-order Euler-Lagrange equation reformulation, 120
First-order functional calculus (first-order logic; predicate first-order calculus), 13
  Hempel, 350
  quantum logic, 633, 635–636, 641
First signal geometry, Reichenbach and, 705
Fisher, Ronald A., 366
  adaptation and adaptationism, 4–5
  Bayesianism, 46, 48, 57
  heritability, 354
  parsimony, 533
  population genetics, 579, 581, 582, 583
  prediction, 597
  statistics
    confidence interval criticisms, fiducial intervals, 806–807
    Fisher and Neyman-Pearson debates, 806
    Fisherian simple significance tests, 804
    inductive behavior philosophy, 805
    Neyman-Pearson (NP) tests, 804–805
    p-values and Bayesian posteriors, 813
Fisherian tests
  Bayesian advances and controversy, 810
  simple significance tests, 804
Fitch, W.M., 254
Fitelson, Brandon
  confirmation theory, 149
  inductive logic, 392, 393
Fitness, 310–315
  adaptation and adaptationism, 3, 5
  altruism, 8–9
  biology, philosophy of, 69
  ensemble properties and population biology, 310–311
  evolutionary biology, philosophy of, 70
  game theory, evolutionary, 328
  and immunology, 365
  inclusive, 9, 311, 318
  innate/acquired distinction, 395
  models, ecological fitness, and evolutionary drift, 313–315
  natural selection, 498–499, 500; *see also* Natural selection
  population genetics, 581–582
  propensity interpretation of, 312–313, 500
  scientific theories, Popper and, 575–576
  solution to design problem, 311–312
Fixed-point lemma, 85
Fixed-point theorem of Brouwer, 508
Flatland, 829, 831–832
Flax, Jane, 300

Fleck, Ludwik, 511
  scientific revolutions, 761
  scientific style, 766
Fluid mechanics, approximation, 24
Fock representation, quantum field theory, 506
Fock space, quantum field theory, 630, 631
Fodor, Jerry, 4
  cognitive science, 128, 129
  connectionism, 153
  evolutionary psychology, 265, 266
  explanation, criticism of Nagel's reductive model, 285
  innate/acquired distinction, nativism, 396, 397, 398
  intentionality, 408
  psychology, cognitive architecture, 616
  psychology, philosophy of, 617
  Putnam and, 626
  Quine and, 660
  unity and disunity of science, 845
Foley, R., 667
Foliation, space-time, 790, 791
Folina, Janet, 569, 867
Folk psychology, 615, 781
Folks, L., 804, 811
Folse, H.J., 141
Forces
  causality, 97–98
  classical mechanics, 119
Ford, Joseph, 204
Ford, K.M., 28
Form, individuality, criteria for, 377
Formal analysis, classical mechanics, 115
Formal causes (Aristotle), 352
Formal genetics, 333–334
Formalist approach
  economic theories, 227
  Hilbert, 357
  Lakatos and, 436
Formalization, proof theory, 358
Formal logic
  Hempel, 350
  Lakatos and, 434, 436
  Reichenbach and, 709
Formal theory of conditional probability, 386
Formatting, visual representation, 869–870
Forster, Malcolm
  confirmation theory, 149
  parsimony, 535, 536, 537
  scientific models, 743
Foster, J.M., 103
Foster, Kenneth, 195
Foucault, Michel, 364, 757, 758
Foundationalist theories
  empiricism, 238
  epistemology, 245–246
  Feyerabend and, 304–305
  Neurath and, 513
  protocol sentences, 610, 611, 612
  verifiability, 851
  Vienna Circle and, 859
  von Neumann and, 504–505, 508
Foundational problems, social sciences, 783
Foundations of Arithmetic (Frege), 13
Foundations of Empirical Knowledge, The (Ayer), 38
Foundations of Geometry (Hilbert), 174, 359
Foundations of Mathematics, The (Ramsey), 672
Foundations of physical theories, 554–555
Foundations of Probability (Kolmogorov), 599
Four-dimensional geometry
  Reichenbach and, 705
  relativistic space-time, 205
  space-time, 786, 829, 831
Four Dissertations (Price), 47
Fourier coefficients, quantum field theory, 630
Fourier transform, quantum field theory, 629
Four-momentum observable, 457
Fox, T.D., 483
Fraenkel, Abraham, 718
Frame invariance, decision theory, 184
Frame problem, computation, 28
Frames of reference, space-time, *see* Inertial frames
Frank, Josef, 858, 861
Frank, Philipp
  Hahn and, 341
  logical empiricism, 461
  logical positivism, 460
  Mach and, 467
  Neurath and, 512
  Schlick and, 726
  unity and disunity of science, 845
  unity of science movement, 848
  Vienna Circle, 858–859
Franklin, Allan, 269, 270, 272
Franklin, Stan., 153
Franks, B., 127
Fraser, S., 354
Fraunhofer, Joseph, 35
Freeman, M., 266
Free will, 63
Frege, Gottlob, 84, 85, 289, 570
  analyticity, 13
  Carnap, 79
  empiricism, 237
  Hilbert and, 357, 360
  logical positivism, 460
  Russell and, 715, 716, 717–718
  unity and disunity of science, 844
  verifiability, 852
Frege-Russell thesis of logicism, 80
French, Steven
  quantum field theory, 631
  scientific models, 741
  theory structure, 827
    semantic view, 825, 826
    syntactic (received) view, 824
Frenkel-Brunswik, Else, 858
Frenz, Horst, 716
Frequency-dependent fitness, game theory, 328
Frequency formula, Bohr, 651
Frequentism
  probability, 601–602, 606, 607
  statistics, 803, 807–809
Frequentists, 53
Fresnel, A.J., 755
Fresnel's wave theory, 45, 590, 689, 755
Freud, Sigmund, 759
Freudian/psychoanalytic theory
  demarcation problem, 191, 192, 574
  Nagel and, 492
  Popper and, 574, 586
  prediction, 586
  scientific revolutions, 759

Friedman, D., 804
Friedman, Michael, 40
  analyticity, 13
  Carnap, 80, 82
  Duhem thesis, 209–210
  empiricism, 238, 241
  explanation, 283
  kinetic theory, 418
  unity and disunity of science, 844
Friedman, Milton, 226
Frigg, Roman, 741, 744
From Copernicus to Einstein (Reichenbach), 704
Frozen accident (Crick), 533
Fruit fly (*Drosophila melanogaster*), 331, 332, 333, 355, 486
F test, ANOVA, 804
Fuchs, Christopher, 454
Function, 315–322
  adaptation and adaptationism, 4
  biology, philosophy of, 69
  directedness, normativity, and value, 320–321
  genetics, 332
  linguistics, Chomsky, 112
  and mind, 319–320
  molecular biology, junk DNA, 482
  naturalistic analysis, 315–316
  naturalizing, 316–319
  normative concepts, biology and, 69
  and physicalism, 559
  prediction, 597
  probability, 599, 600
Functional analysis
  function, 317, 318
  social sciences, 781–782
Functional decomposition, 473
Functional explanation, 276, 285–286
Functional imaging of brain, 616, 617
Functional individuals, 378
Functionalism
  perception, 548, 549
  Ramsey and, 676
  social constructionism, 776
Functional magnetic resonance imaging, 616, 617
Functional organization
  individuality, criteria for, 377
  structural heterogeneity, 377
Functional psychology, behaviorism and, 61
Function-and-argument structure, symbolic models, 153
Function of General Laws in History (Hempel), 463
Fundamental theorem of natural selection, 581
  population genetics, 581
Fundamental theory
  particle physics, 538
  physical sciences, philosophy of, 555
  unity and disunity of science, 846
Funkenstein, Amos, 757
Future-directed space-time curve, 205
Futuristically deterministic theory, 199

G

Gabelbarkheitssatz, 289–290
Gagen, M.J., 487
Gaifman, Haim, 606
Gaito, J., 809
Galavotti, Maria-Carla, 675
Galilean idealizations, 742
Galilean relativity principle, 787, 788
Galilean space-time, Hilbert and, 359
Galilean transformations, 788, 789
Galilei, Galileo, 26
  astronomy, 34
  causality, 92–93
  classical mechanics, 115
  demarcation problem, 189
  experimental method, 268, 269
  Feyerabend on, 307
  Hanson and, 346
  Kuhn's scientific revolutions, 421, 422, 424, 428
  mechanics, 116
  perception, 546
  scientific revolutions, 754, 756, 757
  scientific style, 766
  unity and disunity of science, 843, 844
Galison, Peter, 269, 542
  physical sciences, philosophy of, 557
  scientific style, 767
  unity and disunity of science, 845, 846, 847
  visual representation, 869
Gall, F.J., 616
Galton, Francis, 334, 353, 354
Galvani, Luigi, 515
Gambling, *see* Betting/wager model
Game of Life, 508
Games and Decisions (Luce and Raiffa), 323
Gametes, 331
Game theory, 323–329
  altruism, 10–11
  Bayesianism, 41
  cooperative games, 328
  economic rationality, 227
  evolutionary, 328–329
  evolutionary psychology, 264
  methodological individualism, 479
  Nash equilibrium, refinements of, 327–328
  non-cooperative games, 325–326
  representations of game, 324–325
  social sciences, 781
  unity and disunity of science, 845
  utility, theory of, 323–324
  von Neumann and, 503, 504, 507–508
Game tree, game theory, 324
Gamma-ray telescopes, 36–37
Gamow, George, 36
Gandy, R.O., 834, 837
Gärdenfors, P., AGM model, 249–250
Gardner, H., 114, 123
Gardner, Michael
  kinetic theory, 418
  prediction, 589, 591
Garrod, Archibald, 334
Garzon, Max, 153
Gases/gas laws, kinetic theory, 415–419
Gassendi, Pierre, 353
Gauge invariance, quantum field theory, 632
Gauge theories, quantum field theory, 632
Gauss, Carl Friedrich, 119
  Poincaré and, 568
  principle of least constraints, 120

Gaussian-chain model of polymers, 740
Gaussian curve, kinetic theory, 416
Gaussian law of errors, 56
Gauthier, David, 328
Gavroglu, K., 102
Gazdar, G., 445, 449
Geach, Peter, 767
Geertz, Clifford, 782
Gehring, W., 481
Gelfand-Naimark-Segal (GNS) construction, 631
Gelinas, R., 483
Gell-Mann, Murray, 541
Gender, *see also* Feminist philosophy of science
  Kuhn, 430
  Turing and, 836
Gender differences in brain, 521
Genealogy of the fact, 777
General causation, 90–91
General causes, 90
Generalist accounts
  causality, 95–98
    Bayesian networks and causal models, 96–97
    causal powers, capacities, universals, forces, 97–98
    probabilistic relevance accounts, 95–96
    singular versus general, 90–91
Generality problem, epistemology, 248
Generalizations, *see also* Laws of nature
  approximation, 24–25
  laws of nature, 441–442
  linguistics, theory of, 107
  neurobiology, 519–520
  scientific progress, 750
  universal, 276, 278
Generalizations versus laws, statistical, 495–496
Generalized field theory, 359
General laws
  prediction, 588
  scientific progress, 750
General relativity
  approximate solution to field equations, 24
  Bayesianism, 45
  determinism, 205
  Euclidean geometry, 361
  irreversibility, 412–413
  and locality/local action, 451–452
  old evidence problem, 150
  particle physics, 538–539
  physical sciences, philosophy of, 554
  reductionism, 698
  Reichenbach and, 704
  space-time, 791–794
  and time, 830–831
  underdetermination of theories, 841
General Theory of Knowledge (Schlick), 174, 175, 726, 727, 861
General transport (Boltzmann) equation, 416, 417
Generative entrenchment, 399
Generative grammar
  conceptual issues in, 109–114
    inductive versus abductive learning, 112–113
    intrinsic versus relational properties in linguistics, 111–112
    limits of science, 113
    mind/body problem and question of physicalism, 113–114
    object of study, 110
    science-forming faculty, 113
    against teleological explanation in linguistics, 112
    underdetermination of theory by evidence, 110–111
  development and evolution of, 107–109
  linguistics, philosophy of, 444–450
  psychology, cognitive architecture, 616
Genes
  biological information, 65
  gene maps, 353
  horizontal transfer, 486
  hypothesis, 331
  knock-out experiments, 266
  molecular biology, 483
Genetic algorithms, 731
Genetic code, 335–336
  molecular biology, 481, 483
  parsimony, 532
Genetic drift, 500–501; *see also* Drift, evolutionary
Genetic engineering, 337, 339
Genetic epistemology, 261
Genetic individuals, 378
Genetic information, *see* Biological information; Molecular biology
Genetic recombination, 497, 579
Genetics, 330–339
  adaptation and adaptationism, 7
  altruism, 8
  biological information, 64–68
  biology, philosophy of, 69
  classical, 330–333
  in context, 337–339
  determinism, 197
  and evolution, 254–255
  evolutionary biology, philosophy of, 70
  evolutionary psychology, monomorphic mind thesis, 266
  formal, 333–334
  heritability, 352–355
  individuality, 377–378
  innate/acquired distinction, 399
  innateness, 398–399
  intragenomic conflict, 498
  material, 334–335
  methodological individualism, 479
  molecular/molecular biology, 335–337, 480–483
    genomics, 485–486
    information, 484–485
    philosophy of, 72
  parsimony, 532–533
  population, 310–311, 578–585; *see also* Population genetics
  reductionism, 696, 698, 699, 701
  species
    gene flow, 796, 797, 798, 800
    phenetic, 796
  unity and disunity of science, 846
Genetic variance, 354
Genome
  genetics, 334
  molecular biology, 481
Genomics
  genetics, 337
  human genome project, 485–486

Genotypes
  heritability, 354
  population, 334
  population genetics, 579, 580
Genotypic fitness, 311
Geodesics, space-time, 788, 792
Geographic variation, heredity and, 353
Geology
  scientific revolutions, 755
  unity and disunity of science, 845
  visual representation, 867
Geometric approach to relativity theory, Nagel on, 493
Geometric conventionalism, 169–170, 174
  Hilbert on, 360
  Poincaré, 568–569
Geometric notation, Hanson and, 346
Geometric optics, mechanics and, 120
Geometrization of mechanics, Hertz and, 120–121
Geometry, 80
  axiomatic-deductive method, 268
  chemistry, philosophy of, 103
  classical mechanics, 117
  continuous, 506
  Euclidean, 116
  Hilbert
    finite point of view and recursive epistemology, 360–361
    foundations of, 357–358
    multiplicity of, 360–361
  instrumentalism, 401
  Nagel on, 493
  Poincaré and, 568–569
  Reichenbach and, 705, 706
  theory of relativity of, 706
  underdetermination of, 512
  underdetermination of theories, 841
  unity and disunity of science, 843
George, A., 445, 663
Gergonne, J.D., 169
Germaneness, 239
German organicism, emergence, 231, 232–233
German tradition
  experimental method, 269, 273, 274
  unity and disunity of science, 844
Gerner, K., 704, 705
Geroch, Robert
  determinism, 205–206, 207
  physical sciences, philosophy of, 555
Gestalt-like modes of scientific thought, 738
Gestalt psychology
  Carnap and, 82
  and emergentism, 232
  Hanson and, 345
  Kuhn and, 427–428, 430
  Mach and, 467
  Nagel on, 493
  Schlick and, 726
  unity and disunity of science, 844
  Vienna Circle and, 861
Gestalts
  scientific style, 766
  and world changes, 427–428
Gettier problem, epistemology, 244–245
Ghirardi, G.C., 655
Ghiselin, Michael
  biology, philosophy of, 71
  individuals, species as, 376
  species, 801
Gibbard, Allan, 183
Gibbins, Peter, 633
Gibbs, J.W., 415, 417
Gibbs, Leslie, 256
Gibbs entropy, 411
Gibson, J.J., 549–550
Giedymin, Jerzy, 169
Giere, Ronald N.
  Bayesianism, 43
  Hempel's (Ravens) paradox, 348
  induction, problem of, 382
  Kuhn and, 428, 430
  laws of nature, 443
  prediction, 589, 591
  probability theory, 607
  realism, 689
  scientific models, 741, 743
  theory structure, 824, 825, 826
  visual representation, 867–868
Gieryn, Thomas, 195, 196, 779
Gigerenzer, G., 810
Gilbert, F.S., 165
Gilbert, S., 233
Gillespie, G.H., 312
Gillespie, N., 499
Gillies, A.S., 250
Gillies, Donald., 607, 731
Gilligan, Carol, 298
Gilpin, Michael, 222
Ginzberg, Lev, 72, 73
Gisin, Nicolas, 456
Given, the, 523
Glaister, S., 387
Glashow, S., 542
Glauberman, N., 354
Gleason, A., 640
Gleason, Henry, 216
Gleasonian community, 216, 217–218
Gleason's representation theorem, 656
Glennan, Stuart
  explanation, 282
  mechanism, 469
  reductionism, 699, 700
Glimm, James, 508
Global accounts
  classical mechanics, 121
  parsimony, 532–534
Global entropy, 412
Global inertial frame, 791
Global supervenience
  physicalism, 564, 565
  supervenience, 819–820
Global time-slice, 206
Glock, Hans-Johann, 16
Gluons, 542
Glymour, Clark
  artificial intelligence, 29
  causality, 96
  confirmation theory, 147–148, 150
  instrumentalism, 403

Glymour, Clark (*Continued*)
  prediction, 590
  scientific change, 731
  statistics, 810
Goal-oriented causality, 90
Goals, function, 320
Godambe, V., 807
Gödel, Kurt, 357, 715
  Carnap and, 82, 84, 85
  conventionalism, 175–176
  determinism, space-time model, 205
  explication, 290
  genesis of proof theory, 358–359
  Hahn and, 341, 342, 343
  quantum logic, 641
  Quine and, 659
  Schlick and, 726
  time, 830
  Turing and, 833
  Vienna Circle, 858
Gödel-Lob (GL) logic, 769, 771
Gödel numbering, 508
Gödel's proof of Peano Arithmetic, 358–359
Gödel's theorem, 834, 836, 837
Godfrey-Smith, Peter
  adaptation and adaptationism, 5
  biological information, 65
  function, 318, 319, 320
Goethe, W., 353
Gold, Thomas, 36
Goldberg, S., 793
Goldfarb, Warren, 175
Goldschmidt, Richard, 253, 332
Goldstein, Sheldon, 411, 656
Goldstine, Herman H., 508
Gómez-Pompa, A.C., 163
Gomperz, Heinrich, 858
Good, I.J.
  Bayesianism, 52
  confirmation theory, 149, 150
  Hempel's (Ravens) paradox, 348
  inductive logic, 393
  Jeffrey-Good-Lindley paradox, 809
  Ramsey and, 679
Goodale, M.A., 548
Gooding, David, 269, 272
Goodman, Nelson
  Carnap, 81
  confirmation theory, 146
  empiricism, 239
  epistemology, 246
  explanation, 278
  Hempel, 349
  induction, problem of, 383
  laws of nature, 439
  logical empiricism, 463
  prediction, 587
  probability theory, 609
  visual representation, 864, 865
Goodness of fit, 533–534, 804
Goodnight, C.J., 582
Goodwin, Richard, 508
Gordon, H. Scott, 478, 784
Gotelli, Nicholas, 221
Gottlieb, G., 267
Gould, Stephen Jay, 4
  adaptation and adaptationism, 6
  Chomsky and, 112
  feminist philosophy, 299
  individuals, clonal populations as, 375
  natural selection, 497, 498
Gradualism, natural selection, 498–499
Graham, G., 618
Graham, Kim S., 156
Grammar
  generative, *see* Generative grammar
  linguistics, philosophy of, 445, 448
  Searle and, 770
  theory of, 111
  transformational, 107
Grant, Peter, 256
Grant, Robert E., 252
Graphs, 864
Grattan-Guiness, Ivor, 119, 672
Graves, Gary, 221
Graves, Leslie, 197
Gravitation/gravitational theory, 25, 279
  astronomy and, 35
  classical mechanics, 117–118
  generalized field theory, 359
  Hanson and, 346
  Kuhn's scientific revolutions, 423–424
  and locality/local action, 451
  methodological individualism, 479
  particle physics, 539
  quantum mechanical, 413
  reductionism, 698
  relativistic statistical-mechanical entropy and, 413
  scientific domains, 735
  space-time, 786, 788, 792, 794
Graviton, 543
Gray, Russell D.
  biological information, 66, 67
  developmental biology, philosophy of, 72
  innate/acquired distinction, 398, 399
Greek philosophers
  instrumentalism, 400
  unity and disunity of science, 843
Greeno, Richard, 280
Gregorian telescope, 34
Gregory, P.A., 663, 668
Grelling, Kurt, 673, 704
Gremer, Jennifer K., 254
Grice, H.P.
  analyticity, 15, 19
  Quine, 660
Grice, Paul, 408
Griesemer, J.R., 827
Griffiths, Paul E., 3
  biological information, 66, 67
  biology, philosophy of, 72, 73
  function, 318–319
  innate/acquired distinction, 398, 399
  molecular biology, 484
Griffiths, Robert, 454, 658
Gross, Paul, 778
Grosser, Morton, 118
Group adaptation, 8
Group behavior, Nagel and, 494
Group concept, Poincaré and, 569

Group rationality, 784
Group selection, 8, 70, 498
Group theory, 119
  and classical mechanics, 121
  geometric transformations, 170
Grue paradox, 146
Grünbaum, Adolf
  conventionalism, 170–174
  Reichenbach and, 706
Grundgesetze der Arithmetik (Frege), 718
Grundlagen der Geometrie (Hilbert), 357
Guesses, Lakatos and, 436–437
Guevara, S., 163
Gutnick, M.J., 515
Gutting, Gary, 758
Güzeldere, G., 618
G-value paradox, 486

H

Haack, S., 82, 633
Habermas, Jürgen, 269, 272, 274
Habit, mental, 679
Habits of inference, 679
Hacking, Ian, 269, 271, 273, 309
  anthropic principle, 22
  corroboration, 178
  parsimony, 533
  probability theory, 607
  rational reconstruction, 685
  realism, 686–687, 688, 691
  scientific change, 730
  scientific revolutions, 763
  scientific style, 767
  social constructionism, 775–776
  statistics, 807–808
  unity and disunity of science, 846–847
Haddad, Brent, 215
Hadrons, 539, 542
Haecciety (thisness), quantum field theory, 631
Hagg, R., 506
Hahn, Hans, 82, 83, 341–343
  conventionalism, 169, 175
  logical empiricism, 459
  logical positivism, 460
  Mach and, 467
  Neurath and, 512
  Quine and, 659
  Schlick and, 726, 728
  unity of science movement, 848
  Vienna Circle, 858, 859
Hahn, L.E., 667
Hahn-Neurath, Olga, 342, 858
Hajek, Alan
  inductive logic, 386
  probability theory, 600, 607
Hald, A., 41, 46, 47, 56
Haldane, J.B.S.
  evolution, 253
  population genetics, 579
Hall, A.R., 762
Hall, Bert, 866
Hall, Ned, 92
Hall, W., 59
Hallett, Michael, 438
Halley, Edmund, 147, 594–595
Halliday, T.R., 264
Halvarson, Hans, 457, 506
Hamadani, K., 487
Hamilton, James D., 593
Hamilton, William D.
  altruism, 8, 9
  fitness, 311
Hamilton, William Rowan, 118, 119
Hamiltonian mechanics, 119–121, 650
  classical mechanics, 121
  quantum conditions, 652
  quantum logic, 634, 637
  quantum mechanics, 656, 657
Hamilton-Jacobi equation, 120
Hamilton-Jacobi theory
  classical mechanics, 121
  quantum logic, 635
Hamilton's principal function, 635
Hammer, Eric, 867
Hammond, Peter, 182
Hands, D. Wade, 226
Hankinson, R.J., 92
Hanson, Norwood Russell, 344–347
  causes, 346
  facts, 345–346
  Kuhn and, 427–428, 429
  logic of discovery, 346–347
  observation, 344–345, 527
  protocol sentences, 612
  scientific change, 730
  scientific domains, 734
Hansson, S.O., 250
Haraway, Donna
  feminist philosophy, 298, 300, 302
  immunology metaphors, 365
  social constructionism, 778
Hard artificial intelligence, 30, 32
Hardcastle, Gary L., 40
  logical empiricism versus logical positivism, 458
  Vienna Circle, 858
Hardin, C.L., 618
Harding, Sandra G.
  feminist philosophy, 296, 297, 300, 301
  physical sciences, philosophy of, 556
  social constructionism, 778
Hardy, G.H., 333–334, 716
Hardy-Weinberg law, 333–334, 579–580
Harman, Gilbert
  abduction, 1
  analyticity, 14, 15–19
  Bayesianism, 53–54, 55
  cognitive science, 129
  epistemology, 247, 250
  intentionality, 409
Harper, William
  classical mechanics, 118
  decision theory, 183
  game theory, 325
  statistics, 807
Harré, Rom, 632, 738
Harris, Marvin, 782
Harsanyi, John, 41
Harstock, Nancy, 300

Hartigan, J.A., 59
Hartmann, Heidi, 298
Hartmann, Stephan, 745, 747
Hartshorne, C., 128
Harvey, Paul, 7
Harvey, William, 353
Hatto, Arthur, 754
Haugeland, John, 408, 472
Hausdorff dimension, 122
Hausman, Daniel M., 224, 225, 443
Hawking, S.W., 21
Hayward, R.W., 540
Heat theory, scientific revolutions, 755
Hebb, D.O., 521
Hebbian learning, 152, 521
Heckerman, D., 393
Hegel, G.W.F.
  Lakatos and, 434, 435–436, 437–438
  scientific revolutions, 758, 759
  theories of social change, 755
Hegelian views, rational reconstruction, 685
Hegemony, immunology, 365
Heidegger, Martin, 269, 766
Heidelberger, Michael, 269, 272, 273, 858
Heims, Steve J., 508
Heinzmann, Gerhard, 570
Heisenberg, Werner
  Mach and, 467
  particle physics, 539
  quantum field theory, 629
  quantum mechanics, 652–653
Heisenberg cut, 505
Heisenberg's quantum mechanics
  complementarity, 141–142
  quantum field theory, 630
  von Neumann and, 506
Helm, Georg, 119
Helmholtz, Hermann von, *see* von Helmholtz, Hermann
Hempel, Carl Gustav
  approximation, 25
  behaviorism, 62
  and Bridgman, 76
  causality, 94, 95
  cognitive significance, 83, 132, 133–134, 135–138
  confirmation theory, 145–147
  corroboration, 177
  demarcation problem, 191, 192
  Duhem thesis, 210
  emergence, 233
  explanation, 276, 283–284
    deductive-nomological (D-N) model, 277, 278–279
    functional explanations and D-N model, 286
    inductive-statistical (I-S) model, 279–280
    Railton and, 282
  function, 316, 317, 320
  induction, problem of, 382
  inductive logic, 385
  instrumentalism, 402–403
  intentionality, 406
  kinetic theory, 418
  Kuhn and, 430
  laws of nature, 439
  life and work, 347–352
    explanation, 347–348
    historical background, 348–349
    later work, 351
    problems and changes, 349–350
    theories and observations, 350–351
  logical empiricism, 461, 462–464
  observation, 523–524
  prediction, 587–588
  Ramsey and, 679
  Ravens paradox, 348–349
  realism, 695
  reductionism, 697
  scientific metaphors, 738
  theoretician's dilemma, 62
  theory structure, 822–824
  unity and disunity of science, 845
  unity of science movement, 848
  verifiability, 852, 853, 854
  Vienna Circle, 858, 860
Hempel's (Ravens) paradox
  confirmation theory, 145, 146, 148–149
  Hempel, 348–349
Hendry, R.F., 826, 827
Henkel, R., 808
Hennig, Willi, 798, 799, 801
Henry, John, 776
Hepp, Claus, 506
Hepp-Bell debate, 506
Heraclitus, 843
Herapath, John, 416
Heredity
  Darwin and, 253
  and evolution, 256
  population genetics, 578–585
  reductionism, 699
Heritability, 352–355
  current definitions and problems, 354–355
  historical development, 352–354
  innate/acquired distinction, 395, 398–399
  molecular biology, human genome project and, 485
  natural selection, 497
Herrnstein, R.J., 354
Herschel, Caroline, 35
Herschel, John, 190
Herschel, William, 35
Hertz, Heinrich, 120–121
  Hahn and, 341
  simplicity principle, 360
Hertz, Paul, 704
Hertzsprung, Ejnar, 36
Hertzsprung-Russell (H-R) diagrams, 36
Herzberg, Alexander, 704
Hesse, Mary, 429
  scientific metaphors, 737–738
  scientific models, 742
Heterological contradiction, 673
Heteropathic laws, 231, 232
Heteropsychological domains of objects, 80
Heterozygotes, 330
Heterozygote superiority, 497
Heuristically progressive programs, 714
Heuristic progress, 714
Heuristics
  artificial intelligence, 31
  computational, 28, 29

  epistemic significance of novelty, 591
  functional analysis, 317
  Hempel, 351
  Lakatos and, 434, 435, 436, 437
  reductionism, 696
  research programs, 713–714
  scientific change, 731
  scientific metaphors, 738
Heyes, C.M., 267
Hidden locality, 456
Hidden state, quantum mechanics, 656, 657
Hidden variable theory
  locality/nonlocality, 453–454, 456
  quantum mechanics, 655–656
  realism, 692–693
Hierarchical organization of mechanism, 472
Hierarchy, Poincaré and, 570
Higginbotham, J., 445
Higgs, A.J., 165
Hilbert, David, 715
  conventionalism, 174, 175
  explication, 288
  life and work, 356–361
    finite point of view and recursive epistemology, 360–361
    foundation of geometry and axiomatic method, 357–358
    new foundation of mathematics and genesis of proof theory, 358–359
    physics, contribution to, 359–360
  Ramsey and, 674, 675
  Reichenbach and, 704
  Russell and, 717, 718
  unity and disunity of science, 844
  and variational calculus, 119
  von Neumann and, 504, 505
Hilbert-Einstein equations, 359
Hilbert school, von Neumann and, 504
Hilbert's eta-operator, 86
Hilbert space
  quantum field theory, 630, 631
  quantum logic, 633, 637, 638, 639, 640, 641
  quantum mechanics, 652–653, 655, 656
  von Neumann and, 506
Hill, T.R., 58
Hilpinen, Risto, 87, 855, 856
Hinde, R.A., 264
Hinkley, D.V., 808
Hintikka, Jaakko
  analyticity, 12
  Bayesianism, 51, 52
  inductive logic, 389
  probability theory, 604
  quantum logic, 642
Hipparchus of Rhodes, 33
Historical determinism, Popper and, 572, 575
Historical elements, cognitive significance, 350
Historically deterministic theory, 199
Historicism
  Popper and, 575
  social constructionism, 776
History
  Nagel and, 492, 494–496
  Popper and, 574
  population genetics, 579–582
  quantum logic, 633
  scientific domains, 734–735
  scientific revolutions, 754, 758
    The Scientific Revolution, 754
    social and political contexts of, 755, 756
  scientific style, 765, 767
History, physical
  prediction, 595
  quantum logic, consistent histories approach, 642–643
  quantum mechanics, 658
History of Philosophy (Windelband), 765
History of philosophy, locality, 451
History of science
  rational reconstruction, 684–685
  scientific change, 730–731
  scientific metaphors, 738–739
  scientific revolutions, 755–758
History of Western Philosophy (Russell), 715, 722
Hitchcock, Christopher
  causality, 90
  explanation, 282
  scientific models, 745
Hobbes, Thomas, 268
  game theory, 323
  scientific revolutions, 757
Hochberg, Herbert
  phenomenalism, 552
  verifiability, 853, 854
Hochstrasser, D.F., 486
Hoddeson, L., 538, 539, 541, 542
Hodge, M.J.S., 500
Hodges, A., 835
Hodges, J.L., 43, 803
Hodges, John R., 156
Høffding, Harald, 142
Hoffmann, R., 103, 106
Hogarth, Mark, 207
Holcomb, H.H. III, 267
Holism
  Duhem thesis, 208–210
  epistemology, 247
  feminist philosophy, 301
  Neurath and, 512, 513
  protocol sentences, 612–613
  Quine and, 660, 662, 663
  relational, 234
  verificationism, adjustments to, 40
  Vienna Circle, 859–860
  semantic, 660
Holland, P., 600
Holmes, Oliver Wendell, 762
Holton, Gerald, 429
  scientific metaphors, 738
  scientific style, 766
  unity of science movement, 848
Holzkamp, K., 269
Homeobox/homeotic genes, 338, 481
Homeostatic buffering, genetics, 338
Homozygotes, 330
Hon, Giora, 272, 273
Hooker, Clifford A.
  complementarity, 141, 142, 143
  quantum logic, 633, 640
  reductionism, 697

