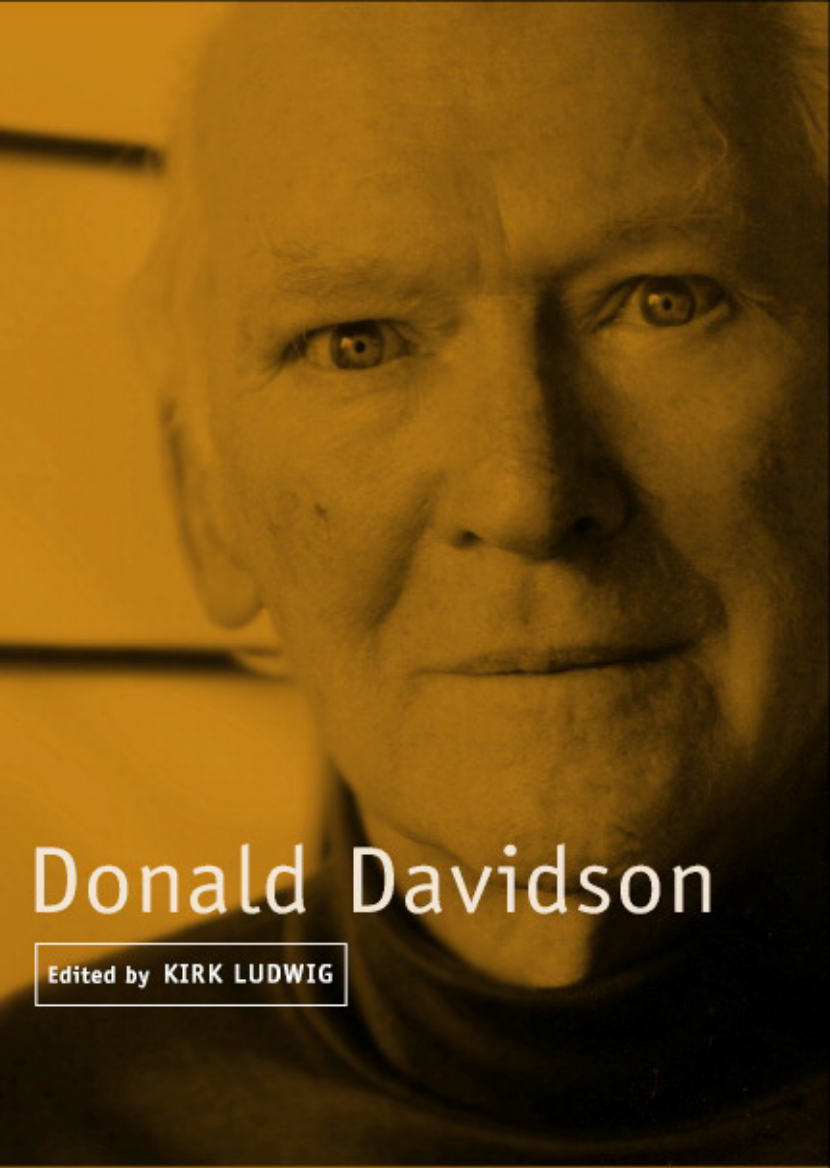

Donald Davidson

Edited by KIRK LUDWIG

CONTEMPORARY PHILOSOPHY IN FOCUS

CAMBRIDGE

more information - www.cambridge.org/9780521790437

This page intentionally left blank

Donald Davidson

Donald Davidson has been one of the most influential figures in modern analytic philosophy. He has made seminal contributions to a wide range of subjects: philosophy of language, philosophy of action, philosophy of mind, epistemology, metaphysics, and the theory of rationality. His principal work, embodied in a series of landmark essays stretching over nearly forty years, exhibits a unity rare among philosophers contributing to so many different topics. These essays – elegant, compact, sometimes cryptic, and difficult – together form a mosaic that presents a systematic account of the nature of human thought, action and speech, and their relation to the natural world, which is one of the most subtle and impressive systems to emerge in analytic philosophy in the last fifty years.

Written by a distinguished roster of philosophers, this volume includes chapters on truth and meaning; the philosophy of action; radical interpretation; philosophical psychology; the semantics and metaphysics of events; knowledge of the external world, other minds, and our own minds; and the implications of Davidson's work for literary theory.

This is the only comprehensive introduction to the full range of Davidson's work, and, as such, it will be of particular value to advanced undergraduates, graduates, and professionals in philosophy, psychology, linguistics, and literary theory.