Hooker, Clifford A. (*Continued*)
  statistics, 807
  underdetermination of theories, 840, 841
Hooker, Joseph D., 252
Hookway, Christopher, 239
Hoppes, D.D., 540
Horan, Barbara L., 197
Horgan, John, 753
Horgan, Terence, 564, 817
Horizon problem, 556
Horizontal gene transfer, 486
Horst, S.W., 130–131, 615
Horwich, Paul, 404, 591, 592
Hoskin, Michael, 32
Housekeeping genes, 486
Howson, Colin
  Dutch Book argument, 212, 593
  induction, problem of, 382
  Lakatos, 438
  parsimony, 534
  prediction, 593
  statistics, 810
How to Solve It (Po'lya), 433–434
HOX gene clusters, 486, 499
Hoyingen-Huene, Paul, 371, 760
Hoyle, Fred, 36
H-R (Hertzsprung-Russell) diagrams, 36
H-theorem, 416, 418
Hubbard, Ruth, 299
Hubble, Edwin, 36
Hubby, J.L., 582
Huber, L., 267
Huber, Peter, 195
Hudson, R.P., 540
Huggett, N.
  particle physics, 543, 544
  quantum field theory, 632
Huggins, William, 35
Hughes, Jon, 858
Hughes, R.I.G.
  Bayesianism, 53–54, 55
  scientific models, 744
Hughes, Thomas, 776
Hull, Clark B., 61, 63
Hull, David L., 71, 73
  biology, philosophy of, 69
  evolutionary epistemology, 258, 259, 260, 261, 262
  individuality, 375
  individuals, species as, 376
  molecular biology, 484
  scientific change, 730
  scientific progress, 753
  species, 801
  unity and disunity of science, 846
Human behavior, *see* Behavior
Human biology, *see* Biology; Cognitive science
Human cognition, *see* Cognition; Cognitive science
Human Genome Project, 485–486, 846
Humanism, scientific, Vienna Circle, 860–861
Human nature, evolutionary psychology, 263–267
Human populations, genetics, 334
Human Understanding (Toulmin), 260
Hume, David, 39
  abduction, 1, 2
  Bayesianism, 48
  causality, 90–91, 93–95
  demarcation problem, 190, 192
  empiricism, 236, 523
  evolution, 252
  game theory, 323
  induction, problem of, 46–47, 379–380, 381, 382
  inductive logic, 384
  locality/local action, 451
  observation, 523, 524
  perception, 545, 547
  phenomenalism, 553
  prediction, 586–587, 592, 594
  propositions and their testing, 524
  Quine and, 666
  Reichenbach and, 708–709
  verifiability, 851, 852
Humor theory, 364
Humphrey, N., 160
Humphreys, G., 160
Humphreys, Paul
  causality, 90, 96
  emergence, 234
  scientific models, 745
Hurley, P., 385, 386
Hurst, L.D., 486
Hurwitz, Adolph, 356
Husserl, Edmund, 80
  explication, 288
  Hilbert and, 360
  phenomenalism, 551, 552
  Reichenbach and, 704
  scientific style, 766
Hutchinson, George Evelyn, 216–217
Hutchinsonian community, 217
Hutchison, C., 483
Huxley, Thomas Henry, 499
Huygens, Christiaan, 34, 589
  mechanics, 116
  probability theory, 602
  space-time, 788
Hydrogen atom, energy levels, 651–652
Hylton, P., 659
Hypatia, 302
Hyperplane, space-like, 206, 642
Hypersurface
  absolute simultaneity, 787
  Cauchy surface, 830
  space-time, 790–791
Hypothesis: formation, testing
  abduction, 1, 2
  auxiliary, 209
  Bayesianism, 42
  confirmation theory, 144–150
  corroboration, 177–179
  Duhem thesis, 210
  experimental method, 270
  Feyerabend, *Against Method*, 306–307
  Hansen, logic of discovery, 346–347
  instrumentalism, 400–401
  neurobiology, 514
  parsimony, 531–538
  phenomenalism, 551–552, 553
  prediction, 589, 592
  protocol sentences, 612
  Quine and, 662, 666–668

- research programs, 714
- Russell and, 720
- scientific progress, 750–751
- scientific style, 766
- unity and disunity of science, 847
- visual representation and, 864, 869

Hypothesis space, 42, 59
Hypothetical constructs, genes as, 331
Hypotheticalists, demarcation problem, 190
Hypothetical realism, scientific progress, 753
Hypothetico-deductivism
- abduction, 2
- confirmation theory, 147
- demarcation problem, 190
- observation, 524

I

Iconic models, 741, 742, 827
Iconic realism, Reichenbach and, 706
Ideal gas, entropy of, 651
Ideal gas law, 417
Idealism
- phenomenalism, 551, 552–553
- Russell and, 717
- scientific style, 765
- social constructionism, 776
- verifiability, 851
- Vienna Circle and, 860

Idealism/realism dispute, Quine and, 660
Idealization, *see* Approximation
Idealized models, 740, 741–742
- unity and disunity of science, 846
- verisimilitude, 857
- visual representation, 866

Ideal physical theory, physicalism, 560–561
Idea of a Social Science (Winch), 782
Ideas
- innate/acquired distinction, 395–396
- verifiability, 851

Identity
- biological
  - epistemology of, 365
  - immune self, 365–369
- linguistic, 444–445
- physicalism, 562–563
  - disjunction option, 562–563
  - multiple realizability claim, 562
  - property identity claim, 562
  - quasi-eliminativist option, 563
  - trope identity option, 563
- quantum field theory, 631

Identity operator, 640
Ideology
- Feyerabend on, 308
- function, 321
- scientific change, 730
- social constructionism, 777
- social sciences, 784
- Vienna Circle and, 860–861

Idiolects, I versus E, 110
Ignorance criterion, Bayes, 46
Ignorance interpretation
- decoherence, 649
- improper mixtures, 647
- quantum probabilities, 648
- quantum states, 644–645

I-language, 445, 446, 447
Iliopoulos, J., 542
Ill-formed problem, decision theory, 182
Illusion
- argument from, 547
- perceptual psychology, 548

Imaging techniques, brain, 617
- psychology, cognitive architecture, 616

Immature field, cognitive science, 123
Immunology, 363–369
- historical antecedents, 364–365
- immune self
  - assaults on, 367–369
  - origins of, 365
  - twentieth century reconstructions of, 366–367
- molecular biology, 481

Impagliazzo, John, 508
Impartiality, sociology of scientific knowledge, 776
Imperfect information, game of, 324, 325
Implementation level
- artificial intelligence, 31
- Marr's computer model of mind, 547–548

Implication logical/L-implication, explicanda/explicata pairs, 288
Implicit approximation, 26
Impoverished models, 740
Impredicative definitions, Poincaré and, 570
Impredicative function, Ramsey, 674
Improper mixture, quantum measurement problem, 647
Inbreeding, 497
- population genetics, 579

Inclusive fitness, defined, 311
Incommensurability, 370–373
- decision theory, 185
- Feyerabend on, 309
- Kuhn on, 420, 429, 430
  - language, 428
  - scientific revolutions, 425
- scientific change, 730
- scientific revolutions, 759

Incompleteness theorems
- Carnap, 84, 85
- second, 175

Incorrigible approximation, 26
Incremental confirmation, 144–145
Independent segregation, law of (genetics), 330
Indeterminate value, Reichenbach's three-valued logic, 710–711
Indeterminism/indeterministic phenomena, *see also* Determinism; Quantum mechanics
- analyticity, 14
- causality, 90
- Chomsky versus Quine, 110–111
- Nagel on, 493
- quantum mechanics, 654
- scientific metaphors, 738

Index of inductive caution, 390
Indifference, principles of
- of Carnap, 390, 492
- of Jaynes, 59
- of Laplace, 52, 602

Indistinguishability, quantum field theory, 631
Individual(s)
  altruism, 8
  ecological community as, 216
  fitness, 311
  Kuhn's scientific revolutions, 427–428
  species as, 376–377, 801, 802
Individual comparison supervenience, 817–818
Individual differences, evolutionary psychology, 267
Individualism, methodological, 478–479
Individualism and Psychology (Burge), 111
Individualistic approaches, linguistics, 111
Individuality, 374–378; *see also* Species
  biological individuals, 374–375
  biological individuals, kinds of, 378
  classes versus individuals, 374
  criteria for, 377–378
  innate/acquired distinction, 395
  meiosis, cost of, 375–376
  methodological individualism, 479
  particulars and bare particulars, 374
  psychology, philosophy of, 615
  quantum field theory, 629, 631
  species as, 376–377, 801
  units of selection, 376
Individual preference, theory of, 227
Individual selectionists, 375
Induction, problem of, 379–383
  abduction, 1, 2
  Bayesianism, *see* Bayesianism
  confirmation theory, 348, 349, 382–383
  conventionalism, 175
  corroboration, 177–179
  demarcation problem, 190
  Hume, 46–47
  Popper and, 572, 573
  prediction, 586–587, 594
  realism, 694, 695
  Reichenbach and, 708–709
  scientific change, 729–732
  scientific style, 767
  statistics, 803
  twentieth century responses, 380–382
Induction, rule/principle of, 708
Inductive behavior philosophy, statistics, 805–806
Inductive inference/logical probability, explicanda/explicata pairs, 288
Inductive learning, generative grammar, 112–113
Inductive logic, 384–394
  Bayesianism, *see* Bayesianism
  Carnap, 86–87, 88
  causality, 95
  classical mechanics, Newton's laws, 117
  confirmation theory, 144–150
  demarcation problem, 192
  explanation, 278, 279, 348
  Galileo, 116
  historical overview, 384–385
  induction, problem of, 382
  Kuhn's scientific revolutions, 425
  logical empiricism, 463
  mathematical analysis of inductive reasoning, 49
  Nagel and, 492
  naive version and received view, 385–386
  prediction, 592–593
  probability, interpretation and role in, 386–391
    epistemic interpretations of probability, 386–387
    logical interpretation, 387–391
    mathematical theory of probability, 386
  probability theory, 603–604
  Ramsey and, 679
  received view, rethinking, 391–392
  Reichenbach rule of induction, 708
  relevance, 392–393
  relevance of relevance, 393
  Russell and, 719
  scientific change, 729–732
  scientific style, 767
  statistics
    Bayesian advances and controversy, 810
    confirmation theory, 803
    roles for probability in inference, 803
  terminology and machinery for, 385
  verisimilitude, 855
Inductive-logical theories of nondemonstrative inference, 384
Inductive simplicity, Reichenbach and, 707
Inductive-statistical (I-S) models, explanation, 276, 279–280
Inductive strength, 385
Inductive verification, 851
Inertia, space-time, 792, 794
Inertial frames
  space-time, 787–788, 789, 790, 791, 792, 793
  time, 830
Inference
  abduction, 1
  Bayesianism, 42
  corroboration, 177–179
  empiricism, 240, 241
  explanation, 279
  explicanda/explicata pairs, 288
  Hanson and, 346
  inductive-logical theories of nondemonstrative inference, 384
  laws of nature, 441
  Neyman-Pearson (NP) tools and, 807
  prediction, 592–593
  quantum logic, 635–636
  statistics
    Bayesian advances and controversy, 810
    roles for probability in, 803
Inference rules, explanation, 284
Inference to the best explanation (IBE), 240, 241, 441
Inferential habits, Ramsey, 679
Inferentialism, intentionality, 409
Infinite degrees of freedom, quantum field theory, 629
Infinite dimensional systems, Von Neumann's operator theoretic methods, 507
Infinite dimensional tensor product, 506
Infinity
  axiom of, 672, 674
  Ramsey, 674
Informal mathematics, Lakatos and, 434
Information
  Akaike information criterion (AIC), 537
  Bayesian information criterion (BIC), 535
  bioinformatics, 485
  biological, 64–68; *see also* Biological information

computation, 28–29
game theory, 324–325
Informational gene concept, problems of, 65–68
Information processing, von Neumann, 508
Information theory
biological information, 66
Kuhn's scientific revolutions, 425
von Neumann and, 508
Inheritance
Darwin's theory of, 253
genetics, 330
heritability, 352–355
innate/acquired distinction, 397–398
non-Mendelian, 486
Initial conditions, Newton on, 118
Innate/acquired distinction, 394–400
biological information, 66
biological notion of innateness, 398
genetic traits as innate traits, 398–399
immunology, 363
nativism, 396–398
recent work, 399
thoughts, origins/sources of, 395–396
Innateness, biology, philosophy of, 72
Innovations, scientific change, 731
Insect-plant resemblance, 4
Instantiation condition, induction, 382
Institute for the Unity of Science, 848
Institutionalism, social sciences, economics, 228
Institutions, social
social constructionism, 774–775
social sciences, 780
Instructions, computation, 28
Instrumentalism, 400–405
artificial intelligence, 32
economics, assumptions of, 226
epistemic, 403–405
intentionality, 409
language of science, 402–403
loci classici, 400–402
Nagel on, 493
parsimony, 537
prediction, 588
Putnam and, 621–622
social constructionism, 776
Instrumentation
chemistry, philosophy of, 102
experimental method, 272
Kuhn's disciplinary matrix, elements of, 421
neurobiology, 517–518
observation, 528
Integrable systems, KAM (Kolmogorov-Arnold-Moser) theorem, 122
Integrative models, psychology, cognitive architecture, 616
Intellectual role, economics, 225
Intelligence
cognitive science domain, 124
Laplace, 198
linguistic ability and, 396
psychology, cognitive architecture, 616
Intelligent behavior, artificial intelligence, 27, 29–30
Intelligent machinery, Turing and, 836–837
Intensionality/intensional notions
analyticity, 14
explicanda/explicata pairs, 288
intentionality versus, 406
scientific metaphors, 738
Intentional causation, 769
Intentionality, 405–410; *see also* Mind/body problem
Brentano's two theses, 407–408
causality, 90
cognitive science domain-specifying assumptions, 125
consciousness, 159, 160, 162
function, 315
linguistics, philosophy of, 447
and mentality, 409
paradoxes of, 406–407
perception, 545, 546
and physicalism, 559
prediction, 598
psychology, philosophy of, 614–615
reducing, 408–409
Searle, speech acts and, 768–770
social sciences, 783
speech acts and, 768–770
Intentional object, 406
Intentional verus nonintentional consciousness, 159, 160
Interacting fields, quantum field theory, 632
Interacting genes, 355
Interactional empiricism, 529
Interaction strength, ecological communities, 218
Interactors, evolutionary epistemology, 261
Interbreeding, potential, 797
Interdependent decision problems, game theory, 323
Interdisciplinary fields
unity and disunity of science, 845
unity of science movement, 848–849
Interest-relativity, Putnam and, 620
Interference, catastrophic, 156
Interfering factors, Duhem thesis, 210
Interfield theories, unity and disunity of science, 846
Interlevel reduction, 697, 699–701
Internalist theories, epistemology, 245
International Encyclopedia of Unifed Science (Neurath), 848, 859
International System of Typographic Picture Education (ISOTYPE), 511, 848
Interphenomena, quantum mechanics, 710
Interpolation, Akaike information criterion (AIC), 537
Interpretation
approximation, 25, 26
Feyerabend, *Against Method*, 307
Hanson, on observation, 345
physical sciences, philosophy of, 555, 557
probability
classical, 602–603
frequentism, 601–602
quantum mechanics, 654–655
social sciences, 782–783, 784
Interpretation of Cultures, The (Winch), 782
Inter-Scientific Discussion Group (Frank), 845, 848
Interspecific interactions, ecological communities, 216–217, 218, 222
Intersubjectivity
Carnap, 83
experimental method, 274
Neurath and, 513
observation sentences, 666–668

Intersubjectivity (*Continued*)
  physicalism, 558–559
  Poincaré and, 570
  protocol sentences, 610
  Quine and, 665–666
Intertheoretic connections, approximation, 25
Intertheoretic reduction, kinetic theory, 418
Intervening sequences, 486
Intervening variables, genetics, 331
Interventionism, irreversibility, 411
Intragenomic conflict, 498
Intralevel reduction, 697, 698–699
Intransitive consciousness, 159
Intrinsic intentionality, 408, 769
Intrinsicness
  conventionalism, 173, 174
  intrinsic versus relational properties in linguistics, 111–112
Introduction to Logic and Scientific Method, An (Cohen and Nagel), 492
Introduction to Mathematical Philosophy (Russell), 716
Introduction to Semantics (Carnap), 85
Introduction to the Study of the Experimental Medicine (Bernard), 269
Introns, 337, 486
Introspection, observation, 524
Intuition/intuitionism
  Euclidean space-time, 117
  Hahn and, 343
  Hilbert, 360
  mathematical
    Hilbert, 357
    Poincaré, 569–570
    Reichenbach, 707
  mechanistic versus ontic levels, 472
  prediction, 589
  Turing and, 836
Intuitive notions of probability, inductive logic, 384, 392–393
Intuitive space, 80
Invariances
  classical mechanics, 119
  complementarity, 143
  game theory, 328
  meaning, 845
  quantum field theory, 632
Invariants, theory of, 392
Invariant submanifolds, classical mechanics, 121
Inverse problem, Bayes's solution, 46
Inverse-square law, 118
  and locality/local action, 451
  research programs, 714
Inverted spectrum hypothesis, 161, 549
Ionizing radiation, particle physics, 538
Irrationalism, scientific change, 730
Irrationality
  Kuhn's scientific revolutions, 425
  Lakatos and, 434
  scientific change, 730
  scientific progress, 749
  Vienna Circle, 860
  Vienna Circle and, 859
Irreducibility, molecular biology, 71
Irrelevance, confirmation theory, 145, 146, 148; *see also* Relevance
Irrelevant conjunction, problem of, 147
Irreversibility, 410–413
  entropy, 412–413
  Hilbert, 359
  kinetic theory, 418–419
  phase space, 411–412
  physical sciences, philosophy of, 554
  probability, 607
  statistical explanation, 412
  statistical mechanics, 411
  time, asymmetry, 831–832
  von Neuman's theory of measurement, 505
Isham, C.J., 543, 642–643
Ishiguro, M., 536
Isidore of Seville, 843
Island biogeography, 164–165
Ismorphism, 827
Isomorphic symbols, 867
ISOTYPE (International System of Typographic Picture Education), 511, 848

J

Jablonka, Eva, 72
Jackson, Frank
  physicalism, 564
  psychology, philosophy of, 617, 618
Jackson, J., 161
Jacob, François, 65, 336, 481
Jacobi, Karl Gustav Jacob, 119
Jacoby, R., 354
Jacquard, A., 581
Jaeger, G., 234, 699
Jakobson, Roman, 444, 848
James, William, 257
Jammer, Max, 93, 633
Janich, Peter, 269, 270, 273
Janis, A., 830
Jansky, Karl, 36–37
Jarret, Jon, 455–456
Jarvie, Ian C., 782
Jasanoff, Sheila, 196
Jaynes, E.T.
  Bayesianism, 41–45, 48, 49, 52, 53, 55–59
    concentration theorem, 55
    principle of equivalence, 56
  probability, 606
  probability theory, 600
Jeans, James, 650
Jeffrey, Richard C,
  Bayesian confirmation theory, 148
  Carnap, 87
  decision theory, 183, 185
  Dutch Book argument, 213
  explanation, 280
  Hempel, 347
  Jeffrey-Good-Lindley paradox, 809
  probability, 605, 606
  Ramsey, 675, 679, 680
Jeffrey conditionalization, 213
Jeffrey-Good-Lindley paradox, 809, 812
Jeffreys, H., 813
  Bayesianism, 57, 59
  mathematical analysis of inductive reasoning, 49

  parsimony, 534, 535
  statistics, 805
Jeffrey's theory, 680
Jenkin, F., 578, 580
Jerne, Niels K., 366, 367, 369
Jevons, W.S., 45, 190
Joergenseon, Joergen, 848
Johannsen, Wilhelm, 330–331
Johnson, M., 739
Johnson, R.W., 54, 55
Johnson, W.E., 49, 51
  inductive logic, 387
  probability theory, 603
  Ramsey and, 671
Johnson's postulate, 51, 52
Jones, C.A., 767
Jordan, M., 616
Jordan, P., 629
Joyce, James M.
  Bayesian confirmation theory, 148
  decision theory, 183, 185
  inductive logic, 386
Judgements
  analyticity, 12
  explicative, 288
  perceptual, 546
Judson, Horace Freeland, 335, 762
Juhl, Cory, 382
Jukes, Thomas, 254, 582
Junk DNA, 7, 482
Junk science, 189
Justification, Kuhn's disciplinary matrices and scientific revolution, 427
Justification, theories of
  coherence theories, 246–247
  epistemology
    doxastic/nondoxastic and internalist/externalist, 245
    externalist, 248–249
  induction, problem of, 381
  Lakatos and, 437
  physicalism, 559, 565–567
  prediction, 587
  scientific change, 730
Justified true belief (JTB) analysis, 244–245, 261
Justus, J., 165

K

Kaas, J.H., 517
Kahneman, Daniel, 184, 227
Kaila, Eino, 858
Kalai, E., 328
Kalai-Smorodinsky solution, 328
Kamlah, A.
  Reichenbach, 705, 706, 708, 709
  Vienna Circle, 862
KAM (Kolmogorov-Arnold-Moser) theorem, 122
Kant, Immanuel, 447
  analyticity, 11–12, 13
  astronomy, 36
  Carnap, 79, 82
  demarcation problem, 190
  and emergentism, 231, 232
  empiricism, 237
  explication, 287, 288
  Hilbert and, 360
  induction, problem of, 380
  knower/known distinction, 142
  Lakatos and, 436
  mind-body problem, 61–62
  Newton's laws, 117
  perception, 547
  phenomenalism, 551, 552, 553
  Poincaré and, 169
  Reichenbach and, 706
  Schlick and, 174, 727
  scientific style, 765
  space-time, 786
  symmetries, 540
  unity and disunity of science, 843–844
  Vienna Circle and, 859
Kantian view
  Mach and, 468
  Poincaré and, 568, 569, 570, 571
  referential semantics, 447
  Reichenbach and, 706
  Schlick, 727
  scientific revolutions, 758, 760, 764
Kaons, 540
Kaplan, David
  Carnap, 86
  cognitive significance, 138, 139
  Putnam, 623
Kaplan, Mark, 185
Kaplan, R., 449
Karr, J.R., 164
Kastler, D., 506
Katz, J.J., 126, 445, 660, 661
Kauffman, Stuart A., 469, 699
Kaufmann, Felix, 858
KEKADA system, 731
Keller, Evelyn Fox
  biological information, 67
  biology, philosophy of, 73
  feminist philosophy, 297, 298, 299
  molecular biology, 485
  scientific metaphors, 739
  unity and disunity of science, 846
Keller, L., 375
Kellert, S.H., 825
Kelly-Gadol, J., 298
Kelsen, Hans, 858
Kelvin, Lord (William Thomson), 417, 755
Kemeny, John, 87, 393
Kempthorne, O, 804, 811
Kepler, Johannes
  astronomy, 34
  Kuhn's scientific revolutions, 422, 423–424, 427
  mechanics, 116–117
  scientific metaphors, 739
  scientific revolutions, 754
Kepler's laws, 25, 116–117
  explanation, 278
  KAM (Kolmogorov-Arnold-Moser) theorem, 122
  Newton and, 423–424
Keynes, John Maynard
  Bayesianism, 49
  Carnap, 86
  inductive logic, 384, 386–387, 390

Keynes, John Maynard (*Continued*)
  prediction, 589
  probability theory, 600, 603
  Ramsey and, 671, 675–676
  Russell and, 716
  Turing and, 835
Kieseppa, Ilkka, 856
Kim, Jaegwon, 226
  causality, 91
  cognitive science, 129
  emergence, 232, 233, 234
  evolutionary epistemology, 262
  physicalism, 560
  reductionism, 699
  supervenience, 816, 817, 818
  unity and disunity of science, 845
Kimura, Motoo, 254, 338
  natural selection, 500–501
  population genetics, 581, 582
Kincaid, Harold
  abduction, 1
  economics, 225, 226
  laws of nature, 443
Kinds, evolution of species, 801;
    see also Natural kinds
Kinematics
  causality, 92
  classical mechanics, 116
  Minkowski space-time, 791
Kinematics, probability, 679
Kinetics, classical mechanics, 116
Kinetic theory, 415–419
  atomism and scientific realism, 417–418
  explanation and reduction, 418–419
  historical background, 416–417
  physical sciences, philosophy of, 554
  reductionism, 696, 698, 700
King, Anthony, 219
King, Jack, 254, 582
Kinkaid, Harold
  laws of nature, 443
  methodological individualism, 479
Kin selection, 311; *see also* Natural selection
Kipling, Rudyard, 6
Kirchhoff, Gustav Robert, 35, 117
Kirchoff's law of radiation, 359
Kirschner, M., 486
Kitagawa, G., 536
Kitcher, Philip
  biology, philosophy of, 69, 71
  determinism, 197
  evolutionary psychology, 264
  explanation, 283, 284, 285
  fitness, 310–311
  Kuhn, 426, 430
  laws of nature, 442
  logical empiricism, 463
  molecular biology, 484
  natural selection, 498
  realism, 695
  scientific progress, 752, 753
  social sciences, 784
  species, 800
  underdetermination of theories, 840, 841
  visual representation, 869
Kleene, S.C., 84
Klein, Felix, 356, 790
Klein, U., 102
Klein-Gordon field, 630
K-mesons, 540
Kneale, M., 384
Kneale, W., 384
Knight, Rob D., 72
Knock-out, gene, 266
Knorr-Cetina, K., 739
Knower/known distinction, 142
Knowledge
  analyticity, 12
  artificial intelligence, 722
  Bayesianism, 55–59
  chemistry, philosophy of, 102
  confirmation theory, Bayesian, 148
  empiricism, 235
  epistemology, 244–251
  evolutionary epistemology, 257–262
  game theory, 326
  inductive logic, epistemic interpretations
      of probability, 386–387
  observation, 524
  perception/perceptual
      states and, 546–547
  phenomenalism, 552
  Popper and, 574–575
  protocol sentences, 82–83
  Putnam and, 620
  Ramsey, value of, 679
  rational reconstruction, 681–685
  scientific, sociology of, 776
  scientific change, 729–732
  scientific models, 745–746
  social constructionism, 776
  social sciences, 784
  sociology of, 784
  verifiability, 851
Koch, C.
  consciousness, 160, 161, 617
  perception, 549
  psychology, philosophy of, 617
Kochen, Simon
  quantum logic, 642
  quantum measurement problem, 648
  quantum mechanics, 655, 656
Kochen-Specker theorem, 453, 656
Koertge, Noretta, 685, 784
Kohler, Wolfgang, 232
Kohne, D.E., 336
Kolmogorov, A.
  inductive logic, 386, 387
  KAM (Kolmogorov-Arnold-Moser)
      theorem, 122
  probability theory, 599–601
Kolmogorov-Arnold-Moser (KAM) theorem, 122
Kolmogorov's axiomatization
  inductive logic, 386, 387–388
  probability, 599–601
Koshland, D.E., 487
Koyré, Alexandre, 754, 756, 763
Kraft, Victor, 304, 858, 860
Krebs, D., 264
Krebs, H.A., 731

- Krimsksy, S., 337
- Kripke, Saul, 88
  - linguistics, philosophy of, 446, 448, 449
  - realism, 695
- Kruse, M., 808
- K-systems, determinism, 203–204
- Kuhlmann, Meinard, 632
- Kuhn, H., 507
- Kuhn, Thomas, 40, 58, 88, 123, 124, 196, 419–431
  - demarcation problem, 191, 192–193, 194
  - disciplinary matrices and scientific revolution, 421–427
    - contexts: discovery, justification, development, 427
    - normal science, 423–424
    - progress, 422–423
    - rationality, 424–425
    - rationality and the social, 425–427
    - Scientific Revolution, The, 427
  - Duhem thesis, 210
  - evolutionary epistemology, 259–260
  - experimental science, studies of, 269
  - experiment and rise of modern science, 268
  - feminist philosophy, 296, 301
  - Feyerabend and, 305–306
  - Hempel and, 347, 351
  - incommensurability, 370–373
  - instrumentalism, 404
  - Lakatos and, 438
  - life and work, 419–420
    - Gestalts and world changes, 427–428
    - influence, 429–430
    - later work, 428–429
    - revolutions and two kinds of science, 420
    - structure, 420–421
  - logical empiricism, 464
  - observation, 527
  - parsimony, 532
  - protocol sentences, 612
  - quantum mechanics, 651
  - Quine and, 668–669
  - rational reconstruction, 684, 685
  - realism, 695
  - research programs, 712
  - scientific change, 730
  - scientific domains, 734
  - scientific metaphors, 738
  - scientific progress, 749, 751–752, 753
  - scientific revolutions, 103, 754, 756, 759, 760–762, 763
  - scientific style, 767
  - social constructionism, 776
  - theory structure, models and, 827
  - theory structure, syntactic (received) view, 823
  - unity of science movement, 848
- Kuhn loss, 758
- Kuipers, Risto, 855, 856, 857
- Kukla, André, 693, 694
- Kulkarni, D., 731
- Kumar, Sudhir, 254
- Kusch, Martin, 682
- Kvart, Igal, 92, 99
- Kyburg, H.E.
  - inductive logic, 386, 387
  - statistics, 810
- Kyprianidis, A., 600

L

- Labinger, Jay, 778
- Lack of coextension thesis, physicalism, 562
- Ladyman, James
  - empiricism, 242
  - theory structure, 827
    - semantic view, 825
    - syntactic (received) view, 824
- Lagrange, Joseph Louis, 119
- Lagrange's principle of virtual work, 120
- Lagrangian field theory model, 541, 544
- Lagrangian solution, three-body problem, 121–122
- Lakatos, Imre
  - connectionism, 151
  - corroboration, 178
  - demarcation problem, 192, 196
  - economics, 226
  - Feyerabend and, 305, 309
  - Kuhn, 421, 423
  - life and work, 433–439
    - methodology of scientific research programs, 438–439
    - proofs and refutations, 433–437
    - transition to research program, 437–438
  - Popper and, 577
  - prediction, 589, 591
  - Quine and, 669
  - rational reconstruction, 681, 684–685
  - realism, 690
  - research programs, 712–715
  - scientific change, 730, 731
  - scientific progress, 752
  - scientific revolutions, 759–760
- Lamarck, Jean Baptiste, 252–253, 353
- Lamarckian models, cybernetics, 262
- Lamb, Willis, 26
- λ – continuum, 50, 51
- Lamb shift, 26
- Lamphere, Louise, 298
- Lanczos, Cornelius, 120
- Lande, Russell, 254
- Landres, P.B., 166
- Landscape, adaptive, 582
- Landsteiner, Karl, 366
- Lanford, Oscar E., 202
- Lange, M.
  - Duhem thesis, 210
  - laws of nature, 443
  - prediction, 591
- Langley, P., 731
- Langman, Rodney, 364
- Langmuir, Irving, 194–195
- Language
  - behaviorism, Chomsky and, 62–63
  - Carnap
    - semantics, 85–86
    - syntactic phase, 83–85
  - cognitive significance, 134–138
  - demarcation problem, 189
  - Duhem thesis, 209
  - empiricism, 238–239
  - evolutionary psychology, 266
  - experimental method, 274
  - explication, 288, 290, 291, 292
  - function, teleological description, 316