Kirk Ludwig is Associate Professor of Philosophy at the University of Florida.

Contemporary Philosophy in Focus

Contemporary Philosophy in Focus offers a series of introductory volumes to many of the dominant philosophical thinkers of the current age. Each volume consists of newly commissioned essays that cover all the major contributions of a preeminent philosopher in a systematic and accessible manner. Comparable in scope and rationale to the highly successful series **Cambridge Companions to Philosophy**, the volumes do not presuppose that readers are already intimately familiar with the details of each philosopher's work. They thus combine exposition and critical analysis in a manner that will appeal both to students of philosophy and to professionals and students across the humanities and social sciences.

PUBLISHED VOLUMES:

Stanley Cavell edited by Richard Eldridge
Daniel Dennett edited by Andrew Brook and Don Ross
Thomas Kuhn edited by Tom Nickles
Alasdair MacIntyre edited by Mark C. Murphy
Robert Nozick edited by David Schmidtz

FORTHCOMING VOLUMES:

Paul Churchland edited by Brian Keeley
Ronald Dworkin edited by Arthur Ripstein
Jerry Fodor edited by Tim Crane
David Lewis edited by Theodore Sider and Dean Zimmermann
Hilary Putnam edited by Yemima Ben-Menahem
Richard Rorty edited by Charles Guignon and David Hiley
John Searle edited by Barry Smith
Charles Taylor edited by Ruth Abbey
Bernard Williams edited by Alan Thomas

Donald Davidson

Edited by

KIRK LUDWIG
University of Florida

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, United Kingdom

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521790437

First published in print format 2003

ISBN-13 978-0-511-06712-9 eBook (NetLibrary)
ISBN-10 0-511-06712-7 eBook (NetLibrary)

ISBN-13 978-0-521-79043-7 hardback
ISBN-10 0-521-79043-3 hardback

ISBN-13 978-0-521-79382-7 paperback
ISBN-10 0-521-79382-3 paperback

For Shih-Ping Lin

Contents

Contributors

JAEGWON KIM is William Herbert Perry Faunce Professor of Philosophy at Brown University. He is the author of *Supervenience and Mind: Selected Philosophical Essays* (Cambridge University Press, 1993), *Philosophy of Mind* (1996), *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (1998), *Supervenience* (2001), and of numerous articles in the philosophy of mind and metaphysics. He is coeditor of *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brant*, with Alvin Goldman (1978); *Emergence or Reduction? Essays on the Prospects of Non-reductive Physicalism*, with A. Beckerman and H. Flohr (1992); and, with Ernest Sosa, of *A Companion to Metaphysics* (1995), *Metaphysics: An Anthology* (1999), and *Epistemology: An Anthology* (2000).

ERNEST LEPORE is Professor of Philosophy at Rutgers University and director of the Rutgers University Center for Cognitive Science. He is the author of *Meaning and Argument: An Introduction to Logic through Language* (2000); coauthor, with Jerry Fodor, of *Holism: A Shopper's Guide* (1992); and the author of numerous articles in the philosophy of language, philosophical logic, philosophy of mind, and metaphysics. He is the editor of *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (1986) and *New Directions in Semantics* (1987). He is coeditor of *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, with B. McLaughlin (1985); *John Searle and His Critics*, with Robert Van Gulick (1991); *Holism: A Consumer Update*, with Jerry Fodor (1993); and *What Is Cognitive Science?*, with Zenon Pylyshyn (1999).

KIRK LUDWIG is Associate Professor of Philosophy at the University of Florida. He is the author of numerous articles in the philosophy of language, philosophical logic, philosophy of mind, and epistemology. His recent publications include "The Semantics and Pragmatics of Complex Demonstratives," with Ernest Lepore, in *Mind* (2000); "What Is the Role of a Truth Theory in a Meaning Theory?," in *Truth and Meaning: Topics*

in Contemporary Philosophy (2001); "What Is Logical Form?," with Ernest Lepore, in *Logical Form and Language* (2002); "Outline of a Truth Conditional Semantics for Tense," with Ernest Lepore, in *Tense, Time and Reference* (2002); "The Mind-Body Problem: An Overview," in *The Blackwell Guide to the Philosophy of Mind* (2002); and "Vagueness and the Sorites Paradox," with Greg Ray, *Philosophical Perspectives* (2002). He is completing a book with Ernest Lepore titled *Donald Davidson: Meaning, Truth, Language and Reality* (forthcoming).

ALFRED R. MELE is the William H. and Lucyle T. Werkmeister Professor of Philosophy at Florida State University. He is the author of *Irrationality: An Essay on Akrasia, Self-deception, and Self-control* (1987), *Springs of Action: Understanding Intentional Behavior* (1992), *Autonomous Agents: From Self-control to Autonomy* (1995), *Self-Deception Unmasked* (2001), *Motivation and Agency* (forthcoming), and of numerous articles in the philosophy of action and philosophy of mind. He is the editor of *The Philosophy of Action* (1997) and coeditor of *Mental Causation*, with John Heil (1993), and of *Rationality*, with Piers Rawling (forthcoming).

PAUL PIETROSKI is Associate Professor of Philosophy and Linguistics at the University of Maryland. He is the author of *Causing Actions* (2000), *Events and Semantic Architecture* (forthcoming), and of numerous articles in the philosophy of language, semantics, and philosophy of mind. His recent articles include "On Explaining That," *Journal of Philosophy* (2000); "Nature, Nurture, and Universal Grammar," with Stephen Crain, *Linguistics and Philosophy* (2001); "Function and Concatenation," in *Logical Form and Language* (2002); and "Small Verbs, Complex Events: Analyticity without Synonymy," in *Chomsky and His Critics* (2002).

PIERS RAWLING is Associate Professor of Philosophy at Florida State University. He is the author of numerous articles in areas ranging from ethics and the philosophy of language to game theory, decision theory, and quantum computing. His recent publications include "Orthologic and Quantum Logic: Models and Computational Elements," with Stephen Selesnick, *Journal of the Association of Computing Machinery* (2000); "The Exchange Paradox, Finite Additivity, and the Principle of Dominance," in *Logic, Probability and the Sciences* (2000); "Achievement, Welfare and Consequentialism," with David McNaughton, *Analysis* (2001); "Davidson's Measurement-Theoretic Reduction of Mind," in *Interpreting Davidson* (2001); "Deontology," with David McNaughton, *Ethics* (2002); "Conditional Obligations," with David McNaughton, *Utilitas* (forthcoming); and

"Decision Theory and Degree of Belief," in *Companion to the Philosophy of Science* (forthcoming). He is coeditor, with Alfred Mele, of *Rationality* (forthcoming).

ERNEST SOSA is Romeo Elton Professor of Natural Theology and Professor of Philosophy at Brown University. He is the author of numerous articles in epistemology and metaphysics. His recent publications include "Reflective Knowledge in the Best Circles," *Journal of Philosophy* (1997); "Man the Rational Animal?," with David Galloway, *Synthese* (2000); "Human Knowledge, Animal and Reflective," *Philosophical Studies* (2001); and "Thomas Reid," with James Van Cleve, in *The Blackwell Guide to the Modern Philosophers from Descartes to Nietzsche* (2001). He is editor or coeditor of fifteen books, including *A Companion to Epistemology*, with Jonathan Dancy (1992); *Causation*, with Michael Tooley (1993); *A Companion to Metaphysics*, with Jaegwon Kim (1995); *Cognition, Agency, and Rationality*, with Kepa Korta and Xabier Arrazola (1999); *The Blackwell Guide to Epistemology*, with John Greco (1999); *Metaphysics: An Anthology*, with Jaegwon Kim (1999); *Epistemology: An Anthology*, with Jaegwon Kim (2000); and *Skepticism*, with Enrique Villanueva (2000).

SAMUEL C. WHEELER is Professor of Philosophy at the University of Connecticut. He is the author of *Deconstruction as Analytic Philosophy* (2000) and of articles in a wide range of areas of philosophy, from ancient philosophy, literary theory, philosophical logic, and the philosophy of language to metaphysics and ethics. His publications include "Metaphor According to Davidson and De Man," in *Redrawing the Lines* (1989); "True Figures: Metaphor, Social Relations, and the Sorites," in *The Interpretive Turn* (1991); "Plato's Enlightenment: The Good as the Sun," *History of Philosophy Quarterly* (1997); "Derrida's Differance and Plato's Different," *Philosophy and Phenomenological Research* (1999); "Arms as Insurance," *Public Affairs Quarterly* (1999); and "Gun Violence and Fundamental Rights," *Criminal Justice Ethics* (2001).