Language (*Continued*)
  generative grammar, I versus E, 110
  Hahn and, 343
  Hanson, on facts, 345–346
  incommensurability, 372
  innate/acquired distinction, 398
  innate/acquired distinction, nativism, 396
  Kuhn, later work of, 428
  Lakatos and, 435–436
  linguistics, philosophy of, 444–450
  logical empiricism, 461
  Neurath and, 512–513
  phenomenalism, 551–552
  philosophy of
    etiological function, 69
    Putnam and, 620
    Schlick and, 728
    social science foundational problems, 783
  of prediction, 585–586
  probability theory, 599–600, 603–604
  protocol sentences, 82–83, 610, 611
  Putnam and, 620, 623–625
  Quine and, 659, 660, 661, 664
  Ramsey's hierarchy of languages, 674
  reductionism, 696–697
  Reichenbach and, 705–706
  Russell and, 720–722
  Schlick and, 726, 728
  scientific domains, 735, 736
  scientific metaphors, 737–739
  Searle, 767–774
  tautological statements, 40
  theory structure, syntactic (received) view, 822, 823
  unity and disunity of science, 843, 844–845
  verifiability, 852, 853
  verisimilitude, 856
  Vienna Circle and, 859, 860
Language, Truth, and Logic (Ayer)
  Ayer, 38, 39, 40
  Carnap, 82
  cognitive significance, 133
  empiricism, 238
  logical positivism, 459, 460
  verifiability, 852
Language acquisition, Chomsky, 109, 111, 112
Language of science, 402–403
  Nagel on, 493
Language-transcendent concepts, analyticity, 14–15
Lankford, J., 33
Laplace, Pierre Simon de
  Bayesianism, 41, 42, 49
    induction, problem of, 47–48
    principle of indifference, 52
  classical mechanics, 118
  determinism, 198
  inductive logic, 384
  Nagel and, 492
  natural selection, 501
  probability theory, 602
  Reichenbach and, 708
Laplace's demon, 118
Laplace's law of succession, 47
Laplace's principle of indifference, 52, 58
Laplace's rule, Bayesianism, 41–42
  inductive logic as generalizations of, 48
  rules of succession, 50
Laplacian determinism
  classical physics, 199, 200
  quantum physics, 204, 205
  relativistic physics, 206
Large N problem, 808–809, 812
Large numbers, weak law of (Bernoulli), 46, 47
Latour, Bruno, 269, 776–777, 779
Lattice logic, orthomodular, 638–639
Lattices
  non-Boolean structure, 633
  nondistributive, 506
  quantum lattice logic, 633
  quantum mechanics, 656
Laubichler, M.D., 355
Laudan, Larry
  demarcation problem, 190, 193, 194
  instrumentalism, 404
  realism, 693–694
  scientific change, 730, 731
  scientific progress, 752, 753
  social constructionism, 776
  underdetermination of theories, 840
Laudan, R., 731
Lauder, George V., 72, 316
Lauraudogoitia, Jon Perez, 201, 202
Lavoisier, Antoine
  chemistry, philosophy of, 102
  scientific domains, 736
  scientific revolutions, 754
  unity and disunity of science, 844
Law-based capacities, causality, 97–98
Law of circular inertia, 307
Law of distributivity, 648
Law of effect (Thorndike), 61
Law of errors, Gaussian/normal, 56
Law of excluded middle, 721
Law of first digits, 58
Law of gravitation, Kuhn's scientific revolutions, 423–424
Law of identity, 721
Law of independent segregation (genetics), 330
Law of likelihood, 533
Law of segregation (genetics), 330
Law of succession (Laplace), 47
Laws of coexistence, 824
Laws of likelihood, 178, 533, 807
Laws of motion, space-time, 786
Laws of nature, 439–444
  accidental generalizations, versus, 439
  best systems account, 94, 442–443, 607–608
  biological, 68–69, 376–377, 443
  causality, 90–91, 92–94
  chemistry, philosophy of, 102
  corroboration, 177–179
  economics, laws of, 225–226, 443
  evolutionary biology, natural selection, 502
  explanation, covering-law models, 276–280
  explanation, 276–278
  Hempel, 348
  instrumentalism, 400–401
  Nagel, 493, 495–496
  natural selection, 501
  possible world accounts, 440–441

  probabilistic, 439–442
  psychology, 62, 443, 564–565
  reductionism, 696–697
  Ramsey, 679–680
  Russell and, 720
  scientific domains, 733
  scientific models, 740, 748
  sophisticated regularity accounts, 442–443
  social science, 443
  species, 801–802
  statistical
    explanation, 276, 278, 279
    versus statistical generalizations, 495–496
  supervenience, 817
  theory structure, 824
  unity and disunity of science, 843, 845, 846
  universals accounts, 441–442
  universals of fact, versus, Ramsey, 680
  verifiability, 853–854
Laws of succession, theory structure, 824
Laws versus generalizations, statistical, 495–496
Lawton, John, 223
Layer-cake model, scientific progress, 750
Laymon, R., 26, 742
Lazarsfeld, Paul, 491
Learning
  generative grammar, conceptual issues in, 112–113
  Hebbian, 152
  innate/acquired distinction, 396, 397
  language, 445
  language and, 428
  neurobiology, 521
  psychology, cognitive architecture, 616
Learning rule, network, 153
Leatherdale, W.H., 739
Leavitt, Henrietta, 36
Lebensweisheit, Fragen der Ethik (Schlick), 727
Leblanc, Hughes
  inductive logic, 386
  probability theory, 600
Lebowitz, Joel L., 411
Lederberg, J., 731
Legal decisions, 195
Legend, scientific progress, 752
Leggett, A.J., 698
Legisimilitude, 856
Leher, K., 247
Lehmann, E.L., 43, 45, 803
Leibnitz shift argument, 555
Leibnitz, Gottfried, 589
  analyticity, 11
  classical mechanics, 119
  explication, 288
  innate/acquired distinction, 395, 396, 397
  Newton and, 118
  parsimony, 532
  physical sciences, philosophy of, 555
  probability theory, 602, 603
  quantum field theory, 631
  Russell and, 722
  scientific domains, 735
  space-time, 788
  unity and disunity of science, 843
Lelas, Srðan, 269, 272
Lemma, fixed-point, 85
Lemma incorporation, 437–438
Leonard, Robert, 507
Leontieff, V., 848
Leplin, Jarrett, 226
  prediction, 590, 591–592
  realism, 687, 689, 692, 693, 695
  underdetermination of theories, 840
Lepore, E., 660
Leptons, 539, 542
LeRoy, Edouard
  conventionalism, 169
  Poincaré and, 571
Lesniewski, S., 659
Lessard, S., 581
Levels
  artificial intelligence, 31–32
  of mechanisms, 472
  reductionism, 697, 700
    interlevel reduction, 699–701
    intralevel reduction, 698–699
    unity and disunity of science, 845
  of selection, *see* Biology, philosophy of; Natural selection
Le Verrier, Urbain, 118
Levi, Isaac, 182, 607
Levin, Michael, 691
Levine, J., 130, 617
Levins, Richard
  ecology, 216, 217, 220, 221
  immunology, 365
  reductionism, 701–702
Levitt, Norman, 778
Lewes, G.H., 231
Lewin, K., 704
Lewis, C.I., 85
  causality, 94
  cognitive significance, 134
Lewis, David, 120
  causality, 99
  conventionalism, 176
  game theory, 326
  laws of nature, 439, 440, 442, 443
  linguistics, philosophy of, 445
  metaphysical realism, 622
  probability, 607
  probability theory, 600, 607, 608
  Putnam and, 622
  Ramsey and, 679
  time, 832
Lewontin, Richard
  adaptation and adaptationism, 5, 6
  Chomsky and, 112, 113
  ecology, 216, 217
  feminist philosophy, 299
  fitness, 311
  heritability, 353, 355
  immunology, 365
  innate/acquired distinction, 398, 399
  natural selection, 497, 498
  population genetics, 582, 584
  unity and disunity of science, 846
Lexical definitions, explication, 287
Lexical items, arrays of, 109
Li, C.C., 581
Liapunov exponents, 203, 204
Liberal outlook, 715

Lichtenberg, A.J., 203, 204
Lie, Sophus, 170
Lieberman, M.A., 203, 204
Life Wisdom (Schlick), 725
Light
  scientific domains, 734
  velocity of, 789, 790–791, 792
Light cone, 790–791
Lightlike vectors, 791
Light quanta, 651
Light-ray geometry, Reichenbach and, 705, 706–707
Likelihood
  laws of, 178, 533, 807
  parsimony, 533, 534
  probability theory, 600
Likelihood approach to evaluation of evidence, 149
Likelihood function
  Bayesianism, 42, 43
  confirmation theory, Bayesian, 148–149
Likelihood principle (LP)
  Bayesianism, 44
  error probability principle versus, 807–809
Likelihood ratio, Bayesianism, 42
Likeness, verisimilitude, 856
Limiting conditions, reductionism, 699
Limit principle, and locality/local action, 452
Lindberg, David C., 545, 754
Lindenbaum algebra, 634, 635
Lindh, A.E., 75, 76
Lindley, Dennis
  Bayesianism, 56, 57, 59
  Jeffrey-Good-Lindley paradox, 809
Lindman, H., 55, 808
Lindsay, R.K., 731
Lineages
  species
    evolutionary, 798
    evolution of, philosophical implications, 802
    pluralism, 800
  species taxa, 796
Lineage species concept, 799–800
Linear coherence theories, 247
Linearity requirement, quantum mechanics, 655–656
Linear organization, mechanisms, 475
Lines of force concept, Faraday, 452
Linguistic ability
  innate/acquired distinction, 398
  intelligence and, 396
Linguistic analysis, Schlick and, 728
Linguistic capability, as adaptive trait, 4
Linguistic conventions
  conventionalism, 176
  Quine and, 660
Linguistic definitions, Reichenbach and, 705
Linguistic empiricism, Neurath and, 513
Linguistic engineering, 290
Linguistic representation
  cognitive science, 128–129
  levels of, in generative grammar, 109
  visual representation and, 869
Linguistics
  cognitive science multidisciplinary approach, 123, 124, 126–127
  inductive logic frameworks, 389
  philosophy of, 444–450
    language faculty and external world, 446–448
    methodological issues, 449
    minimalist program, issues raised by, 449–450
    object of study, 444–446
    rules and representations, 448–449
  protocol sentences, 611
  psychology, philosophy of, 614
  reductionism, 696–697
  scientific revolutions, 755
Linguistic theory
  Chomsky, 106–115
  consciousness, 161
  generative grammar, 107–114
    conceptual issues in, 109–114
    development and evolution of, 107–109
  innate/acquired distinction, nativism, 396
Linguistic theory structure, 827
  semantic view, 825
Linguistic turn, 737
Linkage, genetic, 331–332
Linnaeus, C., 766
Liouville's theorem, 121
Lipsitz, Imre, *see* Lakatos, Imre
Lipton, Peter, 1, 591
Little, Daniel
  economics, philosophy of, 224
  methodological individualism, 479
  social sciences, 780
Lloyd, Elizabeth A.
  biology, philosophy of, 70, 73
  natural selection, 498
  theory structure, 825
Lloyd Morgan, C., 231, 232
Llull, Ramon, 843
Lobatchevskian geometry, 170, 569
Local accounts, parsimony, 532–534
Local flow, classical mechanics, 121
Local hidden variable model, 454, 456
Local inertial frames, 792
Localist networks
  connectionism, 153
  psychology, cognitive architecture, 617
Locality, 451–457
  action at a distance, 455–456
  Bell's theorem, 453–454
  Bell's theorem, interpretation of, 454
  entanglement and EPR, 452–453
  history of philosophy, 451
  in modern physics, 451–452
  nonlocality, subtleties of, 456
  physical sciences, philosophy of, 554
  quantum field theory, 631, 632
  quantum mechanics, 454–455, 656
  reductionism, 700–701
  in relativistic quantum field theory, 456–457
  time, presentism, 831
  Turing and, 835
  von Neumann and, 505
Localization
  individuality, criteria for, 377
  neurobiology, 516–518
  quantum field theory, 631
Local level view, classical mechanics, 121
Local particle density, quantum field theory, 631
Local stability analysis, ecological communities, 219

Location, Bohr and, 142
Loci classici, instrumentalism, 400–402
Lock-and-key fit rule, molecular biology, 483
Locke, John
  empiricism, 236, 237
  perception, 545, 546, 547
  verifiability, 851
Lockwood, Michael, 205
Loeb, Leo, 364
Loewer, Barry
  determinism, 205
  laws of nature, 443
  physicalism, 567
Logic
  abduction, 1–3
  analytical philosophy, 715
  Bayesianism, 42–43
  Carnap, syntactic phase, 83–84
  demarcation problem, 189, 572–573
  explication, 290
  inductive, *see* Inductive logic
  Kuhn's scientific revolutions, 425
  Lakatos and, 434
  logical empiricism, 459
  logical positivism, 460–461
  Nagel and, 492
  neurobiology, 509
  predicate, 709, 715
  Putnam and, 620, 621
  quantum, *see* Quantum logic
  Quine and, 660, 664
  Ramsey, *Foundations of Mathematics, The*, 672
  Reichenbach and, 709
  Russell and, 717–719
  three-valued, 710–711
  Turing and, 833
  von Neumann and, 504, 509
Logical analysis
  falsificationism and methodological rules, 574
  Russell and, 720–722
Logical Atomism (Russell), 719
Logical atomism, 722, 728, 859
Logical calculus, theory structure, 822
Logical conventionalism, 174–176
Logical empiricism, 458–465
  American Structuralism, 107
  analyticity, 13
  approximation, 24–25
  Ayer and, 38–40
  behaviorism, 62
  biology, philosophy of, 69
  Carnap and, 79, 88
  causality, 95
  cognitive significance, 131–132, 134–138
  demarcation problem, 188–197
  dilution: analyticity and cognitive significance, 461–463
  dissolution of, unity, realism, and philosophy of science, 463–464
  distillation: central themes of logical positivism, 459–461
  Duhem thesis, 209
  explanation, covering-law models, 276
  geometry, 360
  Hanson and, 344–345, 346–347
  Hempel and, 347, 350
  incommensurability, 370–373
  instrumentalism, 402
  Kuhn's scientific revolutions, 427
  logical positivism versus, 458
  Neurath and, 510–512
  observation, 523–530
  prediction, 587–588
  protocol sentences, 610–613
  Quine and, 659, 665
  rational reconstruction, 681, 682–684
  rational reconstruction and, 682
  Reichenbach and, 704–705
  Russell and, 720
  scientific domains, 733
  scientific metaphors, 737
  scientific progress, 750
  scientific revolutions, 755, 756
  theory structure, 822–824
  Vienna Circle, 858–859
Logical form of representation, 109
Logical Foundations of Probability (Carnap), 86–87, 287, 384, 393
Logical interpretation of probability, 603–604
  Carnap, 86–87
  explicanda/explicata pairs, 288
  in inductive logic, 387–391
  Ramsey, 675
Logical knowledge, analyticity, 12
Logical methodology, Carnap, 85
Logical necessity, corroboration, 177
Logical positivism, 82–83
  Ayer and, 39
  behaviorism and, 62
  Carnap, 79
  empiricism, 237–238
  kinetic theory, 417
  Kuhn and, 429–430
  logical empiricism and comparison, 458
    distillation: central themes of logical positivism, 459–461
  physicalism, 558
  Poincaré and, 570, 571
  Popper and, 572, 573
  scientific progress, 750
  unity and disunity of science, 845
  verifiability, 851
  Vienna Circle, 858–859
Logical probabilistic empiricism, Reichenbach and, 708
Logical relation, research programs, 713
Logical structure of the world, empiricism, 525
Logical Structure of the World, The (Carnap), 13, 238
Logical syntax, protocol sentences, 612
Logical Syntax of Language, The (Carnap), 13, 39, 84–85, 175, 209
Logical truth, explicanda/explicata pairs, 288
Logical validity, analyticity, 12
Logicism, 80
  Carnap, Rudolf, 84
  Hilbert's formalistic approach versus, 357
  Poincaré and, 569–570
  Ramsey, abandonment of, 674
  Russell and, 718–719
  Vienna Circle, 82
Logico-empirical structure of science, Quine and, 667
Logic of confirmation, *see* Confirmation theory

Logic of consistency, Ramsey, 675, 677–678, 680
Logic of discovery, Hanson on, 345, 346–347
Logic of Modern Physics, The (Bridgman), 76
Logic of science, Vienna Circle and, 859
Logic of Scientific Discovery, The (Popper), 260, 766
Logic of truth, Ramsey, 678–679
Logico-linguistic structure of science, scientific revolutions, 757
Logic-probability relationship, von Neumann and, 504
Logic without Metaphysics (Nagel), 492
Logik der Forschung (Popper), 343, 572
Logische Aufbau der Welt, Der (Carnap), 80, 81, 82, 84
  observation, 523, 524
  phenomenalism, 552
  prediction, 587
  protocol sentences, 610
  rational reconstruction, 682, 684
Logische Syntax der Sprache (Carnap), 39, 84–85, 175
Lolle, S.J., 486, 488
Longino, Helen E.
  feminist philosophy, 297, 299, 300, 301
  Kuhn’s scientific revolutions, 426–427, 430
  underdetermination of theories, 841
Loos, Adolph, 861
Loredo, T.J., 43
Lorentz invariance, 540
Lorentz-invariant quantum mechanics, 642
Lorentz signature, relativistic space-time, 205
Lorentz transformations, 789
Lorenz, Konrad, 264, 265
  evolutionary epistemology, 257
  evolutionary psychology, 263
  innate/acquired distinction, 398
Lorenzen, Paul
  experimental method, 269
  quantum logic, 642
Lorenz model of atmosphere, 740
Loschmidt, L., 417, 418
Loss aversion, decision theory, 184
Loss function, Bayesianism, 41
Lotka-Volterra community models, 223, 740
Lottery, game theory, 323–324
Löwenheim-Skølem theorem, 621, 823
L-truths, verifiability, 852
Luce, R.D., 323
Luckmann, Thomas, 774–775
Lucretius, 252
Ludlow, P., 447, 449
Ludwig, Donald, 215
Ludwig, G., 600
Lukács, György, 433
Lukasiewicz, J., 659
Lummer, O.R., 650
Lumpers versus splitters, species, 801
Luria, Salvator, 335
Lycan, William
  analyticity, 15, 18
  intentionality, 409
Lyell, Charles, 252
Lynch, Michael
  demarcation problem, 196
  social constructionism, 776
  visual representation, 866
Lyre, Holger, 632

M

Maanan, Adrian van, 36
MacArthur, Robert, 164, 218, 220
Macdonald, Cynthia, 626
Mach, Ernest, 117
  classical mechanics, 115, 116
    action principles, 120
    energy conservation, 119
  demarcation problem, 193
  empiricism, 237
  Hahn and, 341, 342
  instrumentalism, 400–401
  kinetic theory, 417
  life and work, 467–469
  Neurath and, 511
  phenomenalism, 551
  Reichenbach and, 707
  Schlick and, 725–726
  scientific style, 765
  space-time, 788, 791
  unity and disunity of science, 844
  Vienna Circle, 858, 861
Machamer, Peter
  biology, philosophy of, 72
  explanation, 282
  mechanism, 469
Mach bands, 467
Machiavelli, N., 754
Machines, Turing, 834–835
  effective calculability, 833
  intelligent machinery, 836–837
  practical machinery, 835–836
Mach’s principle, 117
MacIntyre, Alasdair, 764
Mackay, T.F.C., 354, 355
Mackey, G.W., 507
Mackie, John, 94
Macko, K.A., 516
Maclaurin, James, 399
MacLeod, Colin M., 335
Macrae, Norman, 504
Macrocausal relations, 90
Macrodeterminacy, 231
Macro-level event, causality, 91
Macromolecules, 480
Macrostates
  phase space, 411–412
  reductionism, 701
Magnani, Lorenzo, 744
Magnetic quantum number, 652
Magnets, Stern-Gerlach, 453
Magnitude, kinetic theory, 416
Maher, P.
  Dutch Book argument, 213
  inductive logic, 387, 389
  probability theory, 604
Maiani, L., 542
Maienschein, J., 353
Maintenance of variation, population genetics, 579–580
Majer, U., 358, 359
Makinson, D,, AGM model, 249–250
Malament, David
  conventionalism, 171–173
  physical sciences, philosophy of, 554

  quantum field theory, 631
  quantum logic, 639
  time, 830
Malcolm, S.B., 164
Malebranche, Nicholas, 352
Malfunction
  mechanism discovery, 475, 476
  normative concepts, biology and, 69
Malinowski, B., 782
Malphigi, Marcello, 352
Malsburg, C. von der, 617
Malthusian economics, 364
Mandik, P., 514
Mangel, Mark, 215
Manhattan Project, 504, 846
Manifestability argument, physicalism, 567
Manifold
  relativistic space-time, 830
  space-time, 786, 829, 831
Manipulability, counterfactuals and, 100
Mann, Charles, 762
Mannheim, Karl
  scientific style, 765–766
  sociology of knowledge, 784
Manning, R., 320, 321
Many-minds interpretation, 205
Many-particle Hilbert space, quantum field theory, 630
Many worlds interpretation, 205, 642–643
Map analogy, Nagel and, 494–495
Mapping
  abstract dynamical systems, 203
  brain, 474
  phenotypic space, 796
  theory structure, 827
Maps, celestial, 35
Maps, gene, 353
Marcus, S.J., 521
Marcuse, Herbert, 269
Margoliash, E., 254
Margolis, Howard, 761, 763
Margules, C.R., 164, 165, 166
Markosian, Ned, 831
Markov Chain Monte Carlo (MCMC) method, 810
Markov conditions, causal, 96, 97
Marr, David
  artificial intelligence, 31
  cognitive science, 126–127
  computer model of mind, 547–550
    consciousness, problem of, 549
    criticism from neuroscience, 548–549
    criticism from perceptual psychology, 548
    Gibson's attack on computationalism, 549–550
  evolutionary psychology, 265
  innate/acquired distinction, nativism, 396
Martin, C.B., 98
Martin, Emily, 365
Martin, J., 399
Martin, Michael, 780
Martin, Paul, 267
Martin, R.M., 14, 19
Marx, Karl, 269, 512
  Neurath and, 511
  Popper and, 575
  scientific revolutions, 759
  social sciences, 784
  theories of social change, 755
Marxist theory
  demarcation problem, 191, 192, 574
  feminist philosophy, 300
  Lakatos and, 433, 434
  Neurath and, 511, 513
  Popper and, 574, 586
  prediction, 586
Mass, classical mechanics, 117
Massive modularity thesis, evolutionary psychology, 266
Master control genes, 481
Masterman, Margaret, 421
Material causes (Aristotle), 352
Material conditionals
  defined, 177
  inductive logic, 385
  quantum logic, 639
Material genetics, 334–335
Materialism, *see* Physicalism
Material models, 743
Mates, Benson, 15, 17, 19
Mathematical analysis of inductive reasoning, 49
Mathematical conventionalism, 174–176
Mathematical Foundations of Quantum Mechanics (von Neumann), 505, 506
Mathematical intuition, 357
  Hilbert, 360
  Poincaré, 569–570
  Reichenbach, 707
  Turing and, 836
Mathematical logic, Kuhn's scientific revolutions, 425
Mathematical mechanics, classical mechanics, 121–122
Mathematical models/modeling
  ecological communities, 219, 220–221
  ecology and conservation biology, 72
  evolution, 253
  experimental method, 271
  fitness, population genetics, 310–311
  population genetics, 583
  scientific models, 744
  theory structure, 827
Mathematical space, 80
Mathematical theory of probability, 386
Mathematicism, scientific style and, 765
Mathematics
  analyticity, 13
  approximation, 24, 26
  artificial intelligence as, 29
  Carnap, syntactic phase, 83–84
  classical mechanics, 115
    algebraic calculus, 119
    celestial mechanics and determinism, 118–119
    classical dynamical systems, 121–122
    variational calculus, 119–120
  cognitive science multidisciplinary approach, 123
  connectionism, 151
  conventionalism, 169
  demarcation problem, 188–197, 572–573
  discursive domains, Carnap, 80
  economic theories, 226
  experimental method and, 272
  experimental tradition, merging with, 268
  Hahn, 341–343

Mathematics (*Continued*)
  Hilbert, 356–361
    axiomatic method, 357–358
    finite point of view and recursive epistemology, 360–361
    geometry, 357–358
    proof theory, 358–359
  inductive logic
    Bayesianism, 43
    probability models, 384, 386
  Kuhn's scientific revolutions, 424
  Lakatos, 433–439
  mechanics, 116
  mechanism, representation of, 472–473
  neurobiology, 515
  population genetics, 579, 582
  quantum field theory, 629–633
  quantum logic, *see* Quantum logic
  Quine, 660
  Ramsey, 671, 672, 673
  rational choice program, 781
  reductionism, 698
  Russell, 715, 716–719
  Schlick, 726, 727–728
  scientific style, 766
  theory structure, 827
  Turing, 833, 834, 835
  unity and disunity of science, 843, 844
  Vienna Circle, 858, 860
  visual representation, 867
  von Neumann, 503–510
Mathematische Grunglagen (von Neumann), 360
Mather, John, 122
Mathesis universalis, 843
Mathews, R., 127
Matrix mechanics, 653
Matthen, M., 618
Mattlick, J., 487, 488
Mature science, scientific progress, 751–752
Maudlin, Tim
  irreversibility, 411
  locality, 452, 455–456
  quantum mechanics, 656
  time, 830, 831, 832
  unity and disunity of science, 845, 846
Maull, Nancy
  biology, philosophy of, 71
  reductionism, 699
  unity and disunity of science, 846
Maupertuis, Pierre Moreau de, 118
Maximal specificity, principle of, 279, 603
Maxwell, Grover
  corroboration, 179
  realism, 687
  scientific domains, 734
Maxwell, James Clerk
  causality, 92–93
  instrumentalism, 403
  kinetic theory, 416, 418
  and locality/local action, 452
  scientific metaphors, 737, 739
  scientific revolutions, 755
  scientific style, 766
  unity and disunity of science, 846–847
Maxwell-Boltzmann distribution law, 416
Maxwell-Boltzmann equation, 359
Maxwell's demon, 418
Maxwell's theory, 416
May, Robert, 219, 221
Mayden, R.L., 800
Maynard Smith, John
  adaptation and adaptationism, 5
  altruism, 8, 9, 10
  biological information, 65–66, 67
  developmental biology, philosophy of, 72
  game theory, 329
  molecular biology, philosophy of, 72
Mayo, Deborah G.
  Bayesianism, 43, 45
  experimental method, 270
  induction, problem of, 382
  prediction, 591
  statistics, 808
    error statistics, reforms within, 811
    power analytic movement, 809
Mayr, Ernest
  evolution, 253, 254
  evolutionary epistemology, 258
  molecular biology, 481
  species, reproductive isolation, 797–798
  unity and disunity of science, 846
Mazumdar, Pauline M.H., 364, 366
McCarty, Maclyn, 335
McClay, J., 399
McClelland, James, 151, 156
McClennen, Edward, 182
McCormach, E.R., 738
McCoy, Earl, 72, 216, 220
McCullough, D.R., 167
McCullough, Warren, 27, 508
McDermott, D., 160
McDermott, M., 349
McDowell, J., 547
McGehee, Richard, 122
McGuinness, Brian, 728
McIntosh, Robert, 221
McIntyre, Lee C., 780, 783
McLaughlin, Brian
  causality, 92
  emergence, 231, 232, 233
  physicalism, 564
McLaughlin, P., 317
McMullin, Barry, 508
McMullin, Ernan, 21
  demarcation problem, 189
  locality/local action, 451
  scientific models, 742, 747
McNaughton, Bruce L., 156
McNaughton, J.S., 219
McPherson, Michael S., 224, 225
McTaggart, J.M.E., 717
Mean, regression to, 353–354
Mean free path of molecules, kinetic theory, 416
Meaning
  analyticity, indeterminacy of, 14
  demarcation problem, 191
  linguistics
    American Structuralism, 107
    indeterminacy of, 110–111
    philosophy of, 446

  physicalism, 558
  protocol sentences, 613
  Putnam on, 620, 623–625
  Reichenbach and, 707–708
  scientific style and, 765
  semantic paradoxes, 673–674
  social sciences, 783
  verifiability criterion, 852
  verifiability theory of, 62
Meaning and Necessity (Carnap), 85
Meaning-constitutive (analytic) statements, Duhem thesis, 209
Meaning/intension, explicanda/explicata pairs, 288
Meaning invariance, 845
Meaning postulates, analyticity, 14
Measure functions
  inductive logic, 388, 389
  probability theory, 604
Measurement
  astronomy, 34
  atomic objects, complementarity, 140–143
  complementarity, 141–143
  computation, 28
  Euclidean geometry, 361
  locality/nonlocality, 456
    holism and nonseparability, 455
    relativistic quantum field theory, 457
  neurobiology, 514–516
  Newtonian space-time, 791
  physical sciences, philosophy of, 554
  quantum field theory, 629
  quantum measurement problem, 505–506, 644–649
  quantum mechanics, 654, 655, 656, 657
  Reichenbach and, 710–711
  Turing and, 837
  von Neuman's theory of, 505–506
Measure of confirmation, inductive logic, 393
Measure theory
  and classical mechanics, 121
  and probability theory, 599
  Von Neumann's operator theoretic methods, 507
Mechanical laws, kinetic theory, 415–419
Mechanical mode of causation, emergence, 231
Mechanical worldview, Hilbert and, 359
Mechanics, *see* Classical mechanics; Quantum mechanics
Mechanisms, 469–477
  aspects of, 469–471
    causal, 470
    componential, 469–470
    organizational, 470–471
    phenomenal, 469
  biology, philosophy of, 71–72
  discovery of, 473–474
    experiments in, 475–477
    organization of mechanism, 474–475
  economics, 225
  evolution of epistemic mechanisms (EEM), 258
  explanations, mechanistic, 473
  kinetic theory, 415
  level of, 472
  Marr's computer model of mind, 547
  methodological individualism, 479
  neurobiology, 471–472, 520
  reductionism, 697, 698
  representing, 472–473
Mechanistic explanation, 281–282
  Nagel on, 494
  prediction, 597
  reductionism, 698
  unity and disunity of science, 844
Mechanistic levels, 472
Meckel, Johann Friederich, 353
Medicine
  atomistic body, 364–365
  experimental method, 268, 270, 274
  human genome project and, 485
Medieval science
  experimental tradition, 268
  scientific style, 766, 767
  unity and disunity of science, 843
Meditations (Descartes), 395
Meehl, P.E., 809
Meffe, G.K., 164
Megill, Norman D., 634, 639
Meinong, Alexius, 406
Meiosis
  cost of, 375–376
  genetics, 331
Meiotic drive, 498
Mellor, D.H.
  causality, 90
  Chomsky, 113
  physicalism, 560
  probability theory, 607
  Ramsey, 671, 676
Melnyk, Andrew, 561, 565
Membrane transport, cellular, 486
Memes, evolutionary epistemology, 262
Memory
  connectionism, 156
  neurobiology, 521
  Searle, 770
Mendel, Gregor, 52, 253, 330, 334, 353
Mendeleev's periodic law, 590
Mendelian genetics
  biology, philosophy of, 69
  Darwin and, 499
  molecular biology, 484
  molecular biology, philosophy of, 71
  natural selection, 501
  population genetics, 578–585
  prediction, 596
  realism, 689
  reductionism, 699
Mendelian hybridization, 332
Mendel's laws
  explanation, 278
  natural selection, 501
Menger, Karl
  Carnap and, 82
  Hahn and, 342, 343
  Schlick and, 726, 728
  Vienna Circle, 858, 860
Meno (Plato), 394, 395
Menozzi, L.P., 267
Mental acts, phenomenalism, 551
Mental habits, Ramsey and, 679
Mentality, intentionality, 409

Mental processes
  behaviorism, 62–63
  connectionism, 150–158
  as neural computation, 27
  physicalism, 560
  physical versus, 558
  psychology, philosophy of, 613–619
  Turing and, 833–834, 836
Mental properties, physicalism, 560–561
Mental representations
  cognitive science, 128
  function, 319
  psychology, philosophy of, 614–615
Mental states
  behaviorism, 62–63
  intentionality, 406–407, 409
  perception, 547
  phenomenalism, 551–552, 553
  Ramsey sentences and, 679
  Searle, 767–774
Menzies, Peter, 98
Mermin, N. David, 658
Merton, Robert K., 782, 784
Merzbacher, E., 25, 26
Merzenich, M.M., 517
Meselson, Matthew, 335
Meson, 539
Meson field theory, 539
Meson theory, 538
Messenger RNAs, 335–336
Meta-epistemology, rational reconstruction, 682
Meta-ethics, 558
Metaheuristic algorithms, 731
Metalanguage
  logical positivism, 460–461
  verifiability, 853
Meta-mathematics, von Neumann, 505
Meta-methodology, statistics in, 810
Metaphorical models, *see* Scientific metaphors
Metaphysical anthropic principle (MAP), 23
Metaphysical realism, 621, 622
Metaphysical research program, Darwinism as, 260
Metaphysics
  Ayer and, 39–40
  biology, philosophy of, 69
  Carnap and, 83, 85, 133
  causality, 90, 91
  classical mechanics, space-time, 117
  cognitive significance, 132, 133, 135, 136
  complementarity, M-interpretation, 141, 143
  demarcation problem, 188–197, 572–573
  determinism, 197
  ecology, 215–217
  and empiricism, 235–236, 238
  Feyerabend and, 305, 306
  immune self, 364
  innate/acquired distinction, 395–396
  Kepler and, 117
  Kuhn on, 421, 426, 428, 429
  linguistics, philosophy of, 448
  logical empiricism and, 459, 461
  physicalism, 558, 559; *see also* Physicalism
  physical sciences, philosophy of, 555, 557
  and physics, 117
  Popper and, 572–574, 575–576
  Putnam and, 620, 621, 622
  Quine and, 664
  rational reconstruction, 683
  realism, 686, 688
  Reichenbach and, 709
  Russell and, 720, 721–722
  Schlick and, 727
  scientific revolutions, 756, 757
  species, 802
  unity and disunity of science, 844
  verifiability, 852
  Vienna Circle and, 860
Metaphysics (Aristotle), 33
Metasemantic theory, Putnam, 623, 624, 625, 626
Meta-theoretical approach, quantum field theory, 632
Metchnikoff, Elie, 365, 366, 368
Metchnikoff's theory, 365–366
Methodical constructivism, experimental method, 273
Method/methodology, *see also* Scientific method
  American Structuralism, 107
  approximation, 25, 26
  Ayer versus Carnap, 40
  Bayesianism, 42–43
  Carnap, 85
  chemistry, philosophy of, 105
  Chomsky on methodological dualism, 114
  cognitive science assumptions, 126
  cognitive significance, 132
  demarcation problem, 189, 190, 193, 194, 195
  Duhem thesis, 209
  ecology and conservation biology, 72
  experimental method, 268–275
  feminist philosophy, 297, 300
  Feyerabend and, 306–307
  game theory as, 507
  incommensurability, 371
  inductive, 1
  kinetic theory, 415
  Lakatos and, 438–439
  linguistics, 449
  mathematical, Turing and, 833
  meta-methodology, statistics in, 810
  molecular biology as, 480
  Nagel and, 491, 492
  natural selection, 498
  Neurath, 512–513, 514
  neurobiology, *see* Neuroscience, neurobiology
  parsimony, 531–538
  physicalism, 566
  Popper and, 572–576
  protocol sentences, 611, 612
  Quine and, 660, 665–666
  rational reconstruction, 681–685
  research programs, 712–715
  Russell and, 720
  scientific change, 729–732
  scientific computing and, 508–509
  scientific metaphors, 738
  scientific progress, 752
  scientific revolutions, 759–760
  simplicity principle, 360
  social sciences, 780, 782–783
    functional analysis, 781–782
  Turing and, 833
  underdetermination of theories, 839

  unity and disunity of science, 843, 846
  verisimilitude, 857
  Vienna Circle, 859–860
Method of Finding Curved Lines That Show Some Property of Maximum or Minimum (Euler), 119
Methodological Character of Theoretical Terms, The (Carnap), 524–525
Methodological fiction, confirmation theory, 145
Methodological individualism, 478–479
  economics, 226
  reductionism, 700
  social sciences, 783
Methodological principle, kinetic theory, 415
Methodology of Scientific Research Programmes (Lakatos), 438–439
Metric, conventionalism, 170–171
Meyer, Eduard, 511
Michalos, A., 393
Michelson-Morley experiment, 789
Michod, R.E., 375
Microcausal relations, 90
Microdeterminism, 231
Micro-level accounts, reductionism, 701
Micrometer, 34
Microphysics
  physicalism, 560
  Reichenbach and, 709–710
Microscopic causality, 540
Microscopic superpositions, quantum mechanics, 655
Microstates, phase space, 411–412
Mie, Gustav, 359
Miescher, Friedrich, 334
Mill, J., 399
Mill, John Stuart, 45
  approximate generalizations, 24–25
  demarcation problem, 190
  Duhem thesis, 210
  and emergentism, 231, 233
  experimental method, 269
  Feyerabend and, 305, 307–308
  inductive logic, 384
  Nagel and, 494
  observation, 524
  prediction, 589
  propositions and their testing, 524
  Ramsey and, 680
  scientific change, 729
Miller, David
  scientific progress, 751
  verisimilitude, 855, 856
Miller, Gary, 222
Miller, Richard, 226
Millikan, Robert A., 428–429
Millikan, Ruth G., 69
  cognitive science, 129
  function, 317, 319
Millikan-Ehrenhaft controversy, 428
Millikan oil drop experiment, 428–429
Mills, Susan
  biology, philosophy of, 69
  fitness, 312
  natural selection, 500
Millstein, R.L., 312
Milner, A.D., 548
Min, Chung-Ki, 59
Mind
  cognitive science, 123
  computational theory of, 30, 31
  computer model, Marr, 547–550
    consciousness, problem of, 549
    criticism from neuroscience, 548–549
    criticism from perceptual psychology, 548
    Gibson's attack on computationalism, 549–550
  evolutionary psychology, 263–267
  and function, 319–321
  information, terminology, 64
  perception, 545, 548
  philosophy of
    biological science, 69
    emergence, 231
    etiological function, 69
    neurobiology, 514
    Putnam and, 620, 623–625
    social science foundational problems, 783
  physicalism, 560
  psychology, philosophy of, 613–619
  Putnam and, 623–625
  science of, artificial intelligence, 31
  Searle, 767–774
  theories of, 61–62; *see also* Physicalism
  Turing, 833–834, 837
Mind, theory of, 617
Mind/body problem, 61–62, 63
  consciousness, 161
  generative grammar, conceptual issues in, 113–114
  physicalism, 558, 566–567
Mind-brain relationship, functionalist view, 548
Minimal belief change, Bayesianism, 52–55
Minimalist interpretation of Bell's theorem, 454
Minimalist program
  linguistics, philosophy of, 449–450
Minimality, variational calculus, 118
Minimal physical duplicate, physicalism, 564, 565
Minimal values, probability theory, 600
Minimax theorem, 507
Minimum viable populations, conservation biology, 164
Minkowski, Hermann
  Hilbert and, 356, 359
  space-time, 785–786, 790
Minkowski geometry, 790
Minkowski space-time, 785–786, 790
  causality, 790–791
  conventionalism, 172
  determinism, 205, 206
  quantum logic, 642
M-interpretation of QM, 141
Miracle argument, 1–2
Mirowski, P., 739
Mishkin, M., 516
Mishler, B.D., 799, 800
MIT bag model, 740, 747
Mitchell, S.D., 318
Mitochondrial DNA, 483
Mitosis, 331
Mittlelstaedt, Peter, 642
Mixed integer linear programming, 165
Mixing
  classical mechanics, 122
  dynamical systems, 203

Mixing conditions, ergodic theory, 507
Mixtures, improper, quantum measurement problem, 647
Modal interpretations
  quantum measurement problem, 647
  quantum mechanics, 657
Modality
  analyticity, 14
  Carnap, modal logic, 88
  determinism in quantum physics, 204–205
  laws of nature, 442
  sensory, 550
Modal operators, theory structure, 822
Modal statements, verisimilitude, 856
Modeling, cognitive, 616
Models, *see also* Scientific models; *specific models*
  artificial intelligence, 31–32
  causality, 96–97
  connectionism, 151–153
  determinism, 199
  ecological, 220–223
    evaluation of, 221–222
    explanatory value of, 222–223
    feasibility of, 220–221
  ecology and conservation biology, 72
  fitness, 313–315
  Kuhn's scientific revolutions, 428
  parsimony, 536–537
  particle physics, 538
  prediction, 593
  Ramsey, theory structure, 679–680
  reductionism, 698
  scientific change, 730–731
  scientific metaphors, 737–740
  theory structure, 826, 827
  unity and disunity of science, 846
  unity of science movement, 849
  visual representation, 867
Models, scientific, *see* Scientific models
Models and Analogies in Science (Hesse), 737
Model-theoretic models
  theory structure, 824–825
  verisimilitude, 855–856
Model theory
  Hilbert and, multiplicity of geometries, 360–361
  Putnam and, 621, 622
  Ramsey and, 680
Modernity, scientific revolutions, 755
Modern physics, locality, 451–452
Modularity, sequestered modular templates (SMT) model of cell, 488
Modular language learning, 396
Modular model of mind
  classical ethology, 265
  evolutionary psychology, massive modularity thesis, 266
  function, 319–320
  psychology, cognitive architecture, 616, 617
Moene, Karl Ove, 228
Moffitt, T.E., 399
Molecular basis of heredity, 353
Molecular biology, 480–490
  biological information, 64–68
  biology, philosophy of, 71–72
  classical, 482–483
  constitutive versus methodological, 480
  developmental biology, philosophy of, 72
  and evolution, 253–254, 256
  immunology, 367
  modern era, 485–486
  philosophical interpretations, 483–485
    information, 484–485
    reduction, 485
  philosophical speculations, 487–488
  population genetics, 582–583
  reductionism, 696, 699, 700
  scientific change, 731
  scientific revolutions, 755
Molecular chaos, kinetic theory, 418
Molecular clock, 481
Molecular facts, 722
Molecular genetics, 335–337, 480–481
  biology, philosophy of, 69
  molecular biology, 483
  molecular biology, philosophy of, 71
  population genetics, 583
  reductionism, 701
Molecular orbital theory, 104–105
  approximation, 25
  chemistry, philosophy of, 103, 104
Molecular shape rule, 483
Molecular structure, chemistry, 102, 103
Molecules, kinetic theory, 416, 418
Momentum
  quantum logic, 636
  quantum mechanics, 656
Momentum operators, von Neumann and, 506
Monism, neutral, 728, 844
Monistic century, 844
Monod, Jaçques
  biological information, 65
  fitness, 312
  genetics, 336
  molecular biology, 481
Monomorphic mind thesis, 266–267
Monophyly/monophyletic species, 799
Monotonicity, game theory, 324
Monster adjustment, Lakatos and, 437–438
Montague, Richard, 199
  determinism, 200
Monte Carlo methods, 810
Moore, C., 483
Moore, David S., 267
Moore, F.B.C., 582
Moore, George Edward
  analytical philosophy, 715
  phenomenalism, 551, 552
  Ramsey and, 671
  Russell and, 715, 716
  supervenience, 816
Moran, Dermot, 406, 407
Moran, P.A.P., 581
Morange, M., 65
Morgan, Mary
  scientific models, 744, 745, 746
  theory structure, 827
Morgan, Thomas H., 331–332, 353
Morgenstern, Oscar
  altruism, 10
  game theory, 323, 507
  probability, 605

Mormann, Thomas, 440
Morphemics, 107
Morphogenetic field, 232
Morphology
  heritability, historical development of ideas, 353
  innate/acquired distinction, 395
  phenotypic variation, 331
Morphospecies concept, 375–376
Morris, Charles
  logical empiricism, 463
  Nagel and, 492
  Reichenbach and, 705
  scientific progress, 750
  unity of science movement, 848
  Vienna Circle, 858
Morris, Desmond, 264
Morrison, D., 808
Morrison, Margaret, 273, 744
  scientific models, 745, 746
  theory structure, 827
  unity and disunity of science, 846
Moser, Jürgen, KAM (Kolmogorov-Arnold-Moser) theorem, 122
Mosini, V., 104
Motion
  classical mechanics
    Newton's laws, 117
    variational problem, 120
  instrumentalism, 400
  mechanics, 116
  space-time, 786
Motivational pluralism, 9, 11
Motive
  altruism, 8
  causality, 90
Moulin, Anne Marie, 364
Moulines, C.U., 824, 826
Mount, K.E., 635
Mückenheim, W., 600
Muller, Hermann J., 332, 334, 335
Multilayer feedforward networks, 151–152
Multilevel selection
  heritability, 355
  natural selection, 498
Multilocus models, population genetics, 579
Multimodal experiences, perceptions as, 550
Multiple alleles, reductionism, 698
Multiple criterion synchronization, conservation biology, 167
Multiple realizability
  explanation, 285
  physicalism, 562
  reductionism, 701–702
  unity and disunity of science, 845
Multiplets, spectral lines, 652
Mundale, Jennifer, 474, 514
Munévar, G., 309
Munitz, Milton K., 33
Munz, Peter, 261
Murphy, Nancey, 194
Murray, C.A., 354
Murray, F. J., 506
Murre, Jacob, 156
Musgrave, Alan, 421, 423
  prediction, 591
  scientific change, 730
Muslim astronomers, 33–34
Mutation, 338, 497
  genetics, 330
  natural selection, 501
  prediction, 596
Myerson, Emile, 101
Mysterium cosmographicum (Kepler), 34, 117

N

Naess, Arne, 858, 862
Nagel, Ernest, 491–496
  criticisms of Carnap, 492
  determinism, 199
  emergence, 233
  explanation, 278, 285
  function, 316, 318, 320
  instrumentalism, 404
  kinetic theory, 418
  logical empiricism, 463
  major articles, 492
  molecular biology, 484
  reductionism, 696, 697, 698
  scientific models, 747
  scientific revolutions, 759
  structure of science, 492–496
    foundations of physics and biology, 493–494
    general issues, 492–493
    social sciences and history, 494–496
  unity and disunity of science, 845
  unity of science movement, 848
  Vienna Circle, 858
Nagel, Jennifer, 242
Nagel, Thomas, 617, 618
Nagel-Schaffner model, molecular biology, 484
Naimark, M.A., Gelfand-Naimark-Segal (GNS) construction, 631
Naive inductive logic (NIL), 385–386
Nakamura, Y., 618
Names, proper, 446
Narrative, scientific revolutions, 755, 756
Narrow-sense heritability, 354
Nash, J., 326, 328
Nash equilibrium, 326, 327–328, 329
Naive set theory, 504
Nativism
  biological information, 66
  innate/acquired distinction, 395, 396–398
Natural economy, theory of, 511
Natural history, 253
Naturalism, *see also* Epistemology; Naturalized epistemology; Quine, Willard Van
  and empiricism, 239
  evolutionary epistemology, 258
  function, 321
  Mach and, 468
  Neurath and, 511, 513
  perception, 545
  Quine and, 660
  social sciences, 783
  Vienna Circle, 859
Naturalistic account of mind, 617
Naturalistic analysis, function, 315
Naturalistic linguistics, 448

Naturalistic theory of content, cognitive science, 131
Naturalized epistemology, 666–668; *see also* Epistemology; Evolutionary epistemology
Naturalizing function, 316–319
  mind, 320–321
Natural kind, *see also* Induction, problem of; Species
  confirmation theory, 146
  laws of nature, 442
  species, evolution of, 801–802
  systematic biology, philosophy of, 71
Natural language
  behaviorism, 62–63
  Carnap and, 86
  Chomsky and, 62–63, 114
  explication, 291
  innate/acquired distinction, nativism, 396
  Quine and, 664
  Reichenbach and, 709
  Russell and, 720
Natural law
  explanation, 276–278
  verifiability, 853
Natural measure functions, inductive logic, 388
Natural necessity, 177; *see also* Laws of nature
Natural philosophy
  classical mechanics, 115
  demarcation problem, 189
  Schlick, 726–729
  scientific change, 729
Natural sciences
  causality, 90
  heuristics, 435
  logical positivism and, 459–460
  Putnam and, 620, 624
Natural selection, 497–502
  abduction, 1
  adaptation and adaptationism, 3–7, 497–498
  altruism, 8
  chance in evolution, 500–501
  evolution, 255
  evolutionary biology, philosophy of, 70
  evolutionary epistemology, 257, 260
  evolutionary psychology, 263–267
  explanation, 501–502
  fitness, 310–315
    propensity interpretation of, 500
  function, 318
  gradualism, 498–499
  heritability, 352–355
  individuality, 376–377
  innate/acquired distinction, 395
  laws in evolutionary biology, 502
  population genetics, 578–585
  progress, 499–500
  species, 800
  units of selection problem, 498
Natural selection model of theory change
  evolutionary epistemology, 261
  Popper and, 575–576
Natural signs, sensations as, 545
Natural world, instrumentalism, 400
Nature
  balance of, 73
  experimental method, 271
  instrumentalism, 400
  laws of, *see* Laws of nature
  uniformity of, 379–380, 383
Nature of Physical Theory, The (Bridgman), 76
Nature of Selection: Evolutionary Theory in Philosophical Focus, The (Sober), 70
Nature of Truth According to Modern Logic (Schlick), 725
Natur und Kultur (Schlick), 728
Navier-Stokes equations, 118
N-body problem, classical mechanics, 121–122
Neander, Karen
  biology, philosophy of, 69
  function, 316, 317, 318
  natural selection, 501
Necessity, *see also* Laws of nature
  corroboration, 177–179
  experimental system interactions with environment, 270
  verifiability, 853
  visual representation, 869–870
Nee, Victor, 228
Needham, Joseph, 232
Negative facts, Russell and, 722
Negative holistic coherence theories, 247
Negative linear coherence theories, 247
Negative utilitarianism, Popper and, 575
Negligibility assumptions, scientific models, 741
Nei, Masatoshi, 254, 582
Nelson, Alan J., 598
Nelson, Lynn Hankinson, 297, 299, 300, 301
Nelson, R.J., 517
Nemeth, Elisabeth, 861
Neoclassical economics, ethical values, 228
Neo-Darwinianism, 253
Neo-Fisherians, population genetics, 582
Neo-Humeans, causality, 91, 94
Neo-Kantians
  rational reconstruction, 682
  Reichenbach and, 706, 707
  Schlick, 726
  unity and disunity of science, 844
Neo-Marxism, feminist philosophy, 297
Neo-Wrightians, population genetics, 582
Nerlich, Graham, 452
Nersessian, Nancy, 744, 867
Nervous system, *see* Neuroscience
Networks/network models, 467
  connectionism, 150–158
  molecular biology, 484, 486, 487, 488
  neurobiology, 521
  scientific change, 731
  von Neumann computation theory, 508
Neue Sachlichkeit, 513
Neumann, John von, *see* von Neumann, John
Neural computations, 27
Neural networks, 467, 731; *see also* Connectionism; Networks/network models
Neural plasticity, 549
Neural plausibility, 155
Neurath, Marie, 848
Neurath, Otto, 82, 83
  Carnap and, 82
  cognitive significance, 134
  Duhem thesis, 209
  empiricism, 238
  epistemology, 247
  feminist philosophy, 302

  Hahn and, 341, 342
  life and work, 510–514
    empiricism in social sciences, 510–512
    scientific language and scientific method, 512–513
    unity of science and encyclopedia model, 513–514
  logical empiricism, 459, 462, 463
  logical positivism, 460
  Nagel and, 492
  physicalism, 558
  protocol sentences, 610, 611, 612
  Quine and, 662
  Reichenbach and, 704
  Schlick and, 726, 728
  scientific progress, 750
  unity and disunity of science, 844–845
  unity of science movement, 848
  Vienna Circle, 858, 859, 860, 862
Neurath principle, 513
Neurons, single-cell recordings, 514–516
Neuropsychology
  connectionism, 156
  consciousness, 160
  evolutionary psychology and, 265
Neuroscience, *see also* Cognitive science
  biological information, terminology, 64
  cognitive science, 123, 127, 130
  connectionism, 151, 155–157
  Mach and, 467
  mechanisms, action potential, 471–472
  molecular methods, 480
  neurobiology, 514–522
    ethics, 521
    localization and reduction, 516–518
    representation in brain, 520–521
    theories, 518–520
    theory-laden observations and single-cell recordings, 514–516
  perception, 545
    Marr's computer model of mind, 548–549
    neural correlates of consciousness, 550
    neural plasticity, 549
  physicalism, 560
  psychology, philosophy of, 614, 615
    cognitive architecture, 616, 617
    consciousness, 617
  von Neumann and, 508–509
Neutral analogy, 742
Neutral (null) models
  ecological communities, 222
Neutral monism, 728, 844
Neutral theory
  of molecular evolution, 582
  population genetics, 582–583, 584
Neutrinos
  leptons, 539
  solar, 528
Newell, Allen
  artificial intelligence, 28, 29, 31
  psychology, philosophy of, 615–616
New epistemology, 59
New Essays (Leibniz), 395
Newman, Max, 833, 834, 835, 836
Newman, M.E.J., 488
Newmeyer, Frederick, 107
New objectivity (*Neue Sachlichkeit*), 513
Newton, Isaac
  astronomy, 34, 35
  causality, 92–93
  classical mechanics, 115
  crucial experiment (*experimentum crucis*) idea, 209
  demarcation problem, 189, 191
  explanation of Kepler's laws, 278
  Hanson and, 346
  instrumentalism, 403
  Kuhn's scientific revolutions, 422, 423–424, 427
  locality/local action, 451
  Mach and, 468
  parsimony, 532
  perception, 546
  prediction, 594–595
  research programs, 713–714
  scientific domains, 734, 735, 736
  scientific metaphors, 739
  scientific revolutions, 754, 756, 757, 760
  space-time, 788, 791, 792
  unity and disunity of science, 843–844
Newtonian mechanics/systems, 25, 650
  approximation, 25
  classical mechanics, 117–118
  complementarity and, 142
  conventionalism, 171, 173
  Feyerabend and, 306
  generalized field theory, 359
  instrumentalism, 400
  Kuhn on
    language, 428
    scientific revolutions, 421, 422, 424, 427, 428
  Laplace on, 501
  Mach and, 468
  methodological individualism, 479
  Nagel on, 493
  parsimony, 532
  pendulum and, 423
  realism, 692
  reductionism, 698
  research programs, 714
  scientific revolutions, 758–759, 760
  space-time, 786–788, 791
    and general relativity, 792, 793
    and special relativity, 788, 791
  underdetermination of theories, 841
  unity and disunity of science, 845
Newtonian telescope, 34
Neyman, Jerzy
  Bayesianism, 43
  statistics
    Bayesian advances and controversy, 810
    confidence interval estimation procedures, 806
    inductive behavior philosophy, 805–806
    Neyman-Pearson (NP) tests, 806
    power analytic movement, 809, 810
    stopping rule, relevance of, 808
Neyman-Pearson (NP) tests, 804–805, 806, 814
  Bayesian advances and controversy, 810
  Bayesianism, 43, 44, 45
  Carnap, 88
  power, 809
  statistics, 804–805
    error probability principle versus likelihood principles, 807
    Fisher and, 806

Neyman-Pearson-Wald approach, 807
Niche, ecological, 73, 222
Nickles, Thomas
  demarcation problem, 190, 192
  emergence, 234
  reductionism, 698, 699
Nicod, Jean, 348
  confirmation theory, 145
  induction, problem of, 382
  Ramsey and, 679
Nicod's criterion, 145, 348, 382
Nidditch, P., 135
Nietzsche, Friedrich, 257
Nihilism, 375
Niiniluoto, Ilka
  Bayesianism, 51, 52
  inductive logic, 389
  probability theory, 604
  scientific progress, 751, 752
  verisimilitude, 855, 856, 857
Nixon, K.C., 799
No class theory of classes, 718
No conspiracy condition, Bell's locality condition with, 454
Noether, Emmy, 119
No-go theorem, 453
Nomic necessity
  laws of nature, 442, 443
  verifiability, 853
  verisimilitude, 856
No miracle argument (NMA), 241
Noncoding regions of genome, 486
Noncognitivist ethics, Reichenbach, 704
Noncomparative approaches, confirmation theory, 145
Non-constant sum games, 507
Non-cooperative games, 325–326
Nondoxastic theories, epistemology, 245, 247
Nonessential consciousness, 158–159, 160
Non-Euclidean geometry, 80, 357
  and empiricism, 237
  Hertz and, 120–121
  instrumentalism, 401
  Poincaré and, 568, 569, 570
  Reichenbach and, 706
  underdetermination of theories, 841
Nonexistence, intentionality and, 406
Nonglomerability paradoxes, 59
Nonintentional consciousness, 159, 160
Nonlinear function mapping, emergence, 232
Nonlinear geometry, chemistry, 103
Nonlinear problems
  scientific computing and, 508
  von Neumann's numerical analysis of partial differential equations, 503
Nonlocality
  gravitation, 451
  quantum field theory, 456–457, 632
  quantum mechanics, 656
  subtleties of, 456
  time, presentism, 831
  Turing and, 835
  verifiability, 852
Nonlocality of explanation, 284
Non-Mendelian inheritance, 486, 488
Nonmental, physicalism, 561
Non-null hypothesis, 804–805
Nonprobabilistic approaches, confirmation theory, 145–148
Nonreduction
  causality, 91–92
  chemistry, philosophy of, 103
  probability theory, 607
Nonrelativistic quantum mechanics, quantum logic, 633
Nonscientific knowledge, Putnam and, 620
Nonseparability, *see* Emergence
Nordheim, L., 505
Normal (strategic) form, game theory, 324, 325
Normal/Gaussian law of errors, 56
Normal science
  cognitive science, 124
  Kuhn's disciplinary matrices and scientific revolution, 423–424
Normative concepts/normativity
  biology and, 69
  epistemology, 667
  experimental method, 274–275
  feminist philosophy, 297, 300
  function, 316, 320–321
  intentionality, 409
  linguistic deficits, 109, 111
  rational choice program, 781
  research programs, 712, 715
  scientific change, 731
  scientific revolutions, 755, 764
  social constructionism, 778
  social sciences, 783, 784
Normative function of visualization, 707
Northeim, L., 360
Norton, B.G., 163, 164
Norvig, P., 619
Notation
  Hanson, on facts, 345–346
  neurobiology of computation, 509
Novelty
  emergence, 231
  prediction, 589–592
  realism, 689–690
  research programs, 714–715
Novick, M., 59
Nowak, Leszek, 747
Nozick, Robert, 228, 244
Nuclear forces, particle physics, 539
Nuclear physics
  observation, 525–526, 527
  particle physics, 538
Nucleic acids, 334–335, 480
Nucleon, models of, 740
Nuisance parameters, Bayesianism, 56
Null hypothesis, 45, 809, 812
Null models
  ecological communities, 222
  population genetics, 584
Number, cognitive architecture, 617
Numeration of lexical items, 109
Numerical analysis of nonlinear partial differential equations, von Neumann and, 503
Numerical pheneticists, 796–797
Numerical representations, probability theory, 600
Nussbaum, Martha Craven, 228
Nyberg, L., 517
Nye, M.J., 102, 103

O

Oberdiek, Hans, 272
Objective chance, probability, 606–609
Objective knowledge
  Popper and, 575–577
  protocol sentences, 610
Objective truth, Poincaré and, 571
Objectivism, scientific change, 730
Objectivity
  Bohr, 143
  complementarity, 143
  Kuhn's scientific revolutions, 426–427
  observation, 527
  scientific revolutions, 764
  scientific style, 766
  scientific style and, 765
  strong, feminist perspective, 301
Object of study
  generative grammar, conceptual issues in, 110
  linguistics, philosophy of, 444–446
Objects
  causality, ontological categories, 90
  complementarity, 142
  domains of, 80
  ecological community as, 216
  intentionality, 406
  phenomenalism, 551–553
  scientific models, 743
Obscurity objection, physicalism, 560, 561
Observability, quantum field theory, 630–631
  phenomena, instrumentalism, 400–405
  phenomena, realism, 691–692
  rational, 529
Observables
  decoherence, 649
  quantum logic, 636
  quantum mechanics, 653
Observation, 523–530
  American Structuralism, 107
  Carnap, 523–524, 525
  Carnap on, 83
  classical empiricist program, fundamental weaknesses of, 527–528
  cognitive significance, 135, 136
  conclusions and implications, 528–530
  empiricism, 240
  experimental method, 274
  failures of empiricist program, 524–525
  Hanson on, 344–345
  Hempel on, 350–351
  incommensurability, 371
  instrumentalism, 402, 403
  Kuhn's scientific revolutions, 424
  likelihood approach to evaluation of, 533
  neurobiology, theory-laden, 514–516
  Popper and, 572, 573
  prediction, 588, 593
  quantum field theory, 629, 630–631
  quantum mechanics, 658
  Quine and, 659–660, 665, 666–668
  Ramsey, theory structure, 679
  realism, 687
  reductionism, 402
  Reichenbach and, 705
  research programs, 712
  scientific domains, 732–734, 735, 736
  social constructionism, 777
  theory structure, syntactic (received) view, 823
  underdetermination of theories, 841
  verifiability, 852
Observational-theoretical distinction
  instrumentalism, 404
  observation, 524–525
  realism, 687
  scientific domains, 734
  theories, 822–824
Observation language
  Carnap on, 86
  Feyerabend on, 305
  theories, 822–824
  verifiability, 852
Observation sentences/statements, *see also* Protocol sentences
  defined, 133
  Kuhn's scientific revolutions, 424
  physicalism, 558
  Quine and, 666–668
  theories, 822–824
  verifiability, 852
Observer/observed complementarity, 141–143
Occam's Razor (Hahn), 343
Occam's razor, 531–538
Occasion sentences, Quine, 666
Occupation number representation, quantum field theory, 631
O'Connor, Timothy, 197
Oddie, Graham, 855, 856
Odenbaugh, Jay, 220
O'Donnell, John M., 61
O'Grady, P., 663
Oken, Lorenz, 353
Okruhlik, Kathleen, 302
Olby, R.C., 480
Old evidence problem, 150, 590
Omniscience, Hahn on, 342–343
On Aggression (Lorenz), 263
Oncogenes, 338
'On Denoting' (Russell), 721
One-particle state space, 630
'On Intuition' (Hahn), 343
On Liberty (Mill), 307
On the Origin of the Species (Darwin)
  evolution, 252
  evolutionary epistemology, 257
  heritability, 353
  natural selection, 497
  population genetics, 579
On the Revolutions of the Celestial Spheres (Copernicus), 400
Ontic formulation of Laplacian determinism, 200
Ontic levels versus mechanistic levels, 472
Ontogeny
  evolution of epistemic mechanisms (EEM), 259
  evolution of epistemic theories (EET), 261–262
  phylogenetic evolution versus ontogenic development, 258
Ontology
  Carnap, 86, 88
  causality, 90
  chemistry, philosophy of, 102, 103, 105
  classical mechanics, 115, 118–119
  cognitive significance, 137

Ontology (*Continued*)
  emergence, 230
  empiricism, 236, 243
  experimental action and production, 270–271
  experimental method, 274
  explanation, 284–285
  explication, 287
  kinetic theory, 415, 418
  laws of nature, 439
  logical empiricism, 463–464
  Mach, 468
  methodological individualism, 478–479
  molecular biology, philosophy of, 71
  parsimony, 532
  population biology probabilities, 70
  quantum field theory, 631
  quantum mechanics, 142
  Quine and, 660
  realism, 687
  scientific models, 740, 743–744
  Searle and, 770
Open Society and Its Enemies, The (Popper), 575
Operant conditioning, 61
Operationalism, 62
  Bridgman, Percy, 75–77
  cognitive significance, *see* Cognitive significance
  demarcation problem, 191
  experimental method and, 269
  quantum logic, 642
  scientific revolutions, 758
  verifiability, 852; *see also* Verifiability
Operator theory, quantum mechanics, 504, 507
Operon model, molecular biology, 483
Oppenheim, Paul, 94
  causality, 95
  explanation, deductive-nomological (D-N) model, 276, 278–279
  inductive logic, 393
  laws of nature, 439
  mechanism, 472
  Putnam and, 620
  scientific models, 747
  unity and disunity of science, 845
Oppenheimer, Robert, 539
Opticks (Newton), 93, 734
Optics
  experimental tradition, 268
  mechanics and, 120
  scientific domains, 734
Optimal adaptation, 5
Optimality, evolutionary biology, 70
Oracle, Turing, 834, 837
Orbital angular momentum, 652
Orbital model, chemistry, 103, 104–105
Order of inquiry, American Structuralism, 107
O'Regan, J.K., 548
O'Reilly, Randall C., 156
Organicism, emergence, 231, 232–233
Organic molecules, astrochemistry, 38
Organism
  ecological communities, 216–217
  immune self, 365–369
  individuality, 375
  psychology, philosophy of, 619
  sexual versus asexual reproduction, 376
Organismic biology, 480, 481
Organization, reductionism, 697–698
Organizational aspects of mechanisms, 477
  aspects of organisms, 470–471
  discovering, 474–475
  neurobiology, 514
Original intentionality, 408, 409
Origin of Species (Darwin), 4
Origin of species, de Vries' *Mutationstheorie*, 330
Origins of Modern Science (Butterfield), 754
Orilia, F., 287
Orr, H. Allen, 254
Orthoalgebras, 640, 643
Orthologous genes, 338
Orthomodular lattice logic, 638–639
Orthomodular lattice of linear subspaces, 635
Orzack, Steven Hecht
  adaptation and adaptationism, 5
  evolutionary biology, philosophy of, 70
  natural selection, 498
Osborne, M.J., 326
Oscillators, black body radiation, 651
Oscillatory cycles in the brain, 617
Osiander, Andreas, 400
Ostension, incommensurability, 372–373
Ostwald, Wilhelm, 513
  chemistry, philosophy of, 101
  classical mechanics, 119
  unity and disunity of science, 844
Ott, Edward, 595
Otte, D., 796
Our Knowledge of the External World (Russell), 80, 552
Outcomes
  Dutch Book argument, 212
  game theory, 323
Owen, Richard, 252
Oxtoby, J.C., 504