Introduction

KIRK LUDWIG

Donald Davidson has been one of the most influential philosophers working in the analytic tradition in the last half of the twentieth century. He has made seminal contributions to a wide range of subjects: the philosophy of language and the theory of meaning, the philosophy of action, the philosophy of mind, epistemology, metaphysics, and the theory of rationality. His principal work, spread out in a series of articles stretching over nearly forty years, exhibits a unity rare among philosophers contributing to so many different topics. His essays are elegant, but they are also noted for their compact, sometimes cryptic style, and for their difficulty. Themes and arguments in different essays overlap, and later papers often presuppose familiarity with earlier work. Together, they form a mosaic that presents a systematic account of the nature of human thought, action, and speech, and their relation to the natural world, that is one of the most subtle and impressive systems to emerge in analytic philosophy in the last fifty years.

The unity of Davidson's work lies in the central role that reflection on how we are able to interpret the speech of another plays in understanding the nature of meaning, the propositional attitudes (beliefs, desires, intentions, and so on), and our epistemic position with respect to our own minds, the minds of others, and the world around us. Davidson adopts as methodologically basic the standpoint of the interpreter of the speech of another whose evidence does not, at the outset, presuppose anything about what the speaker's words mean or any detailed knowledge of his propositional attitudes. This is the position of the radical interpreter. The adoption of this position as methodologically basic rests on the following principle:

> The semantic features of language are public features. What no one can, in the nature of the case, figure out from the totality of the relevant evidence cannot be part of meaning. (Davidson 1984a [1979], p. 235)

The point carries over to the propositional attitudes, whose attributions to speakers are inseparable from the project of interpreting their words.

Virtually all of Davidson's major contributions are either components of this project of understanding how we are able to interpret others, or flow from his account of this. Davidson's work in the philosophy of action helps to provide part of the background for the interpreter's project: for an understanding of the nature of agency and rationality is also central to understanding the nature of speech. Davidson's work on the structure of compositional meaning theories plays a central role in understanding how we can interpret others as speakers; it also contributes to an understanding of the nature of agency through applications to the logical form of action sentences and connected investigations into the nature of events. The analysis of the nature of meaning and the attitudes through consideration of radical interpretation leads in turn to many of Davidson's celebrated theses in the philosophy of language, mind, and knowledge.

This introduction briefly surveys Davidson's life and philosophical development (§§1–2), and then provides an overview of major themes in, and traces out connections between, his work on the theory of meaning (§3), the philosophy of action (§4), radical interpretation (§5), philosophical psychology (§6), epistemology (§7), the metaphysics of events (§8), the concept of truth (§9), rationality and irrationality (§10), and the theory of literature (§11). The final section provides a brief overview of the volume.

1. EARLY LIFE AND INTELLECTUAL DEVELOPMENT

Donald Davidson was born on March 6, 1917, in Springfield, Massachusetts. After early travels that included three years in the Philippines, the Davidson family settled on Staten Island in 1924. From 1926, he attended the Staten Island Academy, and then began studies at Harvard in 1935, on a scholarship from the Harvard Club of New York. During his sophomore year, Davidson attended the last seminar given by Alfred North Whitehead, on material from *Process and Reality* (Whitehead 1929). Of his term paper for the seminar, Davidson has written, "I have never, I'm happy to say, received a paper like it" (Davidson 1999a, p. 14; henceforth parenthetical page numbers refer to this essay). He received an 'A+'. Partly inspired by this experience, as an undergraduate Davidson thought that in philosophy "[t]ruth, or even serious argument, was irrelevant" (p. 14).

For his first two years at Harvard, he was an English major, but he then turned to classics and comparative literature. His undergraduate education in philosophy, aside from his contact with Whitehead, came through a tutor

in philosophy, David Prall, and from preparing for four comprehensive exams – in ethics, history of philosophy, logic, and metaphysics. His main interests in philosophy at the time were in its history and its relation to the history of ideas.

He graduated in 1939. That summer he was offered a fellowship at Harvard in classical philosophy. He took his first course in logic with W. V. Quine, on material from *Mathematical Logic* (Quine 1940), which was published that term. Davidson's fellow graduate students at Harvard included Roderick Chisholm, Roderick Furth, Arthur Smullyan, and Henry Aiken.