P

Page, M., 153
Pagel, Mark, 7, 498
Pain, 618
Pairwise interactions, ecological communities, 217
Palmer, S., 129
Paneth, F., 102
Pangenes, 330
Pannekoek, A., 34, 35
Papineau, D.
  cognitive science, 129
  physicalism, 561, 567
Paradigms
  cognitive science, 123
  Kuhn, 420–421, 429
  linguistics, Chomsky and, 107
  Quine and, 668
  scientific progress, 751–752
  scientific revolutions, 760
  scientific style, 765, 767
Paradoxes
  Bayesianism, 58, 59
  causality, 95–96
  confirmation theory, 145, 146
  decision theory, 186, 187

  of intentionality, 406–407
  Poincaré and, 570
  Ramsey, 672–674
  Ramsey and, 673–674
  realism, 692
  space-time, 791
  Zeno's, 292
Paradoxes of confirmation, 145, 383
Parallel architecture, neural machinery of brain, 509
Parallel distributed processing (PDP), 150–151
Parallel mechanisms, 471
Paralogous genes, 338
Parameter independence and interdependence, locality, 455
Parameters, generative grammar, 109
Parametrization, chemistry, 105
Pareto optimality, game theory, 328
Pargetter, Robert, 318, 440
Paris, J., 387
Parity, 540
Park, Robert, 195
Parsimony, 531–538
  Akaike and selecting models, 536–537
  Bayesianism, 534–535
  local versus global accounts, 532–534
  Popper and falsifiability, 535–536
  simplicity and, 534
Parsons, Talcott, 782
Parsons, William, 35
Partial Boolean algebras, quantum logic, 633, 638, 639–640
Partial differential equations, classical mechanics, 120
Partial entailment, 384
Particle density, quantum field theory, 631
Particle interpretation of quantum mechanics, Reichenbach and, 710
Particle ontology, quantum field theory, 631
Particle physics, 538–544
  anthropic principle, 21–23
  early, 538–540
  physical sciences, philosophy of, 557
  quantum field theory, 542–544, 629, 631
    interacting fields, 632
    routes to, 630–631
    and standard model, 542–544
  realism, 687
  S-matrix program and particle democracy, 541–542
  standard model, 22, 542–544
  symmetries and particles, 540–541
  underdetermination of theories, 841
  unity and disunity of science, 846
Particles
  kinetic theory, 416
  quantum field theory, 631
Particulars, individuals, 374
Partition exchangeability, 52
Partition vector, 52
Pascal, Blaise
  Fermat, correspondence with, 599
  probability theory, 602
Pascal's sentence, 357
Pasch, M., 357, 358
Patents/intellectual property, 274
Path analysis, causality, 94–95
Path integral approach, 120
  quantum field theory, 631
Pathology, psychology, 618
Patrides, C.A., 395
Pattern cladists, 799
Patterns of Discovery (Hanson), 427–428
Paul, D.B., 334
Paul, Laurie, 100
Pauli, Wolfgang
  particle physics, 539, 540
  quantum field theory, 631
Pauli exclusion principle, 104
Pauling, Linus, 254
  molecular biology, 482
  population genetics, 582
Pauling model of chemical bonds, 740
Pavicic, Mladen, 634, 639
Peacocke, C., 546
Peano, Giuseppe
  Ramsey and, 673
  Russell and, 717
Peano arithmetic, 357, 358
Peano axioms, 290
Pearce, D.A., 826
Pearl, Judea, 96, 97, 100, 616
Pearson, Karl, 578
  causality, 94–95
  scientific change, 729
  statistics
    chi-square goodness-of-fit test, 804
    inductive behavior philosophy, 805–806
    Neyman-Pearson (NP) tests, 806
    power analytic movement, 810
    stopping rule, relevance of, 808
Peirce, Charles Sanders, 45, 257, 589
  abduction, 1
  Bayesianism, 43
  Chomsky and, 112, 113
  cognitive science representational assumption, 128
  Ramsey and, 671, 680
  scientific change, 729
  statistics
    error probability philosophy, 803
    testing statistical hypotheses, 43
Pellegrini, A.D., 267
Pellionisz, Andras, 151
Pendulum, 792
  Copernican/Newtonian physics, 423
  Kuhn's scientific revolutions, 428
Penrose, Roger, 835, 837
  irreversibility, 412
  particle physics, 543
Perception, 545–550
  experimental method, 274
  Hanson, on observation, 344–345
  innate/acquired distinction, 395–396
  Marr, computer model of mind, 547–550
    consciousness, problem of, 549
    criticism from neuroscience, 548–549
    criticism from perceptual psychology, 548
    Gibson's attack on computationalism, 549–550
  observation, 524, 526–527
  outstanding problems, 550
  overview of philosophy of, 545–546
  phenomenalism, 551–553
  prediction, 588
  protocol sentences, 611
  psychology, philosophy of, 618

Perception (*Continued*)
  Putnam and, 623
  science versus common sense, 546–547
  Searle and, 770
  sense data and argument from illusion, 547
  states, experience, sensation, 545
  von Neumann and, 505–506
Perceptual/articulatory system, linguistics, 449
Perceptual content, epistemology, 245–246
Perceptual psychology, Marr's computer model of mind, 548
Perceptual sets, Hanson, 344
Peres, Asher, 454, 456
Perfect information, game of, 324, 325
Performance level
  artificial intelligence, 31
  linguistics, 110
Perini, Laura, 869
Periodic motion, canonical transformation, 120
Periodic table, 103–104
Perlman, Mark, 3, 69
Permutations, state descriptions, 388
Perrin, Jean, 417
  kinetic theory, 418
  scientific domains, 733
Persistance, ecological communities, 218
Perspectivalism, 832
Perturbation series, celestial mechanics, 118
Perturbation theory, quantum field theory, 632
Perturbative renormalization techniques, 543, 544
Peters, R.H., 221
Peters, Robert, 220
Petroni, N., 600
Pettis, B.J., 504
Pfister, Herbert, 117
Phase space
  classical mechanics, 121
  irreversibility, 411–413
  neurobiology, 521
  and quantum logic, 633–634
  reductionism, 698
  statistical mechanical entropy definition, 412–413
Phenetic species concept, 796–797
Phenomena
  mechanisms, 469, 475, 477
  scientific models, 741–743
Phenomenal consciousness, 161, 162, 618
Phenomenalism, 551–554
  Carnap, 80, 82, 84
  empiricism, 239
  historical background, 551–552
  idealism and realism, 552–553
  instrumentalism, 400
  perception, 545–550
  physicalism, 558
  prediction, 586, 588
  unity and disunity of science, 844
Phenomenal versus access consciousness, 159–160
Phenomenal world, space-time, 786
Phenomenological models, 740, 742–743, 747
  particle physics, 538, 540, 541, 543, 544
  unity and disunity of science, 844
Phenomenology
  Hanson, on observation, 345
  social constructionism, 774
Phenotypic space, 795–796
Phenotypic traits, 331
  adaptation and adaptationism, 4, 5, 6–7
  variance, 354
Philips curve, 741
Phillips, P.C., 582
Philo, 252
Philosophiae naturalis principia mathematica (Newton), 34, 93, 118, 189
  classical mechanics, 117
  parsimony, 532
  prediction, 594
Philosophical Foundations of Quantum Mechanics (Reichenbach), 708, 709–710
Philosophical Investigations (Wittgenstein), 344, 783
Philosophical relativity, 706
Philosophy
  cognitive science multidisciplinary approach, 123
  Quine, 660–666
Philosophy of action, determinism, 197
Philosophy of mind, *see* Mind, philosophy of
Philosophy of Natural Science (Hempel), 350
Philosophy of science
  Russell and, 719–722
  scientific revolutions, 758–760
  Vienna Circle, 859–860
Philosophy of Space and Time (Reichenbach), 704, 705
Phlogiston theory, 102
Phonemics, 107
Phonetic form of representation, 109
Photoelectric effect, 651
Photon density, quantum field theory, 631
Photons
  leptons, 539
  particle physics, 538
  quantization, 651
Phylogenetic cladists, 799
Phylogenetic evolution versus ontogenic development, 258
Phylogenetics
  parsimony, 532
  species, 798–799
Phylogeny, 71
  evolution of epistemic mechanisms (EEM), 258–259
  evolution of epistemic theories (EET), 259–261
  molecular methods, 253
  reproductive isolation, 798
Physical chemistry, approximation, 25–26
Physical definitions, Reichenbach and, 705
Physical domain of objects, 80
Physical environment, consciousness, 160
Physicalism, 558–568; *see also* Mind/body problem
  behaviorism, 62
  Carnap, 82–83, 84
  causality, 91
  consciousness, 161
  emergence, 230–235
  formulation and justification, 559
  generative grammar, conceptual issues in, 113–114
  identity theses, 562–563
    disjunction option, 562–563
    multiple realizability claim, 562
    property identity claim, 562
    quasi-eliminativist option, 563
    trope identity option, 563
  intentionality, 408, 409
  justification, question of, 565–567

    - causal impact argument, 566–567
    - manifestability argument, 567
  - Kuhn's scientific revolutions, 424
  - Neurath and, 513
  - perception, 549
  - physical, specification of, 559–560
  - physics and the physical, 560
  - prediction, 586
  - protocol sentences, 611, 613
  - realizationism, 565
  - Schlick and, 728
  - Searle and, 770
  - skeptical worries, 560–561
    - responses to, 561–562
  - supervenience theses, 563–565
  - verifiability, 851
  - Vienna Circle and, 859
- Physicalism (material-thing-language)
  - Neurath and, 513
  - unity and disunity of science, 844–845
  - Vienna Circle and, 859
- Physical necessity, *see* Laws of nature
- Physical objects
  - phenomenalism, 551–553
  - scientific models, 743
- Physical reductionism, molecular biology, 485
- Physical sciences, philosophy of, 554–557
  - foundations and interpretations of physical theories, 554–555
  - general philosophy of science, 555–556
  - metaphysics, 557
  - physics centrism, 556–557
- Physical space, 80
- Physical systems, kinetic theory, 415–419
- Physical theory
  - chemistry, philosophy of, 102, 105
  - complementarity, 143
  - determinism, 207
  - physicalism, 559–560
  - and theory of mind, 113–114
- Physics
  - anthropic principle, 21–23
  - approximation, 25, 26
  - Bridgman, Percy, 75–76
  - Carnap
    - discursive domains, 80
    - syntactic phase, 83–84
  - chemistry, philosophy of, 104
  - determinism, *see* Determinism
  - experimental method, 269, 270
  - Hilbert, 359–360
  - Mach, 468
  - mechanics and, 116
  - Nagel, structure of science, 493–494
  - observation, 525–526
  - particle, *see* Particle physics
  - physicalism, 559–560
    - causal impact argument, 566–567
    - manifestability argument, 567
  - physical sciences, philosophy of, 554–557
  - prediction
    - classical mechanics and chaotic systems, 594–595
    - quantum mechanics, 595–596
  - probability, 607
  - psychology, cognitive architecture, 617
  - quantum field theory, 629–633
  - quantum measurement problem, 644–649
  - realism, 692–693
  - reductionism, 696, 697, 698, 699, 700–701
  - scientific domains, 736
  - scientific models, 740, 747
  - scientific revolutions, 754, 755, 758, 760
  - scientific style, 766
  - space-time, 785–795
  - standard model of particle physics, 22
  - unity and disunity of science, 844, 846
- Physiological events, protocol sentences, 611
- Physiology
  - experimental science, 268
  - molecular methods, 480
- Piaget, Jean, 257
  - evolutionary epistemology, 261–262
  - scientific revolutions, 757
- Piatelli-Palmerini, M., 114
- Piatesky-Shapirom G., 731
- Piazza, A., 267
- Pickering, Andrew, 777, 779
- Picture theory of language, Vienna Circle, 859
- Pietroski, Paul, 210
- Pimm, Stuart, 218, 219, 221, 223
- Pinch, Trevor
  - experimental method, 269, 272
  - social constructionism, 776
- Pinker, Steven, 4
  - evolutionary psychology, 266
  - linguistics, philosophy of, 445
- Pinnick, Cassandra L., 784
- Pinto-Correia, C., 353
- Pioneer stage of ecological succession, 215
- Pittendrigh, C.S., 69
- Pitts, Walter, 27, 508
- Place prioritization, conservation biology, 164–165
- Planck, Max
  - classical mechanics, 119
  - kinetic theory, 417
  - Kuhn's scientific revolutions, 428
  - Mach and, 467
  - quantum mechanics, 650–651
  - Reichenbach and, 704
  - Schlick and, 725
  - scientific style, 765
  - unity and disunity of science, 844
- Planck's constant, 141–142, 428, 651
- Planck length, 543, 544
- Planetary motion, 34
  - classical mechanics, 116, 117–118
  - KAM (Kolmogorov-Arnold-Moser) theorem, 122
- Plantinga, A., 244
- Plate tectonics, 755
- Plato, 545, 843
  - astronomy, 33
  - innate/acquired distinction, 394, 395, 396
  - Popper and, 575
- Platonism, Cambridge, 395
- Platonism, scientific revolutions, 756
- Pleiotropy, gene, 331
- Plotkin, Henry, 262
- Plücker line geometry, 170
- Pluralism
  - motivational, 9, 11
  - species, 800–801

Pluralism (*Continued*)
  theoretical
    ecology, 221
    Feyerabend, 307–308
  units of selection question, 70
Plurality, scientific style, 765
Podolsky, Boris, Einstein-Podolsky-Rosen relation, 452–453
Podolsky, Scott H., 366, 367, 368
Poincaré, Henri, 568–571
  Carnap, 82
  classical mechanics, 115, 117, 121
  conventionalism, 169–170, 171, 174, 175
  Duhem thesis and, 209
  geometric conventionalism, 568–569
  Hahn and, 341
  Hilbert and, 360
  instrumentalism, 401, 402, 403
  intuition in pure mathematics, 569–570
  Neurath and, 512
  Popper and, 577
  Reichenbach and, 705, 707
  scientific realism, 570–571
  Vienna Circle and, 859–860
Poincaré recurrence, classical mechanics, 121
Pointer observable, 646, 647, 648, 649
Point mass approach, defined, 118
Point of view
  complementarity, 143
  finite, Hilbert and, 360–361
Poisson bright spot, 590
Poland, Jeffrey, 561
Polanyi, Michael
  chemistry, philosophy of, 101
  emergence, 233
  scientific revolutions, 761
  Turing and, 836
Political economy
  of development, 229
  expanding economy model, 508
Political identity, language and, 444–445
Political issues
  neurobiology, 514
  public policy
    demarcation problem, 195–196
    species concept, importance of, 801
Political philosophy, *see also specific individuals*
  atomistic body, 364–365
  economics, 227–228
  Popper and, 575
  Russell and, 720
  Schlick and, 728
Politics, *see also* Sociopolitical contexts
  demarcation problem, 188–197
  ecological models and, 223
  game theory, 507
  genetics/eugenics, 334
  Lakatos and, 433, 434
  logical empiricism, 461, 464
  Mach, 468
  Russell-Einstein Manifesto, 716
  Schlick and, 728
  scientific revolutions, 754
  species concept, importance of, 801
  unity of science movement, 849
  Vienna Circle, 83, 860
Politics, academic
  controversies, *see* Controversies/disputes, scientific
    evolution, 254
Pollack, Jordan B., 154
Pollock, J.L., 245, 246, 247, 250
Polya, George
  induction in mathematics, 43
  Lakatos and, 433–434, 435, 436
Polymerase chain reaction (PCR), 483
Polythetic species taxa, 795–796
Poole, C., 811
Pooley, Oliver, 117
Popescu, Sandu, 456
Popper, Karl R., 571–578
  Bayesianism, 43
  Carnap and, 83, 88
  and corroboration, 45, 177–179
  demarcation problem, 188–197
  determinism, 198, 199, 200, 207
  Duhem thesis and, 208–209
  emergence, 231
  evolutionary epistemology, 257, 258–259, 260, 261
  experimental method, 271, 273
  falsificationism and methodological rules, 574–575
  Feyerabend and, 304
  Hahn and, 341, 342, 343
  inductive logic, 382, 386, 392, 393
  instrumentalism, 400
  Kuhn and, 423, 425
  Lakatos and, 434, 436–437, 438
  later developments: objective knowledge, third world, and verisimilitude, 575–577
  life and work, 572
  Neurath and, 513
  parsimony, 534, 535–536
  Poincaré and, 570
  prediction, 586, 591, 598
  probability theory, 600, 607
  protocol sentences, 610
  Quine and, 666, 668
  research programs, 712–715
  science, history, and society, 575
  science and metaphysics, 572–574
  scientific change, 730
  scientific progress, 750–751
  scientific revolutions, 759
  scientific style, 766
  social sciences, 780, 781, 782, 784
  statistics
    error probability philosophy, 803
    testing statistical hypotheses, 43
  unity and disunity of science, 845
  verifiability, 852
  verisimilitude, 854, 855–856
  Vienna Circle, 858
Population biology
  biology, philosophy of, 69
  conservation biology, 163–168
  ecological communities, 216–217
  evolution, 256
  evolutionary biology, philosophy of, 70–-71
  fitness, 310–311
  innate/acquired distinction, 395, 399
  molecular biology, 481
  natural selection, 501

Population ecology, 215
Population genetics, 578–585
  approximation, 24
  biological information, 66
  defined, 310–311
  and evolution, 253, 254–255
  fitness, 310–315
  genetics, 333–334
  heritability, 352–355
  history of, 579–582
    correlation between relatives, 580–581
    fundamental theorem, 581
    Hardy-Weinberg law and maintenance of variation, 579–580
    Wright versus Fisher, 582
  molecular biology, 481
  molecular biology and neutral theory, 582–583
  prediction, 596
  retrospective models, molecular genetics, and coalescence theory, 583
  species, reproductive isolation, 797–798
Populations
  clonal, as individuals, 375
  ecological communities, 219–220
Population size, population genetics, 583
Population viability analysis, 166–167
Port, R., 616
Port Royal logic, 599, 605
Position
  quantum logic, 636
  quantum mechanics, 656, 657
  visual representation, 865
Position effects, genetics, 332
Position operators, von Neumann and, 506
Position space, quantum field theory, 630
Positive holistic coherence theories, 247
Positive-instance criterion, confirmation theory, 145
Positive linear coherence theory, 247
Positivism, *see also* Logical empiricism; Logical positivism
  classical mechanics, 119, 121
  empiricism, 237–238
  empiricism after, 239–241
  immunology, 364
  kinetic theory, 417
  Mach, 468
  physical sciences, philosophy of, 556–557
  protocol sentences, 610
  Putnam and, 624
  rational reconstruction, 682
  Reichenbach's realism and, 708
  scientific change, 729
  scientific progress, 752
  scientific revolutions, 755, 756
  social constructionism, 774–775
  Vienna Circle, 859
Positron emission tomography, 616, 617
Positrons, 539
Posits (Reichenbach), 708, 709
Possible world accounts, laws of nature, 440–441
Posterior probability, 803, 857
Postmodernism and science, 778
Potential explanation, deductive-nomological (D-N) model, 276
Potential models, theory structure, 826
Potter, Elizabeth, 297, 299, 300, 301
Poulton, R., 399
Poverty of Historicism, The (Popper), 575, 780
Powell, Walter, 228
Power, N-P tests, 805
Power analytic movement, statistics, 809–810, 811
Practical mechanics, defined, 116
Pragmatic accounts/approaches
  Carnap, 83
  Duhem thesis, 209, 210
  empiricism, 235
  explanation, 282–283
  induction, problem of, 381–382
  inductive logic, 392
Pragmatism
  abduction, introduction of term, 1
  anthropic principle, 22
  demarcation problem, 194
  evolutionary epistemology, 257
  instrumentalism, 400
  logical empiricism, 461
  protocol sentences, 612
  Quine and, 660, 663
  Schlick and, 726, 727
  Vienna Circle, 859
Prandtl, Ludwig, 118–119
Pratt, J.W., 813
Precision
  explication, 292
  statistics, initial versus final, 807
Predation, ecological communities, 217
Predicate first-order calculus (first-order functional calculus; first-order logic), 13
  Hempel, 350
  quantum logic, 633, 635–636, 641
Predicate logic
  Reichenbach and, 709
  Russell and, 715
Predicates
  corroboration, 177
  disjunctive, 562
  dispositional versus projectible, 177
  explication, 288
  induction, problem of, 383
  set theory predicate-extensions, 504–505
Predicative functions, Ramsey, 674
Predicative mathematics, Poincaré, 570
Prediction, 585–599
  Bayesianism, 42, 45
  chemistry, philosophy of, 103
  classical mechanics, 119
  cognitive significance, 135
  determinism, 198, 200, 204, 206–207
  ecological community models, 220–221
  ecological models, 221–222
  economic theories, 227
  empirical, models of, 587–589
  empirical sciences, 594–598
    biological and social sciences, 596–598
    classical mechanics and chaotic systems, 594–595
    quantum mechanics, 595–596
  epistemic significance of, 589–592
  explanation and, 277, 348
  functional analysis, 317

Prediction (*Continued*)
  Hempel and, 348, 351
  instrumentalism, 400
  methods of, 592–594
  model selection, 537
  Neurath and, 514
  particle physics, 538, 540
  physical sciences, philosophy of, 555
  problem of induction, 586–587
  quantum logic, 637
  Quine and, 662
  realism, 688, 689–690, 691
  relativistic physics, 206–207
  research programs, 714–715
  scientific progress, 749
  social sciences, 780
  social sciences, Popper and, 575
  special relativity, 207
  statistics, 803
Predictivism, 589
Preference
  confirmation theory, decision theoretical approach, 148
  decision theory, expected utility representation of, 183–185
  Dutch Book argument, 212
  game theory, 323
  probability, 605
  Ramsey and, 677–678
Preformationism, 352
Preformationistic gene, 339
Premise circularity, 381
Presentism, time, 831
Pressey, R.L., 164, 165
Preston, J.M., 309
Price, G.B., 504
Price, G.R., 311
Price, Henry Habberley, 553
Price, Huw, 92
Price, Richard, 46, 47, 52
Prigogine, Ilya, 418
Primary structure, molecular biology, 485
Primary succession (ecological), 215
Primas, H., 103, 234
Primitives, neurobiology, 514
Primitive thisness, quantum field theory, 631
Primitivist accounts
  causality, 98
  probability, 607
Principal principle
  laws of nature, 442
  probability theory, 608
Principia Mathematica (Newton), *see Philosophiae naturalis principia mathematica* (Newton)
Principia Mathematica (Russell, Whitehead), 80, 82
  Hahn and, 342
  phenomenalism, 552
  Quine and, 659
  Ramsey and, 672–674
  rational reconstruction, 682
  Russell and, 717, 718, 719
  verifiability, 852
  Vienna Circle and, 860
Principle of acquaintance, 851
Principle of common cause, 555–556, 709
Principle of equivalence, Jaynes's, 56
Principle of indifference
  Carnap's, 390, 492
  Jaynes's, 59
  Laplace's, 52, 602
Principle of induction, 708
Principle of least constraints (Gauss), 120
Principle of maximal specificity, 279, 603
Principle of tolerance, *see* Tolerance, principle of
Principles and parameters framework of generative grammar, 109
Principles of Mathematics (Russell), 718, 721
Pringsheim, Ernest, 359, 650
Prinzipien der Mechanik, Die (Hilbert), 359
Prior knowledge
  Bayesianism, 55–59
  Hansen, logic of discovery, 346
Prior probabilities, Bayesian updating, 593–594
Priors, probability theory, 600
Prisoner's Dilemma, 10–11, 327, 781, 784
Probabilistic approaches
  confirmation theory, 148–150
  Nagel and, 495–496
  population genetics, 584
  probability, 606
  Ramsey and, 680
  social sciences, 495–496
Probabilistic laws, laws of nature, 440, 441
Probabilistic propensity, fitness, 312–313, 314
Probabilistic relevance accounts, causality, 95–96
Probability, 599–610
  Bayesianism, 45, 48; *see also* Bayesianism
  Carnap, 86–87, 88
  causality, 95
  classical interpretation, 602–603
  confirmation theory, Bayesian, 148–150
  corroboration, 177–179
  decision theory, 181–188
  Dutch Book argument, 210–213
  evolutionary biology, philosophy of, 70
  explanation
    deductive-nomothetic model of probabilistic (DN-P) explanation model, 281–282
    inductive-statistical (I-S) model, 279, 280
    statistical relevance model, 280–281
  explicanda/explicata pairs, 288
  fitness, 312–313
  frequentism, 601–602
  heritability, 353, 354–355
  induction, problem of, 380, 382
  inductive logic, 384, 386–391
    epistemic interpretations of probability, 386–387
    logical interpretation, 387–391
    mathematical theory of probability, 386
    naive version and received view, 385–386
    relevance, 392–393
  inference, roles in, 803
  Kolmogorov's axiomatization, 599–601
  logical, 86–87, 603–606
  logical empiricism, 463
  Nagel and, 492
  natural selection
    chance in evolution, 500–501
    fitness, 500
  objective chance, 606–609

  parsimony, 533, 534–535
  psychology, cognitive architecture, 616
  quantum logic
    classical ensembles and probability densities, 635
    consistent histories approach, 642
    probability amplitudes and Feynman paths, 636–637
  Ramsey and, 671, 675–676
  Reichenbach and, 704, 708–709
  scientific revolutions, 755
  scientific style, 766, 767
  statistics, 803, 804; *see also* Statistics, philosophy of
  subjectivism, 604–606
  verisimilitude, 854
  Vienna Circle, 860
  von Neumann and, 504
Probability amplitudes, quantum logic, 636–637
Probability-and-measure theory, 433
Probability assignment, Bayesianism, 41
Probability densities, quantum logic, 635
Probability distribution
  biology, philosophy of, 69
  Bohmian, 656
  quantum mechanics, 453
Probability function
  conditional, 387–388, 389
  a priori, 391
  probability theory, 599, 600
Probability kinematics, 679
Probability measures, quantum measurement problem, 644–645
Probability model, inductive logic, 386
Probability rules, Bayesianism, 42
Probability space, 599
  von Neumann computation theory, 508
Probability theory
  and classical mechanics, 121
  inductive logic, 384
  old evidence problem, 150
Probability theory of meaning, Reichenbach and, 707, 708
Probability/utility pairs, decision theory, 183
Probability value, complementarity logic, 633
Probing models, 747
Problem of induction, *see* Induction, problem of
Problem of irrelevant conjunction, 147
Problem of Knowledge, The (Ayer), 38
Problem of old evidence, 150, 590
Problem of provisos, 210
Problems of Philosophy (Russell), 720
Problem solving
  evolutionary epistemology, 260
  Popper and, 572
  scientific progress, 751–752
Processes
  causality, ontological categories, 90
  quantum logic, 642
Production, experimental, 270–271
Product rule, Bayesianism, 42
Product state, vector state, 456
Programmed cell death, 338
Program metaphor, molecular biology, 484–485
Programs, artificial intelligence, 31–32
  computation, 28
  hard versus soft, 30
Progress
  Kuhn's disciplinary matrices and scientific revolution, 422–423
  natural selection, 499–500
  scientific, 749–753; *see also* Scientific progress
  scientific change, 729–732
Projectible predicates, 177
Projectibility, confirmation theory, 349
Projection operators
  quantum logic, 633, 639–640
  quantum mechanics, 653
Projection postulate, quantum mechanics, 654
Projective geometry, 357
Project K, 487
Prokaryotic organisms
  genetics, 335
  molecular biology, 481–482, 483
Proof
  Hilbert, 358
  Lakatos and, 435
  visual representation, 867
Proofs and Refutations (Lakatos), 434, 437, 438, 439
Proof theory, Hilbert and, 357, 358–359, 360
Propensities, corroboration, 178
Propensity interpretation of fitness, 312–313, 500
Propensity interpretation of probability, 607
Proper names
  linguistics, 446
  Russell and, 721
Properties
  causality, ontological categories, 90
  linguistics, intrinsic versus relational, 111–112
Property identity claim, physicalism, 562
Proportionality constant (Hubble's constant), 36
Proportionality constant (likelihood principle), 44
Propositional calculus, classical, 634
Propositional reference, belief content, 676
Propositions
  analyticity, 12
  chemistry, philosophy of, 102
  complementarity logic, 633
  empiricism, 524
  inductive logic, 384, 385
  mental representations, 615
  quantum logic, 634, 635
Prospective models, population genetics, 579
Proteins
  genetics, 335
  molecular biology, 481, 483, 484, 485
  proteomics, 337
Protein synthesis, 335
Protein taxonomy, 253
Proteomics, 337
Protocol sentences, 610–613
  Carnap and, 82–83
  cognitive significance, 133
  empiricism, 238
  Hanson and, 344
  Kuhn's scientific revolutions, 424
  Neurath and, 512, 513
  Quine and, 666
  Schlick and, 728
Protons, particle physics, 539
Proto-oncogenes, 338
Proust, Joelle, 12

Provine, William
  evolution, 253
  heritability, 354
  population genetics, 578, 583
Provisos, see also *Ceteris paribus* clauses
  Duhem thesis, 210
  fitness, variances, 313
  laws of nature, 443
  Popper and, 577
Pruitt, R.H., 486, 488
Pseudo-alleles, 332
Pseudo-Euclidean space, 790
Pseudo-explanations, 277
Pseudo-problems
  Carnap and, 83
  Quine and, 660
  Schlick and, 728
Pseudo-propositions
  Ayer and, 39, 40
  Vienna Circle and, 859
Pseudorandom elements, Turing machines, 836
Pseudorationality, Neurath and, 512
Pseudoscience, 194–195, 712; *see also* Cognitive significance; Demarcation, problem of
  demarcation problem, 188–197
  Lakatos and, 434
  prediction, 586
Psi (mechanism levels), 472
Psi (quantum state)
  prediction, 595–596
  quantum field theory, 629, 630
  quantum mechanics, 656
Psillos, Stathis
  empiricism, 241
  instrumentalism, 400
  realism, 695
  theory structure, 826, 827
Psychic unity of mankind (monomorphic mind thesis), 266–267
Psychoanalytic theory, *see* Freudian/psychoanalytic theory
Psychological altruism, 9–10
Psychological association, Hume and, 523
Psychological laws, 443
Psychological mechanisms, abduction, 2
Psychologism
  causality, 91
  observation, 524
Psychology
  altruism, 8–11
  behaviorism, 61–63, 114
  biology, philosophy of, 69
  Carnap and, 80–81
  Chomsky on, 114
  cognitive science multidisciplinary approach, 123
  cognitive significance, 350
  connectionism, 151, 155–157
  consciousness, 161
  and emergentism, 231
  evolutionary, *see* Evolutionary psychology
  experimental method, 268, 270, 274
  feminist philosophy, 296, 298
  function, 319–320
  Hanson and, 344, 345
  induction, problem of, 379
  Kuhn and, 427–428, 430
  Lakatos and, 434
  linguistics, *see* Linguistics
  Mach and, 467–468
  Nagel and, 493, 495–496
  observation, Carnap approach, 523
  perception, Marr's computer model of mind, 548
  philosophy of, 613–619
    behaviorism, 61–63
    cognitive architecture and processing, 615–617
    cognitive science, 127
    consciousness, 158–163, 617–618
    embodied, embedded, and situated cognition, 618–619
    intentionality and mental representation, 614–615
    Mach and, 467–468
    pain, psychopathology, and color, 618
    scientific progress, 750
  physicalism, 83; *see also* Physicalism
  psychology, cognitive architecture, 617
  rational reconstruction, 681–685
  reductionism, 701
  Schlick and, 725
  scientific revolutions, 422
  Searle, 769
  statistics, 808–809
Psychology from an Empirical Standpoint (Brentano), 406, 407
Psychopathology, 618
Psycho-physical parallelism principle, 505–506
Ptak, Pavel, 633
Ptolemy/Ptolemaic systems
  astronomy, 33–34
  demarcation problem, 193
  Feyerabend on, 307
  Kuhn's scientific revolutions, 421, 422, 423, 427, 428
  parsimony, 532
Public policy, *see* Social/public policy
Publishing, scientific, 34–35
Pulmannova, Sylvia, 633, 643
Punctuated equilibrium, 254
Punnett, R., 580
Pure states, quantum measurement problem, 645
Purpose, causality, 90
Putnam, Hilary
  abduction, 1–2
  analyticity, 20
  and Carnap, 88
  and Chomsky, 111–112, 113
  cognitive science, 130
  conventionalism, 171
  empiricism, 241
  explanation, criticism of Nagel's reductive model, 285
  inductive logic, 389, 390
  innate/acquired distinction, 397
  life, work, views, 620–627
    further implications of, 625–626
    meaning and mind, 623–625
    realism and reference, 621–622
    reference and perception, 622–623
  linguistics, philosophy of, 446, 447
  mechanism, 472
  quantum logic, 638, 641
  realism, 688, 694
  scientific domains, 734
  theory structure, syntactic (received) view, 823
  time, presentism, 831
  unity and disunity of science, 845
p-values, 811, 812–814