Quine changed Davidson's attitude toward philosophy. Davidson reports that he met Quine on the steps of Eliot Hall after interviewing as a candidate to become a junior fellow. When Quine asked him how it had gone, Davidson "blurted out" his views on the relativity of truth to a conceptual scheme. Quine asked him (presciently borrowing an example of Tarski's) whether he thought that 'Snow is white' is true iff snow is white. Davidson writes: "I saw the point" (p. 22). In his first year as a graduate student, he took a seminar of Quine's on logical positivism: "What mattered to me," Davidson reported, "was not so much Quine's conclusions (I assumed he was right) as the realization that it was possible to be serious about getting things right in philosophy" (p. 23).[1]

With the advent of the Second World War, Davidson joined the navy, serving as an instructor on airplane spotting. Discharged in 1945, he returned to Harvard in 1946, and the following year took up a teaching position at Queens College, New York. (Carl Hempel was a colleague, whom Davidson later rejoined at Princeton; Nicholas Rescher was a student in one of Davidson's logic courses during this period.) On a grant from the Rockefeller Foundation for the 1947–48 academic year, Davidson finished his dissertation on Plato's *Philebus* (published eventually in 1990 [Davidson 1990b]) in Southern California, receiving the Ph.D. from Harvard in 1949.

In January 1951, Davidson left Queens College to join the faculty at Stanford, where he taught for sixteen years before leaving for Princeton in 1967. Davidson taught a wide range of courses at Stanford, reflecting his interests in nearly all areas of philosophy: logic, ethics, ancient and modern philosophy, epistemology, philosophy of science, philosophy of language, music theory, and ideas in literature, among others.

Through working with J. J. C. McKinsey and Patrick Suppes at Stanford, Davidson became interested in decision theory, the formal theory of choice behavior. He discovered a technique for identifying through choice behavior an agent's subjective utilities (the values agents assign to outcomes) and subjective probabilities (the degree of confidence they have

that an outcome will occur given an action), only to find later that Ramsey had anticipated him in 1926. This led to experimental testing of decision theory with Suppes, the results of which were published in *Decision Making: An Experimental Approach* (Davidson and Suppes 1957).

This early work in decision theory had an important influence on Davidson's later work in the philosophy of language, especially his work on radical interpretation. Davidson drew two lessons from it. The first was that in "putting formal conditions on simple concepts and their relations to one another, a powerful structure could be defined"; the second was that the formal theory itself "says nothing about the world," and that its content is given in its interpretation by the data to which it is applied (p. 32). This theme is sounded frequently in Davidson's essays.[2] The first suggests a strategy for illuminating a family of concepts too basic to admit of illuminating analyses individually. The second shows that the illumination is to be sought in the empirical application of such a structure.

At this time, Davidson also began serious work on semantics, prompted by the task of writing an essay on Carnap's method of extension and intension for the Library of Living Philosophers volume on Carnap (Davidson 1963), which had fallen to him after the death of McKinsey, with whom it was to have been a joint paper. Carnap's method of intension and extension followed Frege in assigning to predicates both intensions (meanings) and extensions (sets of things predicates are true of). In the course of work on the project, Davidson became seminally interested in the problem of the semantics of indirect discourse and belief sentences. Carnap, following Frege, treated the 'that'-clause in a sentence such as 'Galileo said that the Earth moves' as referring to an intension – roughly, the usual meaning of 'the Earth moves'. For in these "opaque" contexts, expressions cannot be intersubstituted freely merely on the basis of shared reference, extension, or truth value. Davidson became suspicious, however, of the idea that in opaque contexts expressions refer to their usual intensions, writing later that "[i]f we could recover our pre-Fregean semantic innocence, I think it would seem to us plainly incredible that the words 'The earth moves', uttered after the words 'Galileo said that', mean anything different, or refer to anything else, than is their wont when they come in other environments" (Davidson 1984 [1968], p. 108).

The work on Carnap led Davidson serendipitously to Alfred Tarski's work on truth. At Berkeley, Davidson presented a paper on his work on Carnap; the presentation was attended by Tarski. Afterward, Tarski gave him a reprint of "The Semantic Conception of Truth and Foundations of Semantics" (Tarski 1944). This led to Tarski's more technical "The Concept