Pylyshyn, Z.W., 28, 31
  cognitive science, 128
  connectionism, 153
  psychology, philosophy of, 615
Pythagoras, 843

Q

Qualia, 617
  intentional theory of, 409
  perception, 546
Qualitative confirmation, 144
Qualitative notions, inductive logic, 385
Qualitative predicates, explanation, 278
Qualitative properties, chemistry, 103
Qualitative results, experimental method, 272
Quantitation
  confirmation theory, 144, 348
  experimental method, 272
  inductive logic, 385
Quantitative genetics, 354–355
Quantitative trait loci (QTL), 334
Quantization, quantum field theory, 629
Quantum chromodynamics, 542, 543
  scientific models, 746–747
Quantum computation, 835
Quantum cosmology, 658
Quantum discontinuity, 651
Quantum electrodynamics, 541
  particle physics, 538
  quantum field theory, 632
Quantum electro-weak dynamics, 542, 543
Quantum entanglement, reductionism, 700–701
Quantum field theory, 506, 629–633
  alternative approaches, 631–632
  effective field theory, 543–544
  field theory and canonical quantization, 629–630
  indistinguishability, 631
  interacting fields, 632
  irreversibility, 413
  locality/nonlocality, 451, 456–457
  particle physics, 538
    S-matrix program and particle democracy, 541–542
    standard model, 542–544
    symmetries and particles, 540–541
  particles or fields, 631
  quantum logic, 642
  routes to, 630–631
  von Neumann and, 506
Quantum lattice logic, 633
Quantum logic, 506, 633–644
  classical ensembles and probability densities, 635
  classical logic versus, 637–638
  classical mechanics and classical logic, 633–634
  consistent histories approach, 642–643
  first-order functional calculus, 635–636
  history of, 633
  operational versus dialogic approaches, 642
  orthomodular lattice logic, 638–639
  partial Boolean algebra logic, 639–640
  prediction, retrodiction, and state functions, 637
  probability amplitudes and Feynman paths, 636–637
  quantum mechanical propositions, 636
  semantic problems, 640–641
  utility of, 641–642
Quantum measurement problem, 644–649
  complementarity, 140–143
  determinism, 204
  false starts and discredited answers, 648–649
  measurements, 646
  physical sciences, philosophy of, 556
  prediction, 595–596
  probability, 607
  problems, 646–648
    proofs of more general problem, 647–648
    state as end of measurement, 646–647
  quantum logic, 636
  quantum measurement problem, as end of measurement, 646–647
  quantum mechanics, 654
  quantum states, 644–646
    mixed versus pure states, 645
    standard interpretation of pure states, 645
    states as probability measures, 644–645
    superpositions, 645–646
  Reichenbach and, 710–711
  Turing and, 835
  von Neumann and, 505–506
Quantum mechanics, 119, 120
  approximation, 25
  chemistry, philosophy of, 103, 104–105
  complementarity, 140–143
  determinism, 204–207
  emergentism, 234
  irreversibility, 413
  kinetic theory, 417
  Kuhn on, 428–429
  locality/nonlocality, 451, 454–455, 456–457
    action at a distance, 455–456
    Bell's theorem, 453–454
    entanglement and EPR, 452–453
    holism and nonseparability, 455
    relativistic quantum field theory, 456–457
    subtleties of nonlocality, 456
  Nagel on, 493
  particle physics, 539
  physical sciences, philosophy of, 554, 555, 556
  prediction, 595–596
  probability, 607
  quantum logic, 633, 642
    consistent histories approach, 642–643
    definition of QM proposition, 636
    lattice logic, 638
    state functions, 637
  realism, 692–693
  reductionism, 700–701
  Reichenbach and, 707, 709–711
  relativistic statistical-mechanical entropy and, 413
  scientific revolutions, 754
  theories, models, interpretations, 649–658
    Bohr's atomic theory, 651–652
    Bohr's complementarity interpretation, 657
    deBroglie-Bohm theory, 656
    hidden variables, 655–656
    interpretations of, 654–655
    miscellaneous, 657–658
    modal intepretations, 657
    old quantum theory, 650

Quantum mechanics (*Continued*)
    quantum conditions, 652
    quantum statistics, 650–651
    Schrödinger, Heisenberg, and Hilbert space formalism, 652–653
  theory structure, 825
  Turing and, 833, 835, 837
  von Neumann and, 504, 505–507
Quantum nonlocality, 831; *see also* Locality
Quantum numbers, magnetic, 652
Quantum physics
  causality, 90
  and classical mechanics, 121
  determinism, 204–205
  logical positivists, 460
  von Neumann and, 503, 505–507
Quantum probability, fitness, 312
Quantum states
  locality, entanglement and EPR, 452–453
  quantum measurement problem, 644–646
Quantum statistics, quantum mechanics, 650–651
Quantum theory, 120
  observation, 526
  physical sciences, philosophy of, 556
  quantum measurement problem, 644–649
  space-time, 794
  Turing and, 837
Quantum unpredictability, 837
Quarks, 542
Quasi-analysis, Carnap, 81
Quasi-deterministic processes, causality, 90
Quasi-eliminativist option, physicalism, 563
Quasi-empirical theory, Lakatos and, 438
Quasi-mathematical models of probability, 384
Quine, Willard Van
  analyticity, 11–21
    Harman's synthesis of case against, 15–19
    historical background, 11–13
    responses of Carnap, Grice, Strawson, and others, 14–15
    two dogmas of empiricism, 13–14
    web of belief, 19–20
  Ayer and, 40
  Carnap and, 86, 88
  Chomsky and, 110–111
  cognitivism, 63
  confirmation theory, 146
  conventionalism, 170, 176
  demarcation problem, 191, 194
  Duhem thesis, 209
  empiricism, 239–240
  epistemology, 246
  explication, objections to, 289, 293
  feminist philosophy, 301
  Hempel and, 350
  intentionality, 408
  Kuhn and, 428
  life and work, 659–669
    observation, theory, and naturalized epistemology, 666–668
    Two Dogmas of Empiricism, 660–666
  linguistics, philosophy of, 448
  logical empiricism, 461–462, 463–464
  phenomenalism, 551–552
  Poincaré and, 570
  Popper and, 577
  prediction, 588
  protocol sentences, 612–613
  Putnam and, 620, 621, 623
  rational reconstruction, 683
  scientific models, 744
  scientific revolutions, 764
  scientific style, 767
  underdetermination of theories, 839, 841
  unity of science movement, 848
  verifiability, 852
  Vienna Circle, 858

R

Radder, Hans, 269, 270, 271, 273
Radiation, Kirchoff's law of, 359
Radical behaviorism, 63
Radical constraints, linguistics, 111
Radical constructivism, Vienna Circle, 861
Radical empiricism, kinetic theory, 417
Radical phenomenalism, instrumentalism, 400
Radical probabilism, Ramsey and, 680
Radical revisability, Quine and, 663
Radical translation, 428
Radioactive processes
  approximation, 25
  particle physics, 538
Radio astronomy, 36–37
Radiotelescopes, 36–37
Radman, Z., 739
Radnitsky, Gerard, 262
Radon-Nikodym theorem, 600
Rafe, R.W., 165
Raff, R., 266
Raiffa, H., 323
Railton, Peter, 95, 281–282
Ramified type theory, 672–673
Rampton, Sheldon, 195
Ramsey, Frank Plumpton, 87
  approximation, 26
  Bayesianism, 41, 49, 148
  confirmation theory, 148
  decision theory, 183
  Dutch Book argument, 210–213
  instrumentalism, 403
  laws of nature, 442
  probability, 604–605
  reductionism, 699
  Vienna Circle, 858
  work of, 671–681
    degrees of belief, 676–677
    impredicative function, infinity, abandonment of logicism, 674
    logic and *The Foundations of Mathematics*, 672
    logic of consistency for degrees of belief, 677–678
    logic of truth, 678–679
    paradoxes and theory of types, 672–674
    probability and partial belief, 675–676
    Ramsey's theorem, 675
    scientific theories, laws, and causality, 679–680
    value of knowledge, 679
Ramsey, J., 105
Ramsey numbers, 675
Ramsey sentence, 403, 679

Ramsey's theorem, 671, 675
Ramsey test, 680
Ramsey theory, 675
Ramus, Petrus, 843
Rand, Rose, 858
Random drift, population genetics, 583
Random elements, Turing machines, 836
Ratcliffe, M., 320
Ratio measure, Bayesian confirmation theory, 148
Rational agent/rational choice/rational preference
  confirmation theory, 148
  decision maker, 185
    decision theory, 181–183
    economic rationality, 227
  game theory, 323, 326, 507
  inductive logic, epistemic interpretations of probability, 386–387
  methodological individualism, 479
  Neurath and, 511
  probability, 605, 606
  social sciences, varieties of social explanation, 780–781
Rational debate, economics, 228–229
Rational degrees of belief, 676
Rationalism
  Feyerabend on, 308, 309
  innate/acquired distinction, 394, 396
  nativism and, 396
  scientific progress, 750
  scientific revolutions, 755
Rationality, *see also* Irrationality
  group, 784
  incommensurability, 371
  Kuhn's disciplinary matrices and scientific revolution, 424–427
  methodology, 507
  rational reconstruction, 684
  scientific change, 730
  scientific progress, 749–752
  scientific revolutions, 760, 764
Rational mechanics, defined, 116
Rational reconstruction, 681–685
  Lakatos and, 684–685
  logical empiricism, 682–684
Rational truths, Schlick and, 726
Ratliff, Floyd, 468
Raum, Der (Carnap), 80
Rauscher, M., 267
Ravens (Hempel's) paradox
  confirmation theory, 145, 146, 148–149
  Hempel, 348–349
  induction, problem of, 382
Rawls, John, 228
Rayleigh, Lord (John William Strutt), 650
Rayleigh-Jeans formula, 650, 651
Raz, Joseph, 185
Realism
  abduction, 1–2
  arguments for, 689–692
  artificial intelligence, 31
  Carnap, 86, 88
  causality, 91
  chemistry, philosophy of, 103
  cognitive simulation, 31
  economics, 226
  empiricism, 241, 243
  experimental science and, 271
  feminist philosophy, 296
  formulation of thesis, 686–689
    components of realism, 686–687
    contrasting formulations, 688
    observational-theoretical distinction, 687
    scope of realism, 687–688
  instrumentalism, 404–405
  kinetic theory, 417–418
  Kuhn's scientific revolutions, 426
  logical empiricism, 463–464
  Nagel on, 493
  phenomenalism, 552–553
  physicalism, 561
  physical sciences, philosophy of, 555
  Poincaré and, 568, 570–571
  Putnam and, 620, 621–622, 626
  rational reconstruction, 686–696
  Reichenbach and, 706, 707–708
  Russell and, 717, 721
  Schlick and, 728
  scientific domains, 734
  scientific metaphors, 738
  scientific models, 740, 747
  scientific progress, 749–750, 753
  scientific revolutions, 764
  skeptical historical introduction, 693–695
  social constructionism, 776
  verisimilitude, 854
Realist model
  artificial intelligence, 32
  causality (covering-law), 97–98
Reality-belief relationship, protocol sentences, 612–613
Realizability, multiple, 701–702
Realizationism, physicalism, 565
Real numbers, probability theory, 600
Reason
  dialectic of, Lakatos and, 437–438
  science wars, 778
  scientific style, 767
Reasoning
  induction, problem of, 379
  scientific style, 767
  visual representation and, 867
Received view (RV), inductive logic, 386, 389–390
Received (syntactic) view, theory structure, 822–824
Recessive traits, 330
Reciprocal altruism, 784
Reciprocity, classical mechanics, 117
Recombination, genetic, 497, 579
Recombination analysis, 334
Reconstructing Scientific Revolutions (Hoyingen-Huene), 371
Reconstruction, rational, 681–685
Recursion, Poincaré and, 570
Recursive arithmetic, 360
Recursive auto-associative memories (RAAMs), defined, 154
Recursive epistemology, Hilbert, 360–361
Recursive rules, connectism, 153
Rédei, Miklo's, 503, 506
Redelmeier, Donald, 185
Redhead, M.L.G., 454, 632, 746
Reducibility, axiom of, 672, 719
Reduction
  causal relations, 90
  intentionality, 408–409

Reduction (*Continued*)
  kinetic theory, 418–419
  molecular biology, 485
  neurobiology, 516–518
Reduction, data, 743
Reduction functions, 699
Reductionism, 696–703
  aggregativity and emergence, 702
  American Structuralism, 107
  analyticity, 13
  approximation and, 25
  biological information, 66
  biology, philosophy of, 69, 71
  Carnap and, 81
  causality, 91–92, 93
  chemistry, philosophy of, 102, 103, 104–105
  Chomsky on, 114
  classical mechanics, 119, 121
  cognitive significance, 136, 137
  economics, laws of, 226
  emergentism and, 230, 234
  explanation, 285
  genetics, 335, 337
  immune self, 364
  instrumentalism, 402
  interlevel reduction, 699–701
  intralevel reduction, 698–699
  methodological individualism, 478–479
  molecular biology, 483, 484
    information, 484–485
    philosophy of, 71
  multiple realizability and supervenience, 701–702
  Nagel and, 492
  neurobiology, 514
  physicalism, 560, 566
  physical sciences, philosophy of, 555
  probability theory, 607, 608
  protocol sentences, 610, 612
  Putnam and, 620
  Quine and, 659, 661–662
  Ramsey on, 672
  scientific change, 730
  scientific models, 740, 747–748
  scientific progress, 750
  scientific revolutions, 759
  social sciences, 783
  supervenience, 816
  theory structure, 826
  unity and disunity of science, 845
Reduction sentences
  observation, 524
  verifiability, 852, 853–854
Reductive explanation, 276
Reductive instrumentalism, 402–403
Reference, *see also* Putnam, Hilary; Quine, Willard Van; Russell, Bertrand
  incommensurability, 372–373
  Putnam, 625
    and perception, 622–623
    realism and, 621–622
  scientific domains, 734
  visual representation, 865
Reference class partition, explanation, 281
Reference sequence problem, probability, 601
Reference theories, Chomsky and, 111
Referential realism, 271
Referential semantics, linguistics, 446–448
Reflexive predictions, 780
Refutation, *see also* Corroboration
  inductive logic, 392
  Lakatos and, 436
  partial, 384
Registration, quantum logic, 636
Regression analysis, heritability, 354–355
Regression to the mean, 353–354
Regularities
  causation, 95, 99
  laws of nature, 439–443
Regularity theory of causation, 93–95
Regulatory genes, 481, 482, 486
Reich, G., 848
Reichenbach, Hans, 83, 87
  Carnap and, 80
  causality, 92, 97
  cognitive significance, 132, 134
  Duhem thesis, 210
  empiricism, 237
  induction, problem of, 381–382
  Kuhn's scientific revolutions, 425, 427
  life and work, 703–712
    direction of time, 709
    ethics, 711
    logics, 709
    probability and causality, 708–709
    quantum mechanics, 709–711
    realism and meaning, 707–708
    space and time, 705–707
  logical empiricism, 461
  physical sciences, philosophy of, 555–556
  prediction, 592–593, 594
  probability, 601
  quantum logic, 633
  Quine and, 659
  rational reconstruction, 683
  scientific change, 730
  verifiability, 851, 853
  Vienna Circle, 858, 862
Reichenbach, Maria, 709
Reid, Thomas, 545
Reidemeister, Hans, 728
Reidemeister, Kurt, 342, 858
Relational holism, 234
Relationalists, space-time, 831
Relatives, correlation between, 580–581
Relative state interpretation, quantum mechanics, 658
Relativism, *see also* Incommensurability; Kuhn, Thomas; Social constructionism/constructivism; Theory/theories, underdetermination of
  democratic, 308
  feminist philosophy, 301
  Feyerabend, 308–309
  scientific change, 730
  scientific style, 765
Relativistic field equations, approximation, 24
Relativistic field theory, and classical mechanics, 115
Relativistic physics
  determinism, 205–207
  field theory, 632
    and classical mechanics, 115

    - equations, approximation, 24
    - locality, 456-457
  - logical positivists, 460
- Relativistic statistical-mechanical entropy, 412–413
- Relativity, conceptual, 620
- Relativity, Epistemological, 706
- Relativity, philosophical, 706
- Relativity principle, Galilean, 787, 788
- Relativity theory
  - and classical mechanics, 117, 121
  - emergence, 234
  - empiricism, 237
  - epistemological, 706
  - Euclidean geometry, 361
  - Feyerabend and, 306
  - general, old evidence problem, 150
  - Mach, 467
  - Nagel on, 493
  - Reichenbach and, 704
  - Schlick and, 727–728
  - scientific revolutions, 754, 760
  - underdetermination of theories, 841
- Relevance
  - confirmation theory, Bayesian, 148
  - explanation, 283
  - inductive logic, 392–393
  - probabilistic, causality, 95–96
  - statistical, explanation, 280–281
- Relevance measures, inductive logic, 393
- Reliabilist theories, epistemology, 248–249
- Religion, demarcation problem, 188–197
- Remnant models, 827
- *Renaissance of Empiricism in the Recent Philosophy of Mathematics, A* (Lakatos), 438
- Renaissance science, 843
- Renormalizabile quantum field theories, 540–541, 632
  - electro-weak theory, 543–544
  - particle physics, 544
- Rensink, R.A., 548
- Renyi, Alfred, 600
- Repeat rate, Bayesian statistics, 54
- Replicators, 261
- Representation(s)
  - Bayesianism, 55–59
  - chemistry, philosophy of, 103
  - cognitive science assumptions, 125–126, 128–129
  - computation, 28
  - decision theory, expected utility representation of preference, 183–185
  - experimental method, 272
  - factual, Hanson and, 346
  - game theory, 324–325
  - information, terminology, 64
  - intentionality, 405–406, 407
  - Kuhn's disciplinary matrix, elements of, 421
  - linguistics
    - Chomsky and, 109, 111
    - levels of, in generative grammar, 107
    - philosophy of, 448–449
  - mechanisms, 472–473
  - neurobiology, 520–521
  - psychology, 617
  - quantum field theory, 631
  - scientific models, 747–748
    - data, 743
    - epistemology, 744–745
    - phenomena, 741–743
    - theory, 743
  - social constructionism, 776–777
  - visual, 863–870
- Representationalism, psychology, 617
- Representational models, 747
- Representation bearer, defined, 128
- Representation in brain, 520–521
- Representation theorem
  - De Finetti, 50, 51
  - Savage, 181, 183
- Reproducibility, experiment, 270, 273
- Reproductive fitness, 69; *see also* Fitness
- Reproductive isolation, species, 797–798
- Reproductive success, 500–501; *see also* Fitness
- Rescher, Nicholas, 261, 753
- Research programs, 712-715
  - chemistry, philosophy of, 105
  - cognitive science, 123–124
  - connectionism, 151
  - Darwinism as metaphysical research program, 260
  - Lakatos
    - methodology of scientific research programs, 438–439
    - proofs and refutations, 433–437
    - transition to research program, 437–438
  - prediction, 589
  - rational reconstruction, 681
  - scientific change, 730, 731
  - scientific progress, 752
  - scientific revolutions, 759–760
  - social sciences, 784
  - Vienna Circle and, 859
- Resilience, ecological communities, 218
- Resistance, ecological communities, 218
- Resnik, David, 196
- Responsible theory choice, 839
- Restrictive interpretation, quantum mechanics, 710
- Retherford, R.C., 26
- Retrodiction, quantum logic, 637
- Retrospective models, population genetics, 579, 583
- Reversibility objections
  - irreversibility, 411
  - kinetic theory, 418
- Revision, Quine and, 662–663
- Revolutionary versus revolution, 762
- Reward-based learning, neurobiology, 521
- Rey, Georges, 210
- Réyni, Alfred, 433
- Rheinberger, Hans-Jörg
  - biology, philosophy of, 72
  - experimental method, 269
  - genetics, 337
  - molecular biology, 484
- Rhine, J.B., 194–195
- Ribonucleic acid, 335–336
- Ribosomes, 336
- Richard's finitely undefinable decimal, 673
- Richardson, Alan
  - Ayer, 40
  - Carnap, 82
  - logical empiricism versus logical positivism, 458
  - rational reconstruction, 682
  - Vienna Circle, 858

Richardson, Robert C.
  mechanism, 469, 473, 475
  reductionism, 699
Richness assumptions, decision theory, 183
Rickert, H., 511, 844
Ricketts, Thomas, 175
Ricklefs, Robert, 222
Riehl, A., 704
Riemannian geometry, 569
Rimini, A., 655
Rindler wedges, 457
Ring species, 797
Riordan, M., 542
Rise of Scientific Philosophy, The (Reichenbach), 711
Risk aversion, decision theory, 184
Risk perception, decision theory, 187–188
RNA
  alternative splicing, 482
  biological information, 64–65
  genetics, 337
  molecular biology, 488
    noncoding DNA transcription, 486
    regulatory networks, 487
RNA editing, 337
Road since Structure, The (Kuhn), 371, 420
Roberts, John, 443
Roberts, R., 483
Robinson Crusoe example, 513
Robotics
  artificial intelligence, 27, 29
  connectionism, 151
Robot objection, 773
Rock, Paper Scissors, 329
Roe, S.A., 353
Roeper, Peter, 386, 600
Rogers, Ben, 39, 40
Rolls, Edmund T., 156
Roozebloom, W., 86
Rorty, Richard, 660–661, 764
Rosaldo, Michelle Z., 298
Rosen, D.E., 799
Rosen, Deborah, 95
Rosen, Gideon, 242
Rosen, Nathan, Einstein-Podolsky-Rosen relation, 452–453
Rosenberg, Alexander
  biology, philosophy of, 69, 71
  determinism, 197
  economics, philosophy of, 224, 227
  fitness, 312, 314
  natural selection, 501
  prediction, 597, 598
  social sciences, 783
  unity and disunity of science, 846
Rosenberg, Charles, 152
Rosenberg, R., 809
Rosenfeld, Edward, 157
  artificial intelligence, 27
  connectionism, 151
Rosenkrantz, R.D., 42, 51, 52, 808
Rosenthal, D., 158
Rotating dust space-times, 830
Rothley, K.D., 167
Rothschild, Michael, 188
Rougier, Louis, 848
Rouse, Joseph
  demarcation problem, 193, 196
  scientific revolutions, 762
Roush, Sherrilyn, 21
Routman, E.J., 355
Royall, Richard M.
  Bayesianism, 44, 45, 56, 149
  confirmation theory, 149
  parsimony, 533
  statistics, 808
Rozebloom, W.W., 139–140
Rubenstein, A., 326
Rudy, J.W., 156
Ruetsche, Laura, 301, 302
Rule of induction, 708
Rules
  inductive, 381–382
  linguistics, philosophy of, 448–449
  social sciences, 783
  verifiability, 853
Rules of inference, empiricism, 240
Rules of probability, Bayesianism, 42
Rules of succession, 50
Rumelhart, James, 151, 157
Ruse, Michael, 195
  altruism, 8
  biology, philosophy of, 69, 73
  evolutionary epistemology, 261
  function, 320
  natural selection, 500
Russell, Bertrand
  Ayer and, 39
  Carnap and, 84
    Frege-Russell thesis of logicism, 80
    logicism, 82
  causality, 93
  empiricism, 237
  epistemology, 244
  explication, 287
  Hahn and, 341, 342
  Hertzsprung-Russell (H-R) diagrams, 36
  Hilbert and, 360
  induction, problem of, 380
  life and work, 715–723
    mathematics, 716–719
    philosophy of science, 719–722
  logical positivism, 460
  Nagel on, 492
  phenomenalism, 551, 552, 553
  Poincaré and, 570
  Putnam and, 622–623
  Quine and, 659
  Ramsey and, 671, 672–674, 679
  rational reconstruction, 682
  Schlick and, 726, 727
  Turing and, 833
  unity and disunity of science, 844
  verifiability, 851, 852
  Vienna Circle and, 860
Russell, Henry Norris, 36
Russell, S., 616, 619
Russell-Einstein Manifesto, 716
Russell's paradox, 717, 718

Russell's set, 673
Rutherford, Ernest, 651
Rydberg's constant, 651
Ryder, Lewis H., 630

S

Saari, Donald G., 122
Sabel, Charles F., 229
Sahlin, Nils-Eric, 671, 675
Saint Hilaire, Etienne Geoffroy, 252
Sakamoto, Y., 536
Salam, A., 542
Salmon, Merrilee, 780
Salmon, Wesley C.
  causality, 90, 95, 96, 99
  conventionalism, 172, 176
  corroboration, 178
  explanation, 280, 281, 282, 283
  induction, problem of, 381, 382
  inductive logic, 392
  kinetic theory, 418
  Kuhn's scientific revolutions, 425
  locality/local action, 451
  logical empiricism, 463
  logical empiricism versus logical positivism, 458
  mechanism, 469
Saltation/saltationism, natural selection, 499
Sampling/sampling theory
  Bayesianism, 42, 44–45, 57
  statistics
    large N problem, 808–809, 812
    roles for probability in inference, 803
    stopping rule, relevance of, 808
Sankey, Howard, 760
Sansom, Roger, 72
Santillana, G., 848
Sarkar, Sahotra, 339
  approximation, 26
  biological information, 65, 67
  biology, philosophy of, 69
  Carnap, 80, 84
  conservation biology, 164, 165, 166, 167
  determinism, 197
  ecology and conservation biology, 73
  emergence, 233, 234
  empiricism, 238
  explanation, 285
  genetics, 334, 335
  heritability, 354, 355
  molecular biology, 481, 483, 484, 485
    networks, 488
    philosophy of, 72
    reductionism, 487
  population genetics, 582
  reductionism, 696, 697, 698, 699, 700, 701
  scientific models, 748
  scientific revolutions, 759
  time, 830
Sartre, J.P., 552
Satisfaction criterion, confirmation theory, 146–147
Sauer, T., 359
Sauer, W., 82
Savage, C.W., 82
Savage, Leonard J.
  Bayesianism, 41, 55, 148
  confirmation theory, 148
  decision theory, 181–188
  probability, 605
  Ramsey and, 679, 680
  statistics, 807, 808
Scale models, 740, 827
Scaling, symbolic models, 154
Scerri, Eric, 104
  approximation, 25
  prediction, 590
  realism, 689
Schaffer, Simon
  experimental method, 268, 269, 272
  scientific revolutions, 758
Schaffner, Kenneth. F., 69, 71, 484, 514
  approximation, 25
  reductionism, 697, 698, 699, 700
  theory structure, syntactic (received) view, 823
  unity and disunity of science, 846
Scheffler, Israel
  cognitive significance, 135
  incommensurability, 372
  scientific revolutions, 760
Scheibe, E., 359
Scheines, R., 96
Schematic drawings, 866
Scher, S., 267
Scheutz, M., 130
Schiebinger, Londa, 296, 297, 302
Schiller, Ralph, 635
Schilpp, Paul
  evolutionary epistemology, 260
  explication, 290
  Popper, 572
  Quine, 667
Schleichert, Hubert, 728
Schlesinger, George N., 591
Schlick, Moritz
  Carnap, 80, 82, 83
  cognitive significance, 132, 133
  conventionalism, 169, 174–175
  Duhem thesis, 209
  empiricism, 238
  Hahn and, 342
  life and work, 725–729
    final year, 729
    between natural and cultural philosophy, 726–729
  and logical empiricism, 459
  logical positivism, 460–461
  Neurath and, 513
  protocol sentences, 610, 612
  Quine and, 659
  Reichenbach and, 705
  Vienna Circle, 858, 859, 860, 862
Schlick Circle, 728, 858, 862
Schmidt, Erhard, 504
Schmidt, J.J., 826
Schmidtz, David, 227
Schmoller, Gustav, 511
Schrödinger, Erwin, 356, 646
  on biological information, 65
  Mach and, 467
  physical sciences, philosophy of, 554

Schrödinger, Erwin (*Continued*)
  quantum mechanics, 652–653
  Reichenbach and, 709
Schrödinger equation
  classical mechanics, 120
  determinism in quantum physics, 204–205
Schrödinger evolution, determinism, 204–205
Schrödinger field, quantum field theory, 630
Schrödinger QM, von Neumann and, 506
Schrödinger's diabolical device, 554
Schummer, J., 103
Schutz, Alfred, 774
Schwarz, G., 535, 536
Schwarzchild assumption, 24
Schweber, S.S., 541, 543
Sciana, D.W., 543
Science
  analytical philosophy, 715
  artificial intelligence, science versus engineering, 29–30
  demarcation problem, 188–197
  evolution of epistemic theories (EET), 260
  generative grammar, conceptual issues in, 113
  language of, 402–403
  social constructionism, 778
Science, technology, and society (STS), 784
Science and Subjectivity (Scheffler), 372
Science-forming faculty
  Chomsky and, 113
  generative grammar, conceptual issues in, 113
Science in a Free Society (Feyerabend), 305
Science of Mechanics (Mach), 468
Science Question in Feminism, The (Harding), 300
Science wars, 764, 778–779, 784
Scientific authority, 195
Scientific change, 88, 729–732
  Darwinian approach, 261
  experiment, 268
  incommensurability, 370–373
  irrationality and relativism, 730
  Lakatos, 433–439
  modeling discovery, 730–731
  objectivism, traditional, 730
  space-time, 788
Scientific Conception of the World: The Vienna Circle, 459, 727, 858, 860–861
Scientific determinism, 198
Scientific domains, 732–737
  causality, 90
  cognitive science, pretheoretical specification of domain, 123–124
  Hansen, logic of discovery, 346
  properties of domains, 735
  reductionism, 696
  seventeenth to nineteenth centuries, 734–735
  types of domain change, 735–736
  unification and transformation of science, 736
Scientific explanation, *see* Explanation
Scientific Explanation and the Causal Structure of the World (Salmon), 282
Scientific humanism, Vienna Circle, 860–861
Scientification of technology, 268
Scientific metaphors, 737–740
  evolutionary epistemology, 257
  molecular biology, 484
  population genetics, 582
Scientific method, *see* Method/methodology
Scientific models, 740–749; *see also* Models
  artificial intelligence, 31–32
  causality, 96–97
  connectionism, 151–153
  determinism, 199
  ecological, 220–223
    evaluation of, 221–222
    explanatory value of, 222–223
    feasibility of, 220–221
  ecology and conservation biology, 72
  epistemology, 744–746
  fitness, 313–315
  Kuhn's scientific revolutions, 428
  model theory
    Hilbert, multiplicity of geometries, 360–361
    Putnam and, 621
  ontology, 743–744
  and other debates, 747–748
  parsimony, 536–537
  particle physics, 538
  prediction, 593
  reductionism, 698
  scientific change, 730–731
  scientific metaphors, 737–740
  scientific progress, 750
  scientific style, 766
  semantics, representational function of, 741–743
    data, 743
    phenomena, 741–743
    theory, 743
  and theory, 746–747
  theory structure, 826, 827
  visual representation, 867
Scientific outlook/temper, 715
Scientific progress, 749–753
  analytical philosophy, 715
  Kuhnian view, 751–752
  Lakatos, 433–439
  logical positivists, 750
  Popperian view, 750–751
  post-Kuhnian development, 752–753
  price of progress, 753
  realism, 695
  reductionism, 698–699
  research programs, 714
  scientific change, 730
  scientific revolutions, 755
Scientific rationalism
  Feyerabend and, 309
Scientific Rationality (Hempel), 351
Scientific rationality, *see also* rationality
  incommensurability, 371
  Kuhn and, 424–427
  Popper and, 574–575
  scientific change, 730
  scientific progress, 749–752
Scientific realism, *see* Realism
Scientific reasoning, visual representation and, 867, 869
Scientific Revolution, The
  parsimony, 532
  scientific revolutions, 754
Scientific revolutions, 754–765
  Bayesianism, 58
  demarcation problem, 191

  history of science, 755–758
  importance of debate about, 764
  incommensurability, 370–371
  Kuhn, 419–431; *see also* Kuhn, Thomas
  Kuhn on structure of, 760–762
  Kuhn's disciplinary matrices and, 421–427
    contexts: discovery, justification, development, 427
    normal science, 423–424
    progress, 422–423
    rationality, 424–425
    rationality and the social, 425–427
    Scientific Revolution, The, 427
  philosophy of science, 758–760
  Poincaré and, 571
  revolutionary versus revolution, 762
  scientific metaphors, 738
  types and characteristics, 762–763
Scientific statements, versus nonscientific statements, 535–536
Scientific style, 765–767, 846
Scientific testing, protocol sentences, 612
Scientific theory, *see* Theory/theories; *specific theories*
Scientific world conception, 464, 727
  Ayer and, 38
  logical empiricism, 461
  logical positivism, 459, 460
  Vienna Circle, 858, 860–861
  von Neumann and, 505
Scientism
  Putnam and, 620
  Quine and, 668; *see also* Quine, Willard Van
  social constructionism, 778
Scope, symbolic models, 153
Screening off, explanation, 281
Scriven, Michael
  causality, 95
  explanation, 277, 280, 282
Searle, John
  cognitive science, 130
  intentionality, 406, 408, 409
  life and work, 767–774
    speech acts and intentionality, 768–770
    strong artificial intelligence, critique of, 771–773
    structure of consciousness, 770–771
    structure of speech acts, 768
  psychology, philosophy of, 617
  social constructionism, 775
  social sciences, 780, 783
Searle, R., 160
Secondary succession (ecological), 215–216
Second law of thermodynamics, 554, 700
Second-order Euler-Lagrange equations, 120
Second Scientific Revolution, The, 755
Sectors, quantum field theory, 632
Seeing that, Hanson and, 345
Segal, I.E., Gelfand-Naimark-Segal (GNS) construction, 631
Segregation, law of (genetics), 330, 596
Seidenfeld, T., 807
Sejnowsky, Terrence, 152
Selection, 312, 375; *see also* Evolution; Natural selection; Units of selection
Self-consciousness, 160, 161–162
Self-evidence, Russell and, 719
Self-interest, economic rationality, 227
Selfish gene, 498
Selfish Gene, The (Dawkins), 378
Self-modifying machines, Turing and, 836
Self-nonself distinctions, immune self, 364, 365–369
  assaults on, 367–369
  origins of, 365
  twentieth century reconstructions of, 366–367
Self-referentiality, causal, 769–770
Self-reproducing automata, 508–509
Selke, T., 812, 813
Sellars, Wilfred, 239
  phenomenalism, 551
  physical sciences, philosophy of, 555
Selten, R., 327
Selye, Hans, 739
Semantic paradoxes, Ramsey and, 673–674
Semantic rules, verifiability, 852
Semantics
  Carnap, 85–86, 88
  cognitive science representations, 128–129, 130
  complex symbols, 153
  determinism, 199
  empiricism, 235
  explication, 293
  Hempel, 350, 351
  incoherence of untranslatability, 372
  incommensurability, 371
  instrumentalism, 400, 401, 402
  Kuhn on language, 428
  linguistics, philosophy of, 446–448
  molecular biology, 489
  observation sentences, 666–668
  probability theory, 600
  Putnam, meaning and mind, 623–625
  quantum logic, 634, 635, 640–641
  Quine and, 660
  Ramsey, theory structure, 679–680
  realism, 686, 691
  representational function, 741–743
    data, 743
    phenomena, 741–743
    theory, 743
  scientific models, 740, 746
  verifiability, 853, 854
  visual representation, 866
Semantics of physics, Reichenbach and, 705–706
Semantic view of theory structure, 822–826
  critics of, 826–827
  models of, 824–826
  theory of natural selection, 314
Semantic holism, Quine, 660
Semi-Kantianism, Poincaré and, 568
Semiotic-semantic property, cognitive science, 130
Sen, Amartya, 229
  decision theory, 184
  economics, 225, 227
Sense data
  American Structuralism, 107
  and argument from illusion, 547
  Carnap, versus elementary experiences, 82
  empiricism, 93, 242, 525
  Hanson, on observation, 345
  Mach bands, 467
  observation, 524, 529
  perception, 545–550
    argument from illusion, 547
    observation, 527–528

Sense data (*Continued*)
        outstanding problems, 550
        states, experience, sensation, 545
    phenomenalism, 551, 552, 553
    protocol sentences, 610
    psychology, philosophy of, 616
    Quine, versus occasion sentences, 666
    scientific domains, 733
Sensibilia, 722
Sensory modalities, 550
Sensory physiology
    Mach, 467
    neurobiology, 518–519
Sentences
    argument, 284
    axiomatic method, 357
    cognitive significance, 136, 137
    confirmation theory, Bayesian, 148–150
    empiricism, 238
    generative grammar, 108
    Harman's case against analyticity, 15, 16–17
    Hempel, 348, 351
    intentionality, 405–406
    logical empiricism, 461
    observation, 666
    Popper and, *see* Popper, Karl R.
    probability theory, 599–600
    Putnam and, 622
    Quine and, 660
    reduction, 524
    Reichenbach and, 707–708
    relations among, Hempel and, 350
    Russell and, 720–721
    well-formed, 109
Sequence of events, probability, 601–602
Sequences, visual representation, 865
Sequential analysis
    statistics, 808
    Wald, 44
Sequestered modular templates (SMT) model of cell, 488
Serial formatting, visual representation, 865
Serres, Antoine Étienne Reynaud Augustin, 353
Set theory/set-theoretical models, 342, 343
    explication, 288
    Hilbert and, 360
    Poincaré and, 570
    quantum logic, 634, 640
    Ramsey and, 680
    Russell and, 718
    Schlick and, 726
    scientific models, 744
    theory structure, 827
    verisimilitude, 855
    von Neumann and, 504–505
Severe test of theory, 591
Severity assessments, 811–812, 814
Sex chromosomes, 331
Sextant, 34
Sextus Empiricus, 193
Sexual reproduction
    individuality, 375–376
    reproductive isolation, 797–798
Shäfer, Lothar, 862
Shafir, Eldar, 184, 185
Shagrir, Oron, 820
Shannon, C.E., 484, 508
Shannon entropy, 54
Shape, molecular biology, 483
Shapere, Dudley, 371–372
    Kuhn, 420
    observation, 526
    realism, 687
    scientific domains, 736
    scientific revolutions, 760
Shapin, Steven, 268
Shapley, Harlow, 36
Sharp, P., 483
Sheppard, Paul, 254
Shift argument, Leibnitz, 555
Shimony, Abner, 87
    emergence, 234
    locality, 454, 455
    quantum measurement problem, 647
    quantum mechanics, 655
    reductionism, 701
Shoemaker, Sydney
    causality, 98
    consciousness, 161
    laws of nature, 442
    physicalism, 565
Shomar, T., 827
Shore, J.E., 54, 55
Shrader-Frechette, Kristin, 72, 216, 220
Sidgwick, Henry, 716
Signal processing, 425
Signal relation, Reichenbach and, 705
Signal-to-noise ratio, 517
Significance, cognitive, *see* Cognitive significance
Significance, statistical
    significance levels
        Bayesianism, 45
        N-P tests, 805
        stopping rule, relevance of, 808
    significance tests
        controversy about, 808–809
        fallacy of non-statistically significant results, 809
        Fisherian, simple, 804
        Neyman-Pearson and Fisher debates, 806
        social sciences, 808–809
Significance criterion, cognitive significance, 133, 134–138
Silpher, V.M., 36
Silverstein, Arthur, 366
Simberloff, Daniel, 216, 220, 222
Similarity approaches, verisimilitude, 856
Simmel, Georg, 704
Simon, Herbert A.
    artificial intelligence, 28, 29
    cognitive science, 123
    economics, 227
    psychology, philosophy of, 615–616
    scientific change, 731
Simple recurrent network (SRN), 152–153
Simplicity postulate (Jeffreys), 535
Simplicity, *see also* Unity and disunity of science
    geometry, 360
    laws of nature, 442
    parsimony, 533–534; *see also* Parsimony
    Ramsey and, 680
    Reichenbach and, 707

Simpson, George G.
  evolution, 253
  species, 798, 800
Simpson's paradox, 95–96, 281
Simulacrum account of explanation, 748
Simulations
  artificial intelligence, 29, 30
  experimental method, 271
  scientific computing and, 508
  scientific models, 745
Simultaneity, 830
  absolute, 787
  conventionalism, 170–171
  Einstein's definition of, 705
  epistemology, 791
  Newtonian, 786
  relativity of, 789
  verifiability, 852
Sincerity condition, Searle's, 768, 769
Singer, Isadore, 508
Singer, M.C., 254
Single-cell recordings, neurobiology, 514–516
Singular causation, 90–91
Singular causes, 90
Singularist accounts, causality, 95, 98–99
Singularities
  classical mechanics, 122
  determinism, 201, 202
  physical sciences, philosophy of, 555
Sister-group relations, species, 798
Situated cognition, 125, 618–619
Situatedness, feminist philosophy, 302
Skeptical solution, prediction, 587, 592
Skepticism
  innate/acquired distinction, 395
  physicalism, 560–561
    responses to, 561–562
  probability, 606
Skinner, B.F., 61, 62
  Chomsky and, 114
  evolutionary epistemology, 261
  unity of science movement, 848
Sklar, Lawrence
  irreversibility, 410, 411
  kinetic theory, 419
  physical sciences, philosophy of, 555
  time, 830
Skyrms, Brian, 10
  Bayesianism, 41
  decision theory, 183
  induction, problem of, 381
  inductive logic, 384, 385, 386, 387
  laws of nature, 439
  probability theory, 600, 609
  Ramsey, 675, 677, 679
  Ramsey and, 680
Slater, L., 102
Slatkin, Montgomery, 254
Slovic, P., 227
Small-population paradigm, conservation biology, 164
Smart, J.J.C., 560
S-matrix program, particle physics, 541–542
Smith, Barry, 626
Smith, C.A.B., 43
Smith, Cantwell, 619
Smith, Dorothy, 298
Smith, George E., 118
Smith, M., 264
Smith-Birnbaum inequality, 44
Smolensky, Paul, 153, 154
Smoluchowski, M., 417
Smorodinsky, M., 328
SMT model of cell, 488
Smyth, P., 731
Sneath, P.H.A., 796
Sneed, J.
  approximation, 25
  theory structure, 824, 826
Snell's law, 276–277, 400
Soames, S.
  cognitive science, 127
  linguistics, philosophy of, 449
Sober, Elliott, 254, 255, 497
  adaptation and adaptationism, 5
  altruism, 8, 9, 10
  Bayesian confirmation theory, 149
  biology, philosophy of, 70, 73
  causality, 91
  fitness, 311, 312, 313
  function, 319
  instrumentalism, 404
  linguistics, philosophy of, 449
  natural selection, 498, 500, 501
  parsimony, 532, 535, 536, 537
  scientific models, 743
Sobstyl, Edrie, 302
Social, science as, 302
Social behavior
  evolutionary psychology, 263–267
  Nagel and, 495–496
  neurobiology, 521
Social constructionism/constructivism, 774–779
  construction of science, 776–778
  demarcation problem, 189, 194
  diffusion of constructionism, 775–776
  experimental method, 271
  feminist philosophy, 296, 300
  science wars, 778–779
  scientific change, 730
  scientific revolutions, 756, 764
  social sciences, 784
Social Construction of Reality, The (Berger and Luckmann), 774
Social Darwinism, 500, 728
Social engineering, piecemeal (Popper), 575
Social environment/reality/context
  cognitive significance, 350
  consciousness, 160
  demarcation problem as social problem, 194–196
  economic theory, assumptions of, 228
  feminist philosophy, 296
  function, 321
  Kuhn's scientific revolutions, 424
  logical empiricism, 464
  Mach, 468
  Popper, social philosophy of, 575
  scientific change, 730
  scientific style, 765
  Searle, 767–774
  and self-consciousness, 160

Social explanations, 479
Social/public policy
  Carnap, role of science, 83
  demarcation problem, 195–196
  genetics/eugenics, 334
  neurobiology, 514
  species concept, importance of, 801
Social sciences
  atomistic body, 364–365
  Carnap, 83
  causality, 90
  cognitive science multidisciplinary approach, 123
  economics, 224–230
  evolutionary psychology, 263–267
  experimentation/experimental method, 268, 270, 274
  feminist philosophy, 296
  game theory, 507
  logical positivism, 460
  methodological individualism, 478
  Nagel, 491, 494–496
  Neurath, 510–514
    empiricism, 510–512
    unity of science and encyclopedia model, 513–514
  philosophy of, 780–785
    foundational problems, 783
    new directions, 783–784
    varieties of social explanation, 780–783
  Popper, 574, 575
  prediction, 596–598
  scientific revolutions, 755
  scientific style, 765–766
  statistics, 808–809
  unity and disunity of science, 846
Social Theory after Cognitive Science (Turner), 783
Society
  feminist philosophy, 297
  Kuhn's disciplinary matrices and scientific revolution, 425–427
  methodological individualism, 478–479
Sociobiology, 70, 263, 264; *see also* Adaptation and adaptationism; Altruism; Evolutionary psychology
  adaptation and adaptationism, 6
  biology, philosophy of, 69
  feminist philosophy, 298–299
Sociology
  atomistic body, 364–365
  cognitive science multidisciplinary approach, 123
  experimental method, 274
  feminist philosophy, 298
  methodological individualism, 478
  Nagel and, 494, 496
  scientific knowledge, 776
  scientific progress, 750
  scientific revolutions, 758
Sociology of knowledge, 784
Sociology of science, experimental science, 269
Sociology of scientific knowledge, 776
Sociopolitical contexts, *see also* Politics
  game theory, 507
  Mach, 468
  Neurath and, 511
  scientific revolutions, 754, 755, 756
  unity and disunity of science, 845
Socrates, 397
Soft artificial intelligence, 30, 32
Sokal, Alan, 777, 778
Sokal, R.R., 796
Solar neutrino experiment, 528
Solar system model of atom, 651
Solipsism
  logical empiricism, 523
  phenomenalism, 553
Solomon, Miriam, 301
Sommerfeld, Arnold, 652, 704
Sophisticated regularity accounts, laws of nature, 442–443
Sorting techniques, neurobiology, 515
Sosa, Ernest, 247
Soskice, J.M., 738
Soulé, M.E., 163
Sound and speech perception, 550
Sovereign Reason (Nagel), 492
Space
  Carnap and, 80, 82
  causality, 82, 94
  classical mechanics
    conservation of energy, 119
    variational problem, 120
  empiricism, 237
  instrumentalism, 400
  Mach and, 468
  mechanism representation, 472
  Nagel on, 493
  physical sciences, philosophy of, 555
  probability, 599
  reductionism, 697–698
  Reichenbach and, 705–707
  species, evolution of, 801
  visual representation, 864–865, 866, 868
Space, absolute, 787
Space, phenotypic, 795–796
Space-like hyperplane, determinism, 206
Space-like separated events, 452
Space-like vectors, 791
Space-time, 785–795
  classical mechanics, 117, 118
  complementarity, 141
  conventionalism, 171–173
  determinism, 200, 205–207
  Feyerabend and, 306
  in general relativity, 791–794
  Hanson and, 346
  Hempel and, 349–350
  Hilbert and, 359
  instrumentalism, 400
  locality, 452
  locality/local action, 451
  locality/nonlocality, 457
  logical positivism and, 460
  Mach and, 468
  Newtonian, 786–788
  particle physics, 539
  physical sciences, philosophy of, 554, 555
  quantum field theory, 630, 631
  quantum logic, 642
  realism, 692–693
  reductionism, 698
  Reichenbach and, 705–707
  special relativity, 788–791
  time, 829–832

  underdetermination of theories, 841
  von Neumann and, 506
Space and Time in Contemporary Physics (Schlick), 728
Space-time manifold, 544
Spann, Othmar, 860
Spatial interlevel reduction, 697
Spatial metrics, Newtonian space-time, 791
Spatial organization of mechanisms, 470–471, 474
Spatial properties, Hanson and, 346
Spatiotemporal continuity, individuality, 377
Spatiotemporal existence, intentionality and, 406
Spatiotemporal location
  Bohr and, 142
  individuality, criteria for, 377
  species, 801–802
  species taxa, 796
Spatiotemporal quantity, 790
Special consequence condition, confirmation theory, 147
Special relativity
  determinism, 207
  emergence, 234
  Hilbert and, 359
  inertial frames in, 790
  locality/nonlocality, 452, 455, 456–457
  particle physics, 538–539
  reductionism, 698
  relativistic space-time, 205
  space-time, 786, 788–791, 792
  time, 829–830
Speciation, mechanism of, 253
Species, 795–802
  evolutionary species concept, 798
  evolution of, 801–802
  importance of, 801
  individuality, 374–378
  origins of, Devries' *Mutationstheorie*, 53
  phenetic species concept, 796–797
  phylogenetics, 798–799
  pluralism, 800–801
  reproductive isolation, 797–798
  species category, 796
  species taxa
  systematic biology, philosophy of, 71
  unified species concept, 799–800
Species-deletion stable, ecological communities, 219
Species diversity
  evolution, 254
  heritability, historical development of idea, 353
Species richness, ecological communities, 218
Specific heat anomaly, kinetic theory, 416–417
Specificity
  immune self, 366
  maximal, principle of, 279, 603
  scientific domains, 733
Specker, Ernest, 453
  quantum logic, 642, 643
  quantum measurement problem, 648
  quantum mechanics, 656
Spectrometer/spectroscopy
  astronomy, 35–36, 37
  chemistry, philosophy of, 102, 103
  dumbbell model and, 417
  quantum conditions, 652
  wave mechanics, 652–653
Speculation
  cognitive significance, 132
  demarcation problem, 189
  Lakatos and, 436
  Schlick and, 727
Speech acts, Searle on
  and intentionality, 768–770
  structure of, 768
Spencer, Herbert, 310, 500
Sperry, R.W., 231, 702
Spin angular momentum, 104
Spin properties, subatomic particles, 539
Spin statistics theorem, 631
Spirtes, P., 96, 97
Splicing, RNA, 482
Splitters versus lumpers, species, 801
Spontaneous generation, 352
Spontaneous symmetry breaking, 542
Sprott, D., 807
Spurious causes, 94
Stability
  classical mechanics, 116
  ecological communities, 218–219
  ecology and conservation biology, 73
  game theory, evolutionary, 329
Stace, W.T., 133
Stachel, John, 639, 643, 830
Stadler, Friedrich, 858, 861, 862
Stadler, L.J., 332
Stahl, Leslie, 335
Stalker, Douglas, 383
Stallo, John B., 417
Stalnaker, Robert, 408, 623
Stalnaker/Lewis semantics, 600
Stampe, Dennis, 408
Standard dynamical law of quantum mechanics, 655
Standard interpretation of pure states, quantum measurement problem, 645
Standard Model of particle physics, 22
  anthropic principle, 21–23
  and quantum field theory, 542–544
Standards, experimental, Kuhn's disciplinary matrix, elements of, 421
Standards for rationality, probability, 606
Standard Social Science model (SSSM), 264
Standpoint approaches, feminist philosophy, 302
Stanford, Kyle
  realism, 693, 694, 695
  underdetermination of theories, 839, 841
Stanford School, unity and disunity of science, 846
Stanley, Julian C., 269
Stapp, Henry, 456
State change, discontinuous, 648
State consciousness, defined, 158
State-dependent utility theory, 182–183
State descriptions
  inductive logic, 388
  probability theory, 603, 604
State functions, quantum logic, 637
Statements
  analyticity, Harman's case against, 15–16
  empiricism, 238–239, 524
  Hempel, 348, 524
  observation, 133
  protocol sentences, 610

Statements (*Continued*)
  Reichenbach and, 706
  Schlick and, 726
  scientific versus nonscientific, 535–536
  verisimilitude, 857
State reduction, determinism, 204
States, perceptual, 545
States, quantum
  locality, entanglement and EPR, 452–453
  quantum measurement problem, 644–646
State space
  dynamical systems theory, 154
  locality/nonlocality, 457
  quantum field theory, 630
  quantum mechanics, 656
Static representation, mechanism, 472
Statics
  dynamics versus, 116
  language, Lakatos and, 436
  reduction of dynamics to, 120
Statistical character of neuronal activity, 509
Statistical descriptions, complementarity, 143
Statistical distribution function, kinetic theory, 416
Statistical explanation
  explanation, 276, 278–281
  irreversibility, 412
Statistical generalizations of social phenomena, Nagel and, 494
Statistical mechanical entropy, 412
Statistical mechanics, *see also* Kinetic theory
  classical, 121, 650
  irreversibility, 411
  kinetic theory, 415, 418
  physical sciences, philosophy of, 554
  probability, 602, 607
Statistical methods, experimentation in social sciences, 269
Statistical physics
  classical mechanics, 122
  probability, 607
Statistical relevance model, 280–281, 283
Statistics
  Bayesian, *see* Bayesianism
  Carnap, 88
  experimental method, 270
  heritability, 353, 354–355
  induction, problem of, 382
  kinetic theory, 416
  Nagel on laws versus generalizations, 495–496
  scientific revolutions, 755
  scientific style, 766
Statistics, philosophy of, 802–815
  first wave, 1930–1960, 803–807
    alternative (non-null) hypothesis, 804–805
    confidence interval estimation procedures, 806
    dualities between one- and two-sided intervals and tests, 806
    Fisher and Neyman-Pearson debates, 806
    Fisherian simple significance tests, 804
    Fisher's criticism of confidence intervals, fiducial intervals, 806–807
    inductive behavior philosophy, 805–806
  probability, two roles in inference, 803
  second wave, 1960–1980, 807–810
    error probability principle versus likelihood principles, 807–809
    power analytic movement, 809–810
    significance testing controversy, 808–809
    stopping rule, relevance of, 808
  third wave, 1980 to present, 810–815
    Bayesian advances and controversy, 810
    error statistics, reforms within, 811
    fallacies of rejection, large N problem, 812
    meta-methodology, 810
    p-values and Bayesian posteriors, 812–814
    severity assessments, 811–812
Stauber, John, 195
Steady State cosmology, 36, 37
Stebbing, Susan, 848
Stegmüller, W., 87, 824, 826
Stein, Howard
  Carnap, 88
  conventionalism, 172, 174
  instrumentalism, 403, 404
  quantum mechanics, 654
  space-time, 787
  time, presentism, 831
Steinle, Friedrich, 269, 273
Stengers, I., 102
Stephens, C., 486
Stephens, L., 618
Sterckx, Sigrid, 274
Sterelny, Kim, 220
  adaptation and adaptationism, 3
  biological information, 66, 67
  biology, philosophy of, 73
  fitness, 310–311
  innate/acquired distinction, 397
  natural selection, 498
Stereospecificity, molecular biology, 483
Stern-Gerlach experiments, 637
Stern-Gerlach magnets, 453
Stewart, John, 367, 368
Stich, Stephen P., 397, 615
Stigler, Stephen, 46, 384
Stiglitz, Joseph, 188
Stimulus-response pairs, 614
Stipulation
  conventionalism, 174
  explication, 287
  Harman's case against analyticity, 15
Stochastic models of social science, Nagel and, 495–496
Stohr, Adolph, 342
Stokes, G., 577
Stoll, Robert R., 633, 634
Stöltzner, Michael, 120, 503, 505, 506
Stolzenberg, Gabriell, 778
Stone, Abraham, 656
Stone-von Neumann uniqueness theorem, 506
Stopping rule, relevance of, 808
Stosszahlansatz, kinetic theory, 418
Stotz, Karola, 72
Straight rule of induction, 381–382
Strange attractor, 595
Strangeness, 540
Strategic agent, game theory, 323
Strategic (normal) form, game theory, 324, 325
Strategy, game theory, 507
  evolutionary, 328
  noncooperative games, 325–327
Strauss, Martin, 633, 639
Strawson, Galen, 94

Strawson, P.F., 98
  analyticity, 15, 19
  explication, objections to, 290–292
  induction, problem of, 380, 381
  linguistics, 111
  Quine, 660
Stringency, explanation, 284
String theory, 413, 794
Strong, Donald, 222
Strong anthropic principle (SAP), 22
Strong artificial intelligence, 30, 31–32, 771–773
Strong equivalence, artificial intelligence, 30–31, 32
Strong interactions, models of, 541, 542
Strong nuclear force, particle physics, 539
Strong objectivity, feminist philosophy, 301
Strong program, sociology of scientific knowledge, 776; *see also* Social constructionism/constructivism
Strong reduction, 697
Structural decomposition, 473
Structural genes, molecular biology, 481
Structural heterogeneity, individuality, criteria for, 377
Structural identity thesis, 277
Structuralism/structural theory, 774
  approximation, 25
  theory structure, 826
Structural psychology, behaviorism and, 61–62
Structure
  consciousness, 160
  linguistic, 109
  molecular biology, 485
  scientific progress, 751
  scientific style, 766
  transformational grammar, 107
Structure description, probability theory, 604
Structure-function rule, molecular biology, 483
Structure of science
  Kuhn's disciplinary matrix, elements of, 421
  Nagel and social sciences, 494–496
Structure of Science, The (Nagel), 492, 493, 495, 496
Structure of Scientific Revolutions, The (Kuhn), 756, 760–762
  evolutionary epistemology, 259–260
  incommensurability, 370–371
  Kuhn, 419, 420, 421, 422, 428, 429, 430
  logical empiricism, 464
  research programs, 712
  scientific change, 730
  scientific revolutions, 754
  social constructionism, 776
  unity of science movement, 848
Structure of Scientific Theories, The (Suppe), 421
Strutt, J.W. Lord Rayleigh, 650
Student t test, 804
Studies in the Logic of Confirmation (Hempel), 348
Study models, 747
Stufflebeam, R., 514
Stump, David, 571
Stumpf, C., 704, 767
Sturgeon, Scott, 560
Sturtevant, Alfred H., 332
Styles of Scientific Thinking in the European Tradition (Crombie), 766
Suárez, Mauricio, 744, 827
Subgame perfect equilibrium, 327
Subjective expected utility maximization, 181
Subjective Bayesianism, probability, 606
Subjective expected utility representations, 183
Subjective perception, von Neuman's theory of measurement, 505–506
Subjective probabilities, 179
  confirmation theory, Bayesian, 148, 149–150
  Ramsey, 675
  Ramsey and, 680
Subjective states, perception, 547
Subjective utility theory, Ramsey and, 680
Subjectivism
  Bayesians, 55
  probability, 604–606, 607
Subjectivity
  causal self-referentiality, 769–770
  inductive logic, 392
  protocol sentences, 611
Subjunctive conditionals, 177, 178
Subjunctive conditionals, probability theory, 607
Suboptimal adaptation, 5
Subsistence, intentionality and, 406
Subspaces, quantum logic
  classical mechanics and, 634
  Hilbert space, 633
  orthomodular lattice logic, 638–639
Substance, individuality, criteria for, 377
Subtraction method, neurobiology, 517–518
Subvenient properties, 816, 819
Succession, ecological, 215–216
Succession laws, theory structure, 824
Succession rules, Bayesianism, 50
Successive-approximation methodologies, demarcation problem, 190
Sufficient parameter concept, 701
Sullivan, Peter, 672
Summers, Stephen, 457, 506
Superorganisms, ecological communities as, 216
Superposition
  improper mixtures, 647
  quantum measurement problem, 645–646, 647
  quantum mechanics, 654, 655
Superselection rules, quantum field theory, 632
Superstring theory, 544
Supervenience, 815–820; *see also* Mind/body problem
  biology, philosophy of, 69
  cognitive science, 131
  connectionism, 155–156
  determination and explanation, 816–817
  economics, laws of, 226
  emergence, 232, 233
  global, 819–820
  individual comparison supervenience, 817–818
  neutrality, 817
  physicalism, 563–565, 567
  reductionism, 701–702
  taxonomy, 817
  unity and disunity of science, 845
Suppe, Frederick
  Duhem thesis, 210
  feminist philosophy, 296
  Kuhn, 421
  neurobiology, 515
  theory structure, 827
    semantic view, 824
    structuralist view, 826
    syntactic (received) view, 823, 824

Suppes, Patrick, 96, 351
  causality, 90
  determinism, 198, 199, 200, 207
  scientific models, 743
  theory structure, semantic view, 824, 825
Sur, M., 517
Sure-thing principle, decision theory, 185–188
Surface structure, transformational grammar, 107
Surrogacy, conservation biology, 165–166
Sustaining technologies, 762
Sutton, Walter S., 353, 699
Svozil, Karl, 643
Swiss army knife model (massive modularity thesis), 266
Swoyer, C., laws of nature, 442
Syllogism, disjunctive, rule of, 717
Symbolic logic
  empiricism, 237
  Poincaré and, 570
  Russell, 715–723
  Russell and, 718–719
  Vienna Circle and, 859
Symbolic models, connectionism, 153–154
Symbolic theory, classical mechanics, 121
Symbol processors, cognitive science, 128
Symbols
  computation, 28
  Russell and, 721
  semantic paradoxes, 673–674
  unity and disunity of science, 847
  visual representation, 864–865, 867
Symmetry
  game theory, 328
  particle physics, 539, 540–541
    spontaneous symmetry breaking, 542
  physical sciences, philosophy of, 555
  sociology of scientific knowledge, 776
Symons, D., 264
Sympatry, 254
Synonymy
  Carnap, 86
  language-transcendent concept of, 14–15
  Quine and, 661, 662
Syntactic characterizations, Hempel, 350
Syntactic instrumentalism, 402–403
Syntacticism, Carnap, 84–85
Syntactic phase, Carnap, 83–85
Syntactic relations
  Hempel, 348, 351
  protocol sentences, 612
  scientific models and theory, 746
Syntactic (received) view, theory structure, 822–824
Syntax, 107
  arithmetization of, 84
  explanation, 276
  protocol sentences, 612
  visual representation, 866
Synthesis, evolutionary, 253
Synthetic/analytic disjunction
  Carnap, 86
  Duhem thesis, 209
  empiricism, 238–239
  Quine and, 659
Synthetic *a posteriori* propositions, Ayer and, 39
Synthetic geometry, 357
Synthetic statements, verifiability, 852
Systematic biology, 71, 253
Systematic calculus, Nagel on, 493
System of Logic, A (Mill), 231

T

Tabery, J., 488
Tabula rasa, 236
Tacit knowledge, *see* Experiment
Takacs, D., 163
Talmage, David, 366
Tanner, Nancy, 298
Tarski, Alfred
  Carnap, 84–85
  cognitive significance, 135
  explication, 290
  Popper and, 576
  Quine and, 659
  Vienna Circle, 858, 862
Tatum, Edward L., 334
Tauber, Alfred I., 366, 367, 368, 485
  immunology, 364, 365
  molecular biology, 485
Taussky-Todd, Olga, 858
Tautologies
  analyticity, 12
  Ayer and, 40
  biology, 69
  conventionalism, 175
  Hahn, 342
  phase space, 634
  quantum logic, 635, 639, 641–642
Taxa, monophyletic, 799
Taxa, species, 795–796, 802
Taxonomic incommensurability, 371
Taxonomy, *see also* Individuality; Species
  molecular methods, 253
  scientific style, 766
  supervenience, 817
Taylor, A., 399
Technology, *see also* Experiment
  artificial intelligence, 29–30
  chemistry, philosophy of, 102, 105–106
  demarcation problem, 196
  experimental method, 268, 269, 271–272, 274
  philosophy of, 275
  reductionism, 700
  revolutions in, 761–762
  scientific style, 766
  Turing and, 833–834
Teleofunctional theories, cognitive science, 129
Teleological causality, 90
Teleological explanation, 285–286
Teleology, *see also* Explanation; Function
  biology, philosophy of, 69
  developmental biology, philosophy of, 72
  function, 315–316, 319, 320–321
  generative grammar, conceptual issues in, 112
  Nagel and, 494
  psychology, philosophy of, 615
  rational reconstruction, 685
Teleology Revisited (Nagel), 492, 494
Teleosemantics, 66
Teleportation, 456