of Truth in Formalized Languages" (Tarski 1983 [1932]). Tarski shows there how to provide a recursive definition of a truth predicate for a formal language that enables one to say for each sentence of the language, characterized in terms of how it is built up from its significant parts, under what conditions it falls in the extension of the truth predicate. Tarski's work struck Davidson as providing an answer to a question that had puzzled him, a question concerning accounts of the semantic form of indirect discourse and belief sentences: how does one tell when a proposed account is correct? The answer was that it was correct if it could be incorporated into a truth definition for the language in roughly the style outlined by Tarski. For this would tell one, in the context of a theory for the language as a whole, what contribution each expression in each sentence in the language makes to fixing its truth conditions. Moreover, such a theory makes clear how a finite being can encompass a capacity for understanding an infinity of nonsynonymous sentences. These insights were the genesis of two foundational papers in Davidson's work on natural language semantics, "Theories of Meaning and Learnable Languages" (Davidson 1984 [1966]) and "Truth and Meaning" (Davidson 1984 [1967]). In the former, Davidson proposed as a criterion for the adequacy of an analysis of the logical form of a sentence or complex expression in a natural language that it not make it impossible for a finite being to learn the language of which it was a part. In the latter, he proposed that a Tarski-style truth theory, modified for a natural language, could serve the purpose of a meaning theory for the language, without appeal to meanings, intensions, or the like.

Another important influence on Davidson during his years at Stanford was Michael Dummett, who lectured a number of times at Stanford during the 1950s on Frege and the philosophy of language.

During the 1958–59 academic year, Quine was a fellow at the Center for Advanced Study in the Behavioral Sciences at Stanford, where he put the finishing touches on the manuscript of *Word and Object* (Quine 1960). Davidson, who was on a fellowship from the American Council of Learned Societies that year, accepted Quine's invitation to read the manuscript. Quine's casting, in *Word and Object*, of the task of understanding linguistic communication in the form of an examination of the task of radical translation had a tremendous impact on Davidson. The radical translator must construct a translation manual for another's language solely on the basis of a speaker's dispositions to verbal behavior, without any antecedent knowledge of his thoughts or what his words mean. The central idea, that there can be no more to meaning than can be gleaned from observing a speaker's behavior, is a leitmotif of Davidson's philosophy of language. The

project of radical interpretation, which assumes a central role in Davidson's philosophy, is a direct descendant of the project of radical translation.[3] As we will see, Davidson brings together in this project the influence of both Tarski and Quine.

While at Stanford, Davidson also became interested in general issues in the philosophy of action, in part through his student Dan Bennett, who spent a year at Oxford and wrote a dissertation on action theory inspired by the discussions then going on at Oxford. The orthodoxy at the time was heavily influenced by Wittgenstein's *Philosophical Investigations* (Wittgenstein 1950). It held that explaining an action by citing an agent's reasons for it was a matter of *redescribing* the action in a way that placed it in a larger social, linguistic, economic, or evaluative pattern, and that, in particular, action explanation was *not* a species of causal explanation, which was taken to be, in A. I. Melden's words, "wholly irrelevant to the understanding" of human action (Melden 1961, p. 184). Davidson famously argued, against the orthodoxy, in "Actions, Reasons, and Causes" (Davidson 1980 [1963]), that action explanations *are* causal explanations, and so influentially as to establish this position as the new orthodoxy.

This interest in action theory connects in a straightforward way with Davidson's work on decision theory. Davidson's work on semantics and action theory came together in his account of the logical form of action sentences containing adverbial modification. Additionally, Davidson's work on action theory and decision theory, as noted earlier, provides part of the background and framework for his work on radical interpretation.

Davidson's first ten years at Stanford were a period of intense intellectual development, though accompanied by relatively few publications. During the 1960s, Davidson published a number of papers that changed the philosophical landscape and immediately established him as a major figure in analytic philosophy. Principal among these were "Actions, Reasons, and Causes" (Davidson 1980 [1963]), "Theories of Meaning and Learnable Languages" (Davidson 1984 [1966]), "Truth and Meaning" (Davidson 1984 [1967]), "The Logical Form of Action Sentences" (Davidson 1980b [1967]), "Causal Relations" (Davidson 1980a [1967]), "On Saying That" (Davidson 1984 [1968]), "True to the Facts" (Davidson 1984 [1969]), and "The Individuation of Events" (Davidson 1980 [1969]). (Details of these contributions are discussed below.) In 1970, Davidson gave the prestigious John Locke Lectures at Oxford University on the topic, "The Structure of Truth."

Davidson taught at Princeton from 1967 to 1970, serving as chair of the Philosophy Department for the 1968–69 academic year. He was appointed professor at the Rockefeller University in New York in 1970; he

moved to the University of Chicago as a University Professor in 1976, when the philosophy unit at Rockefeller University was disbanded. In 1981, he moved to the Philosophy Department at the University of California at Berkeley.

2. WORK CIRCA 1970 TO THE PRESENT

Davidson's work during the late 1960s and 1970s developed in a number of different directions.

(1) *Philosophy of action.* In a series of papers, Davidson continued to defend, refine, and elaborate the view of actions as bodily movements and action explanations as causal explanations originally introduced in "Actions, Reasons, and Causes." These papers included "How Is Weakness of the Will Possible?" (Davidson 1980b [1970]), "Action and Reaction" (Davidson 1970), "Agency" (Davidson 1980a [1971]), "Freedom to Act" (Davidson 1980a [1973]), "Hempel on Explaining Action" (Davidson 1980a [1976]), and "Intending" (Davidson 1980 [1978]). The work on the semantics of action sentences led to additional work on the semantics of sentences containing noun phrases referring to events – specifically, "Causal Relations" (Davidson 1980a [1967]), "The Individuation of Events" (Davidson 1980 [1969]), "Events as Particulars" (Davidson 1980a [1970]), and "Eternal vs. Ephemeral Events" (Davidson 1980b [1971]).

(2) *Philosophical psychology.* The publication in 1970 of "Mental Events" (Davidson 1980c [1970]) was a seminal event in the philosophy of mind. In it, Davidson proposed a novel form of materialism called anomalous monism. Davidson advanced an argument for a token-token identity theory of mental and physical events – according to which every particular mental event is also a particular physical event – that relied crucially on a premise that denied even the nomic reducibility of mental to physical properties. This was followed by a number of other papers elaborating on this theme, including "Psychology as Philosophy" (Davidson 1980 [1974]), "The Material Mind" (Davidson 1980b [1973]), and "Hempel on Explaining Action" (Davidson 1980a [1976]). Another paper from this period on the philosophy of psychology is "Hume's Cognitive Theory of Pride" (Davidson 1980b [1976]), which interprets Hume's theory of pride in the light of Davidson's causal theory of action explanation.

(3) *Natural language semantics.* Davidson elaborated and defended his proposal for using a Tarski-style truth theory to pursue natural language semantics in "In Defense of Convention *T*" (Davidson 1984a [1973]) and

extended a key idea (*parataxis*; see Chapter 1, §7, for a brief overview) of the treatment of indirect discourse introduced in "On Saying That" (Davidson 1984 [1968]) to quotation and to sentential moods (the indicative, imperative, and interrogative moods) in "Quotation" (Davidson 1984c [1979]) and "Moods and Performances" (Davidson 1984b [1979]), respectively. In addition, he edited, with Gilbert Harman, two important collections of essays on natural language semantics: *Semantics of Natural Language* (Davidson and Harman 1977) and *The Logic of Grammar* (Davidson and Harman 1975).

(4) *Radical interpretation*. Among the most important developments in Davidson's work in the philosophy of language during the 1970s was his elaboration of the project of radical interpretation, already adumbrated in "Truth and Meaning" (Davidson 1984 [1967]). Radical interpretation can be seen as an application of the insight – prompted by Davidson's work in decision theory during the 1950s – that a family of concepts whose members resist reduction to other terms one by one can be illuminated by examining the empirical application of the formal structure that they induce. The relation of the project of radical interpretation to understanding linguistic communication and meaning is taken up in "Belief and the Basis of Meaning" (Davidson 1984a [1974]) and, in the context of a defense of the claim that thought is not possible without a language, in "Thought and Talk" (Davidson 1984 [1975]). "Reply to Foster" (Davidson 1984 [1976]) contains important clarifications of the project and its relation to using a truth theory as a theory of interpretation; it responds to a critical paper by John Foster (Foster 1976), which appeared in an important collection of papers edited by Gareth Evans and John McDowell (Evans and McDowell 1976). "Reality without Reference" (Davidson 1984b [1977]) and "The Inscrutability of Reference" (Davidson 1984a [1979]) are applications of reflections on radical interpretation to the status of talk about the reference of singular terms and the extensions of predicates in a language. Davidson draws the startling conclusion (first drawn by Quine [1969]) that there are many different reference schemes that an interpreter can use that capture equally well the facts of the matter concerning what speakers mean by their words.