Telescope, 34–35
Teller, Paul, 631
  Dutch Book argument, 213
  emergence, 234
Tempelton, A.R., 799–800
Template models of antigen-antibody reactions, 366
Temporal ordering, *see* Space-time; Time
Tennyson, Alfred, 716
Tensor product networks, defined, 154
Tensors, classical mechanics, 121
Teratology, 353
Terms
  reduction sentences, Carnap and, 524
  theory structure, syntactic (received) view, 823
Terms (concepts) versus propositions (statements), 524
Testability, *see also* Cognitive significance; Demarcation, problem of; Verifiability
  demarcation problem, 190
  ecological diversity/stability hypothesis, 219
  economics, 226–227
  evolutionary epistemology, 260
  Lakatos and, 435
  Nagel on, 493
  Popper and, 574–575
  scientific progress, 751
  unity and disunity of science, 844
Testability and Meaning (Carnap), 83
Testing/tests
  Bayesianism, 43–45
  observation, 528
  protocol sentences, 612
  scientific models, 740
  statistical, *see* Statistics, philosophy of
  verifiability, 853
Tevatron, 538
Thagard, Paul, 1, 761
Thau, Michael, 406
Themata
  scientific metaphors, 738
  scientific style, 765, 766
Theological mathematics, 673
Theorems, Lakatos and, 435, 438
Theoretical chemistry, 103
Theoretical holism, Neurath, 512
Theoretical language, Carnap, 86
Theoretically progressive programs, 714
Theoretical pluralism
  ecology, 221
  Feyerabend, 307–308
Theoretical terms, reduction sentences, 524
Theoretician's dilemma, 62, 402–403
Theoretician's Dilemma, The (Hempel), 350
Théorie de la probabilité (Laplace), 708
Theory change/succession
  evolutions, *see* Scientific progress
  natural selection model, 261
  reductionism, 698–699
  revolutions, *see* Scientific revolutions
Theory choice, laws of nature, 442–443
Theory of conditional probability, 386
Theory of content determination, 129, 131
Theory of coordinative definitions (Reichenbach), 705–706, 707
Theory of equivalent descriptions (Reichenbach), 706, 710
Theory of Games and Economic Behavior (von Neumann), 507
Theory of invariants, 392
Theory-laden factual representation, Hanson on, 346
Theory-laden observation
  Hanson on, 344–345
  neurobiology, 514–516
  Popper and, 573
  social constructionism, 777
Theory of laws, Lewis, 94
Theory of meaning, demarcation problem, 191
Theory of mind, 617
Theory nets, 826
Theory of practical rationality, 183
Theory of reference
  incommensurability, 372–373
  Putnam, 622
Theory of Relativity and A Priori Knowledge (Reichenbach), 705
Theory of scientific method, Popper and, 572–576
Theory/theories, *see also specific theories*
  approaches to structure of, 822–828
    semantic view, 822–826
    structuralist view, 826
    syntactic (received) view, 822–824
  artificial intelligence, 31–32
  biology, philosophy of, 70–71
  Carnap on, 86
  chemistry, philosophy of, 102, 104–105
  cognitive science, pretheoretical specification of domain, 123–124
  cognitive significance, 134–140
  complementarity, 143
  confirmation, 144–150
  deterministic, 199
  ecology, 220–223
    evaluation of models, 221–222
    explanatory value of models, 222–223
    feasibility of models, 220–221
  epistemic change, 244, 249–250
  evolutionary biology, philosophy of, 70–71
  evolution of epistemic theories (EET), 258
  experimental method and, 271, 272, 273
  feminist philosophy, 296
  Feyerabend and, 305–307
  generative grammar, conceptual issues in, 110–111
  Hanson and, 344, 346
  Hempel, 350–351
  incommensurability, 371, 372
  instrumentalism, 400–402
    epistemic, 402, 403–405
    reductive, syntactice, and eliminative, 402–403
  intentionality, 408–409
  Kuhn's scientific revolutions, 419–431
  laws of nature, 442–443
  logical empiricism, 462–463
  methodological individualism, 478
  molecular biology, philosophy of, 71
  Nagel on, 493
  Neurath and, 514
  neurobiology, 518–520
    framework, 520
    theory-laden observation, 514–516
  and observation, 527
  operationalism, 76–77
  parsimony, 531–538

Theory/theories (*Continued*)
  particle physics, 538–539, 543
  phenomenalism, 551–552
  physical, foundations and interpretations of, 554–555
  physicalism, causal impact argument, 566–567
  Popper and, 572–575
  prediction, 592; *see also* Prediction
  Quine and, 660, 664, 665, 666–668
  Ramsey, 679–680
  realism, 687, 688, 689–690
  reductionism, 696–697
  research programs, 712–715
  scientific domains, 733–734, 736
  scientific metaphors, 737–740
  scientific models, 740, 743, 746–747
  scientific progress, 749–753
  underdetermination of, 839–842
    conventionalism, 169
    feminist philosophy, 296
  unity and disunity of science, 844
  verisimilitude, 855, 856, 857
  visual representation, 867
Theory of types
  Ramsey, 672–674
  Russell, 715, 718
Theory of World Views (Dilthey), 765
Thermodynamics, 525
  experimental-mathematical, 268
  Hilbert and, 359
  irreversibility, 410–413
  kinetic theory, 415–416, 418
  Nagel on, 493
  physical sciences, philosophy of, 554
  quantization and, 651
  reductionism, 696, 700
  unity and disunity of science, 845
  von Neumann and, 507, 508
Thiéffry, D., 483
Third dogma of empiricism, 764
Third world, Popper and, 575–577
Thirring, Walter, 121, 122
Thisness, quantum field theory, 631
Thistle, A.B., 222
Thomas, J.W., 166
Thompson, D'Arcy W., 109
Thompson, G.I., 508
Thompson, P., 825
Thomson, William, Lord Kelvin, 417
Thorndike, Edward, 61
Thorne, Kip S., 830
Thought
  adaptive economy of, 844
  artificial intelligence, 722
  innate/acquired distinction, 395–396
  vision and, 272
Thought collective (*Denkkollectiv*), 766
Thought experiments, 271
  irreversibility, 411
  Lakatos and, 436
  Maxwell's demon, 418
  physicalism, 560–561
  Quine and, 668–669
  Reichenbach double slit, 710
  scientific models, 745
  Searle, 772–773
  Turing test, 836
Thought style (*Denkstil*), 765
Three-valued logic, Reichenbach, 710–711
Thyer, Bruce A., 63
Tiao, George C., 59
Tichy, Pavel, 855
Tiles, Mary, 272
Tilman, David, 219
Time, 829–832; *see also* Space-time
  absolute, 205
  asymmetry of, 831–832
  beginning of, 831
  and general relativity, 830–831
  Hilbert, 359
  instrumentalism, 400
  Mach and, 468
  mechanism
    organization, 470–471
    representation of, 472
  physical sciences, philosophy of, 554
  presentism, 831
  quantum logic, 634
  Reichenbach and, 705–707
  spatiotemporal location
    Poincaré, 569
    species, 801–802
    species taxa, 796
  spatiotemporal quantity, 790
  special relativity, 829–830
  species, evolution of, 801–802
  temporal metrics, Newtonian space-time, 791
  temporal ordering
    Carnap, 81
    Ramsey on causal laws, 680
  temporal organization of mechanisms, 470–471, 474, 475
  temporal orientation, space-time, 830
  visual representation, 864
Time, direction of, Reichenbach and, 709
Time asymmetry, irreversibility, 411
Time axis representation, Hanson and, 346
Time evolution, dynamical systems theory, 122
Timelike curves, 830
Timelike vectors, 791
Time-reversed processes
  determinism, 201, 202
  prediction, 595
Time-slice generation, 830
Timing, psychology, philosophy of, 617
Timms, Edward, 858
Tinbergen, Niko, 263, 264
Tipler, F.J., 21
Titchener, Edward, 61
Tolerance, principle of, 862
  cognitive significance, 136
  conventionalism, 175
  Duhem thesis, 209
  explication, 290
  Vienna Circle, 862
Tonsor, S.J., 582
Tooby, John
  evolutionary psychology, 264, 265, 266
  function, 319–320

Tooley, Michael, 97, 98
  Dretske-Tooley-Armstrong account, 98
  laws of nature, 440, 441, 442, 443
Topology
  and classical mechanics, 121
  molecular biology, 484, 487
  relativistic space-time, 205
  space-time, 793
Total evidence requirement, epistemology, 248–249
Totalitarianism, Vienna Circle, 860
Toulmin, Stephen
  evolutionary epistemology, 259, 260, 261
  Kuhn, 429
  scientific change, 730
  scientific progress, 753
  scientific revolutions, 761
  scientific style, 766
Toumey, Christopher, 195
Townsend, James T., 155
Toy models, 747
Traaweek, Sharon, 430
Tractatus Logico-Philosophicus (Wittgenstein), 38, 83, 460
  conventionalism, 175
  demarcation problem, 189, 191
  Hahn and, 342
  Ramsey and, 671
  Schlick and, 728
  Vienna Circle, 859
Tradition
  Feyerabend on, 308–309
  scientific progress, 751–752
Training, Kuhn's scientific revolutions, 426
Traits
  adaptation and adaptationism, 3–7
  evolutionary biology, philosophy of, 70
  genetics, 330, 334
  innate/acquired distinction, 398–399
  natural selection, 497
  phenotypic, 796
Transcendent individuality, quantum field theory, 631
Transcription, DNA, 335
Transcription factors, 337
Transfer RNA, 336
Transformational grammar, 107
Transformation of science, *see also* Scientific change
  scientific domains, 736
  scientific revolutions, *see* Scientific revolutions
Transformation of species, 353
Transformations, Galilean, 788, 789
Transformations, Lorentzian, 789
Transgenic animals, 339
Transients, dynamical systems theory, 154
Transitive versus intransitive consciousness, 159
Transitivity
  causality, 92
  decision theory, 185
  game theory, 324
Translation, indeterminacy of, 14
Translation, language, 428
  incoherence of untranslatability, 372
Translation, protein synthesis, 335
Transport phenomena, Maxwell's theory, 416
Treatise of Human Nature (Hume), 47
Tree switching, 761
Treves, Allesandro, 156
Trials, probability, 601
Triggering, learning versus, 397
Trivers, R.L., 9
Triviality objection, physicalism, 560
Troland, Leonard, 334
Trope identity option, physicalism, 563
Trustworthiness argument, 247
Truth
  analyticity, 12
  Carnap, syntactic phase, 84–85
  confirmation theory, Bayesian, 148–150
  correspondence theory of, 576
  explicanda/explicata pairs, 288
  Hempel and, 347
  innate/acquired distinction, 395–396
  logical positivism, 460–461
  logic of, 675, 678–679
  parsimony, 531–538
  Poincaré and, 571
  Ramsey, 675, 678–679
  Reichenbach theory of simplicity, 707
  scientific progress, 753
  scientific revolutions and, 763
  scientific style, 767
  social constructionism, 776
  theory structure, semantic view, 826
  verifiability, 852
  verisimilitude, 854–857
Truth by Convention (Quine), 663, 664
Truth functional operations
  quantum logic, 634, 643
  Ramsey on, 672
Truthlikeness (verisimilitude), 854–857
Truth theory of meaning, 707
Truth value, complementarity logic, 633
T-symmetry violations, irreversibility, 411
t'Hooft, G., 542
Tuana, Nancy, 299, 739
Tucker, A., 507
Turbulence, boundary layer theory and, 119
Turelli, M., 582
Turing, Alan
  artificial intelligence, 27
  Bayesianism, 52
  life and work, 833–838
    intelligent machinery, 836–837
    machines, 834–835
    mind, 833–834
    practical machinery, 835–836
    unfinished work, 837
  von Neumann, 508
Turing machines, 29, 508, 833, 834
  cognitive science, 127–128
  determinism, 199
Turing paradox, 837
Turing's oracle, 834
Turing test, 836
Turner, J.R., 254
Turner, Stephen, 783
Tversky, Amos, 184, 227
Twin paradox, 791
Two-dimensional representations, mechanism, 472
Two-dimensional state space, quantum mechanics, 656

Two Dogmas of Empiricism (Quine), 13–14, 350, 461, 462, 660–666
Two-factor theories, cognitive science, 129
Two-force models, particle physics, 539
Two-particle models, particle physics, 539
Two-place function, probability as, 600
Tye, M., 162, 409
Type I error, 804, 805
Types, theory of
  Ramsey, 672–674
  Russell, 715, 718
Typology
  scientific style, 765
  species, 795–796

U

Über die Grundlagen der Geometrie (Hilbert), 357
Überwindung der Metaphysik durch Logische Analyse der Sprache (Carnap), 39–40
Uebel, T.E.
  Neurath, 511
  physicalism, 558
  Schlick, 726
  Vienna Circle, 861
Uffink, Joseph, 410
Ulises-Moulines, C., 25
Ullian, J.S., 767
Ultraviolet catastrophe, 650, 651
Ultraviolet spectrum, astronomy, 35
Ultraviolet telescopes, 36–37
Umkehereinwand, kinetic theory, 418
Unconditional probability, inductive logic, 386, 387
Underdetermination of theories, 839–842
  bottom-up arguments, 841–842
  Duhem-Quine thesis, 841; *see also* Duhem thesis
  empirical equivalence, 840–841
  empiricism, 239
  instrumentalism, 401
  Neurath and, 512
  Poincaré and, 568–569
  Quine and, 662
  realism, 692–693
  social constructionism, 777
  social sciences, 784
Understanding *(verstehen)*
  social sciences, 782
Ungerleider, L.G., 516
Unification approach, explanatory, 283–285
Unified descriptions of fundamental forces, 544
Unified species concept, 799–800
Unified theories, unity and disunity of science, 846
Uniformity of nature assumption, 379–380
  induction, problem of, 383
Uniqueness theorem, Stone-von Neumann, 506
Units of evolution
  individuals as, 378
  species, 796
Units of selection, 498, *see also* Evolution; Natural selection
  altruism, 8–9
  biological information, 66
  evolutionary biology, philosophy of, 70
  genes as, 9, 498
  groups as, 8, 498
  heritability, 355
  individuality, 376
  individuals as, 8, 375
  multilevel, 498
  natural selection, 498
  species as, 498
Unity
  evolutionary, 797
  logical empiricism, 463–464
  organic, Nagel on, 493–494
Unity and disunity of science, 842–847
  Carnap, 83
  feminist philosophy, 301
  incommensurability, 370–373
  Putnam and, 620
  Quine and, 659
  reductionism, 697
  scientific change, 730
  scientific domains, 736
  scientific progress, 750
  scientific style, 765, 767
  Vienna Circle, 860
Unity of consciousness, Searle and, 770–771
Unity of science movement, 845, 848–849
  Carnap, 83
  Kuhn, 420
  logical empiricism, 463
  logical positivism, 460
  Neurath, 511, 513–514
  Putnam and, 620
  Quine and, 659
  scientific domains, 736
  scientific progress, 750
  scientific style, 765
  social sciences, 780
  Vienna Circle, 860, 861
  Vienna Circle and, 859
Universal Darwinism, 258
Universal field equations, 359
Universal forces, Reichenbach and, 707
Universal generalizations, 276, 278
Universal grammar, Searle and, 770
Universal gravitation, 117–118, 279
Universalism, Vienna Circle, 860
Universals
  causality, 97–98
  laws of nature, 441–442
  unity and disunity of science, 844
Universal Turing machines, 508, 834
Unpredictability, quantum, 837
Unpredictability in classical physics, determinism, 203–204
Unrevisability, Quine and, 662–663
Untranslatability, 372
Untranslated RNA, 488
Urbach, Peter
  Dutch Book argument, 212
  induction, problem of, 382
  prediction, 593
  statistics, 810
Urrutia, A.O., 486
Uthurusamy, R., 731
Utilitarianism, Popper and, 575

Utility
  confirmation theory, decision theoretical approach, 148
  decision theory, 182
  empiricism, 238
  game theory, 323–324
  quantum logic, 641–642
  Ramsey and, 679, 680
  theory of
    economic rationality, 227
    game theory, 323–324
Utility function, 323
  probability, 605
  Ramsey and, 680
Utopian views, von Neumann, 503–504

V

Vacuum state, quantum field theory, 629, 630, 632
Vaihinger, Hans, 743
Valence bond theory, chemistry, 104
Validation
  induction, problem of, 381
  Nagel on, 493
  protocol sentences, 612
Validity, analyticity, 12
Vallentyne, Peter, 440
Value, function, 320–321
Value-free science, natural selection, 500
Value-free social science, economics, ethics and, 228
Value independence, decision theory, 184
Value-laden social sciences, scientific style, 765–766
Value-neutral science, unity of science movement, 849
Values
  Putnam and fact/value dichotomy, 620
  scientific style and, 765
  social sciences, 783, 784
van Brackel, J., 103
van der Walls, J.D., 417
van Essen, D.C., 516
van Fraassen, Bas
  Bayesianism, 53–54, 55
  causality, 93
  Dutch Book argument, 213
  empiricism, 240–241, 242, 243
  explanation, 282–283
  inductive logic, 390
  instrumentalism, 403
  laws of nature, 440, 441, 442, 443
  locality, quantum mechanics, 454
  observation, 528
  parsimony, 537
  physical sciences, philosophy of, 554
  prediction, 588
  probability, 601
  quantum mechanics, 657
  realism, 688, 689, 692, 693
  scientific models, 748
  theory structure, 827
    semantic view, 824, 825
    structuralist view, 826
    syntactic (received) view, 824
  underdetermination of theories, 840
van Gelder, Timothy, 154, 616
van Leeuwen, Cees, 155
van Maanen, Adriaan, 36
van Sweiten (*immunitas*), 364
van Vleck determinant, 635, 637
Varela, Francisco, 367, 368
Variability
  ecological communities, 218
  evolutionary psychology, monomorphic mind thesis, 266
  heritability, historical development of ideas, 353
  natural selection, 500–501
Variables
  causality, ontological categories, 90
  probability theory, 600
Variance, genetic
  fitness, 313
  heritability, 354
Variation, axiomatic method, 357
Variation, maintenance of, population genetics, 579–580
Variational calculus, 119, 121
Variational principles, classical mechanics, 119–121
Variolization, 364
Varzi, A.C., 287
Vásquez-Yanes, C., 163
Vector/axial-vector (V-A) model of weak interactions, 541
Vector fields, classical mechanics, 121
Vector state, Bell's inequality, 456
Vehicles versus replicators, 261
Velocity
  kinetic theory, 416
  of light, 789, 790–791, 792
Velocity, of light, 789, 790–791, 792
Veltman, M., 542
Venn, John, 46, 601, 607
Venn diagrams, 867
Veracity, visual representation, 868
Verbal Behavior (Skinner), 114
Veridical experiences, perceptual experiences as, 545
Verifiability, 851–854
  Ayer and, 39
  Carnap and, 83
  cognitive significance, 132–135
  demarcation problem, 191
  disposition, laws, reduction sentences, 853–854
  Quine and, 659
  Vienna Circle, 860
Verifiability theory of meaning, 62
Verification
  explicanda/explicata pairs, 288
  logical empiricism versus logical positivism, 458
  physicalism, 558
  Popper and, 573, 574
Verificationism
  analyticity, 13
  Ayer, 38–40
  Carnap, 83
  cognitive significance, 133
  demarcation problem, 191
  induction, problem of, 382
  Kuhn's scientific revolutions, 424–425
  linguistics, American Structuralism, 107
  perception, *see* Perception

Verificationism (*Continued*)
  Popper and, 573, 574
  prediction, 587
  Quine and, 659, 661–662
  Reichenbach and, 707–708
  Schlick and, 728
  scientific domains, 733
  Vienna Circle and, 859
Verisimilitude, 854–857
  application, 857
  epistemological and methodological approaches, 857
  history, 854–855
  Popper, 855–856
  Popper and, 575–577
  scientific progress, 751, 752
  similarity approaches, 856
Vermaas, Pieter, 657
Verner, J., 166
Versuche über Planzenhybriden (Mendel), 330
Vertical inferences, 2
Verver, S., 155
Viability analysis, conservation biology, 166–167
Vicious circle principle (Ramsey), 672
Victor, J.L., 486, 488
Vienna Circle
  analyticity, 13
  Ayer, 38–40
  Carnap, 79, 80, 82–83, 86
  classical mechanics, 117, 121
  conventionalism, 169, 175
  demarcation problem, 191
  Duhem thesis, 209
  empiricism, 237–238
  feminist philosophy, 302
  Hahn, 341–343
  Hempel and Berlin circle and, 347
  history, members, philosophy, 858–863
    logical positivism and/or logical empiricism, 858–859
    recent reassessments, 861–862
    scientific conception of the world and scientific humanism, 860–861
    scientific philosophy and philosophy of science, 859–860
  kinetic theory, 417
  logical empiricism, 459–461, 464
  Mach, 467–468
  Neurath, 510–514
  Popper and, 572
  protocol sentences, 610, 611, 612
  Quine and, 659
  Reichenbach and, 707, 709
  Schlick and, 726, 727
  unity and disunity of science, 844
  unity of science movement, 848
  verifiability, 851–852
  von Neumann, 505, 506
Viennese positivism, *see* Logical positivism
Vigier, J., 600
Vihvelin, Kadhri, 99
Vinci, Thomas, 591
Vindication, induction, 381
Vin Staudt, K.G.C., 357
Virtual particles, quantum field theory, 629
Virtual work, Lagrange's principle of, 120
Vision
  evolutionary psychology, 265–266
  perception, 550
    Gibson's attack on computationalism, 549–550
    Marr's computer model of mind, 548
    visual agnosia, 546
  and thought, 272
Visionary ideas, von Neumann, 503
Visual cortex, 617
Visual education, Neurath, 511
Visual field
  blind spot, 548
  emergentism, 232
Visualization
  Reichenbach and, 706, 707
  Vienna Method, 848
Visual representation, 863–870
  evidential roles, 867–869
  necessity, 869–870
  scientific reasoning and, 867
  visual symbols, 864–865
Vital forces, 90
Vocabularies, reductionism, 696–697
Volition, scientific style, 765
Volitional bifurcation, 707
von Aster, E., 704, 709
von Baer, Karl Ernest, 353
von Bekesy, G., 698
von Bertalanffy, Ludwig, 232
von Ekhardt, B., 123–124, 127, 129, 130, 131
von Helmholtz, Hermann
  Carnap, 80
  classical mechanics, 115, 119
  Reichenbach and, 706
von Kries, J., 603
von Liebig, J., 273
von Mises, Ludwig, 511
von Mises, Richard
  classical mechanics, 119
  Hahn and, 341
  probability, 601, 606, 607
  Vienna Circle, 858
von Neumann, John, 503–510
  altruism, 10
  Bayesianism, 41
  game theory, 323, 507–508
  Hilbert and, 360
  life and work
    computers, automata, and the brain, 508–509
    education and early career, 504
    game theory and expanding economy model, 507–508
    quantum physics, 505–507
    set theory and foundational program, 504–505
  locality, Bell's theorem, 453
  probability, 605
  quantum logic, 633, 638
  quantum measurement problem, 647
  quantum mechanics, 653, 654, 655–656
  Turing and, 833, 835
  Vienna Circle, 858
von Neumann, Nicholas, 504
von Neumann algebras, type II, 506
von Neumann-Dirac collapse formulation of quantum mechanics, 595
von Neumann machines, 127, 508

von Neumann-Morgenstern utility function, 323
von Weiszsäcker, Carl
  observation, 525, 526
  quantum logic, 633
Voo-doo science, 195
Vortex theory, Kelvin, 417
Vrba, Elizabeth, 4

W

Waddington, C.H.
  evolutionary psychology, 266
  innate/acquired distinction, 399
  molecular biology, 481
Wade, M., 582
Wager model, *see* Betting/wager model
Wagner, Günther P.
  developmental biology, philosophy of, 72
  evolutionary psychology, 266
  heritability, 355
Waismann, Friedrich, 82, 83
  probability theory, 603
  quantum logic, 641
  Quine and, 659
  Schlick and, 726, 728
  Vienna Circle, 858, 859
Wald, Abraham
  Bayesianism, 41, 44
  Neyman-Pearson (NP) tests, 806
  Neyman-Pearson (NP) theory extension, 44
Wald, Robert M.
  irreversibility, 413
  physical sciences, philosophy of, 555
Wallace, Alfred Russel, 252–253, 500
Walsh, Denis, 6, 318
Walter, Maila L., 75
Warrant, limits of, 239
Warsaw Group, Vienna Circle, 862
Wartofsky, Marx, 849
Wassermann, L., 535
Waters, Kenneth, 484
  biological information, 67–68
  biology, philosophy of, 71, 72
  reductionism, 699
Waterston, J.J., 416
Watkins, John, 479
Watkins, J.W.N., 226
Watson, G., 166
Watson, James D., 254, 366
  biological information, 65
  experimental method, 271
  genetics, 332
  molecular biology, 480–481
  realism, 689
  theory structure, models and, 827
Watson, John B., 61–62
  behaviorism, 63
Watson-Crick model of DNA, 333, 335, 336, 337
Watts-Tobin, R.J., 336
Wave aspect, quantum field theory, 631
Waveforms, hydrogenic, 104
Wave function
  complementarity, 141
  prediction, 595
  quantum mechanics, 653
  Reichenbach and, 710–711
Wave mechanics, 652–653
Wave motion, and locality/local action, 452
Wave optics, scientific revolutions, 755
Waves, complementarity, 142
Wave theory
  of Fresnel, 45
  underdetermination of theories, 841
Wayne, Andrew, 543, 632
Weak anthropic principle (WAP), 22, 23
Weak artificial intelligence, 30, 31–32
Weak axiom systems, 358
Weak constraint, quantum theory, 646
Weak equivalence, artificial intelligence, 30–31, 32
Weak forces, unified description, 542
Weak global supervenience, 820
Weak interactions, 542
  effective field theory, 543–544
  models of, 541
  observation, 527
Weak interactions rule, molecular biology, 483
Weak law of large numbers (Bernoulli), 46, 47
Weak nuclear force, 543
Weasel, Lisa H., 296, 302
Weatherbee, Scott D., 254
Weaver, Warren, 480
Weber, Max, 478, 511, 784
Weber, T., 655
Web of belief, Quine
  analyticity, 19–20
  Duhem thesis, 209
  logical empiricism, 462
  Poincaré and, 570
  protocol sentences, 612–613
Weibel, Peter, 862
Weierstrass, Karl, 119
Weighted connections, 151, 152, 153
Weinberg, S.
  particle physics, 542, 543
  unity and disunity of science, 846
Weinberg, William, 333–334
Weinert, F., 827
Weininger, S., 103, 105
Weinreich, Max, 110, 444
Weiskrantz, L., 618
Weiss, Paul, 128, 232, 233
Weldon, W.F.R., 578
Well-formed sentence, 109
Werner, Reinhard, 456, 457
Werner state, 456
Westman, Robert, 754
Weyl, H.
  Hilbert and, 356, 359
  Ramsey and, 674
  von Neumann and, 504
Weyl momentum operators, 506
What is Infinity (Hahn), 343
Wheeler, Q.D., 799
Wheeler, W.M., 231
Whewell, William, 45
  demarcation problem, 190
  prediction, 589
Whig fallacies, scientific revolutions, 756
White, Roger, 22

- Whitehead, Alfred North, 715, 716
  - Carnap, 80
  - Carnap, Rudolf, 82
  - logical positivism, 460
  - phenomenalism, 552
  - Quine and, 659
  - Ramsey and, 672–674
  - rational reconstruction, 682
  - Russell and, 718
  - verifiability, 852
  - Vienna Circle and, 860
- Whitlock, M., 582
- Why questions, explanation, 283
- Wien, Max, 80
- Wiener, N., 848
- Wien's formula, 650, 651
- Wiggins, D., 377
- Wigner, Eugene
  - quantum measurement problem, 647
  - quantum mechanics, 654
- Wilce, Alexander, 640, 643
- Wiley, E.O., 798, 800
- Wilkerson, Timothy E., 71
- William of Occam, 531
- Williams, George C.
  - altruism, 8–9
  - biological information, 65
- Williams, K.L., 486
- Williams, Mary B.
  - fitness, 314
  - prediction, 596
- Williams, S.M., 582
- Wilson, Curtis, 118
- Wilson, David Sloan, 8
  - altruism, 9
  - evolutionary biology, philosophy of, 70
  - natural selection, 498
- Wilson, E.O., 263
  - conservation biology, 164
  - unity and disunity of science, 846
- Wilson, Jack, 364, 375, 378
- Wilson, Jessica, 91
- Wilson, Mark, 118
- Wilson, M.R., 796
- Wilson, Robert A.
  - biology, philosophy of, 71
  - psychology, philosophy of, 614, 615
  - species, 800
- Wilson cloud chambers, 538
- Wimsatt, William
  - ecology, 221, 223
  - emergence, 232
  - explanation, 285
  - heritability, 355
  - innate/acquired distinction, 399
  - mechanism, 472
  - molecular biology, 484
  - molecular biology, philosophy of, 71
  - population genetics, 584
  - reductionism, 697, 699, 700, 701, 702
  - scientific models, 747
  - visual representation, 868
- Winch, Peter, 782
- Windellband, W., 765, 766, 844
- Winnie, John, 15, 19
- Wisan, W.L., 766
- *Wissenschaftkliche Weltauffassung: Der Wiener Kreis*, 459, 727, 858
- *Wissenschaftslehre* (Bozano), 12
- *Wissenschaftslogik*, 859
- Witmor, D. Gene, 567
- Wittgenstein, Ludwig, 716
  - Ayer and, 38, 39
  - Carnap and, 83
  - conventionalism, 169, 175
  - demarcation problem, 189, 191
  - Feyerabend and, 304, 305
  - Hahn and, 342, 343
  - Hanson and, 344
  - linguistics, philosophy of, 448
  - logical positivism, 460–461
  - Nagel on, 493
  - probability theory, 603
  - Putnam and, 620
  - Ramsey and, 671, 680
  - Schlick and, 726–727
  - social sciences, 783
  - Turing and, 835
  - unity and disunity of science, 844, 846
  - Vienna Circle, 858, 859, 861
- Wolff, Caspar Friederich, 353
- Wollaston, William Hyde, 35
- Wöllflin, H., 766
- Wolpert, L., 808
- Wood, R., 531
- Woodward, James
  - causality, 100
  - explanation, 278
  - laws of nature, 442
  - neurobiology, 514
- Woody, A., 25, 104
- Wooley, R.G., 234
- Woolgar, Steve, 269, 776–777
- World, metaphysical realism, 622, 623
- World changes, Gestalts and, 427–428
- World conception, scientific, *see* Scientific world conception
- Worldline, 786, 790
- World stages (Popper), 258–259, 575–577
- World-structure, 786
- Worldview
  - commonsense
    - phenomenalism, 552, 553
    - Quine, 660
  - mechanical, Hilbert and, 359
  - paradigms, 752
  - scientific, 505, 765; *see also* Scientific world conception
  - scientific revolutions and, 760, 762
  - scientific style and, 765–767
  - unity and disunity of science, 844
- Wormholes, 830
- Worrall, John
  - demarcation problem, 192
  - instrumentalism, 401
  - prediction, 590, 591
  - scientific revolutions, 760
  - statistics, 810

Wray, C.A., 486
Wright, Chauncey, 257
Wright, Crispin, 19, 626
Wright, Larry
  causality, 90
  function, 317
  social sciences, 782
Wright, Sewall
  causality, 94–95
  evolution, 253, 255–256
  population genetics, 578, 582–583
Wu, C.-S., 540
Wylie, Alison, 297, 299, 300, 301, 302

X

Xia, Zhihong, 122
X-ray crystallography, 480
X-ray telescopes, 36–37

Y

Yarzi, Achille, 869
Yockey, H.P., 484
Young, J.M., 486, 488
Young, Thomas, 755
Yukawa, Hideki, 538, 539

Z

Zabell, S.L.
  Bayesianism, 47, 52, 56, 59
  Ramsey, 675
Zabell's rule, 52
Zahar, E.G.
  demarcation problem, 192
  prediction, 589, 591
  research programs, 715
Zanghi, Nino, 656
Zeeman effect, 652
Zeitgeist, scientific style, 765
Zeitlin, Jonathan, 229
Zellner, Arnold, 59
Zeno's paradoxes, 292
Zermelo, Ernest
  game theory, 323
  Hahn and, 342
  Russell and, 718
Zermelo-Fraenkel (ZF) set theory, 504
Zero-dimensional linear subspace, quantum logic, 641
Zero-sum games, 507
Ziehen, Theodor, 682
Zihlman, A., 298
Zilsel, Edgar, 610, 858, 860
Zuckerandl, Emile, 254, 582
Zwart, Sjoerd, 855, 856
Zytkow, J.M., 731