(5) *Epistemology*. "On the Very Idea of a Conceptual Scheme" (Davidson 1984b [1974]) originated in the last of Davidson's six John Locke Lectures in 1970 and was delivered in the published form as his presidential address to the Eastern Division meeting of the American Philosophical Association in 1973. An influential paper, it argues against the relativity of truth to a conceptual scheme and against the possibility of there being radically different

conceptual schemes. "The Method of Truth in Metaphysics" (Davidson 1984a [1977]) is concerned with the relation between semantic theory and the nature of reality. In it, Davidson argues for two connected theses about the relation between our thought and reality. The first is that the ontological commitments of what we say are best revealed in a theory of truth for the languages we speak. The second is that massive error about the world, including massive error in our empirical beliefs, is impossible. The second thesis rests in part on conclusions reached in reflections on the project of radical interpretation, especially reflections about the need to employ in interpretation what is called the Principle of Charity, an aspect of which is the assumption that most of a speaker's beliefs about his environment are true.

(6) *Metaphor*. The last development in Davidson's work during the 1970s is an important and original account of the way in which metaphors function. In "What Metaphors Mean" (Davidson 1984 [1978]), Davidson argued that it is a mistake to think that metaphors function by virtue of having a special kind of meaning – metaphorical meaning; instead, they function in virtue of their literal meanings to get us to see things about the world. "Metaphor makes us see one thing as another by making some literal statement that inspires or prompts the insight" (Davidson 1984 [1978], p. 261).

Two collections of Davidson's papers appeared during the 1980s – *Essays on Actions and Events* (Davidson 1980a) and *Inquiries into Truth and Interpretation* (Davidson 1984b). These works collected many of his papers, respectively, on the philosophy of action and the metaphysics of events, and in the theory of meaning and philosophy of language. In 1984, an important conference on Davidson's work (dubbed "Convention D" by Sydney Morgenbesser), which brought together more than 500 participants, was organized at Rutgers University by Ernest Lepore, out of which came two collections of papers – *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Lepore and McLaughlin 1985) and the similarly subtitled *Truth and Interpretation* (Lepore 1986). A collection of essays on Davidson's work in the philosophy of action, with replies by Davidson, edited by Bruce Vermazen and Merrill Hintikka, *Essays on Davidson: Actions and Events* (Vermazen and Hintikka 1985), appeared in 1985.

Davidson's work during the 1980s can be divided into five main categories. (1) In the first category are those papers following up on issues in action theory – "Adverbs of Action" (Davidson 1985a) and "Problems in the Explanation of Action" (Davidson 1987b). (2) In the second are papers on the nature of rationality and irrationality – "Paradoxes of Irrationality" (Davidson 1982), "Rational Animals" (Davidson 1985 [1982]),

"Deception and Division" (Davidson 1985b), and "Incoherence and Irrationality" (Davidson 1985c). (3) The third category combines elements of work on the determination of thought content and epistemology. "Empirical Content" (Davidson 2001a [1982]), "A Coherence Theory of Truth and Knowledge" (Davidson 2001 [1983]), "Epistemology and Truth" (Davidson 2001a [1988]), "The Conditions of Thought" (Davidson 1989), "The Myth of the Subjective" (Davidson 2001b [1988]), and "What Is Present to the Mind?" (Davidson 2001 [1989]) are all concerned with the thesis that the contents of our thoughts are individuated in part by their usual causes in a way that guarantees that most of our empirical beliefs are true. "First Person Authority" (Davidson 2001 [1984]) and "Knowing One's Own Mind" (Davidson 2001 [1987]) are concerned to argue that knowledge of our own minds can be understood in a way that does not give primacy to the subjective, and that the relational individuation of thought content is no threat to our knowledge of our thoughts. (4) The fourth category of papers includes those that develop earlier work in the philosophy of language. "Toward a Unified Theory of Meaning and Action" (Davidson 1980b) explicitly combines decision theory with Davidson's earlier work on radical interpretation, and "A New Basis for Decision Theory" (Davidson 1985d) outlines a procedure for identifying logical constants by finding patterns among preferences toward the truth of sentences. In "Communication and Convention" (Davidson 1984 [1983]), Davidson takes up the question of what role convention plays in communication, and in particular the question of whether it is essential to communication at all. "Communication and Convention" already contains the main themes, if not so provocatively stated, of Davidson's later and more controversial "A Nice Derangement of Epitaphs," in which he argues that "there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed" (Davidson 1986c, p. 446). "James Joyce and Humpty Dumpty" (Davidson 1991b) is another excursion into literary theory. (5) The fifth category is work on issues in ethical theory from the standpoint of radical interpretation, the Lindley Lectures, *Expressing Evaluations* (Davidson 1984a), and "Judging Interpersonal Interests" (Davidson 1986b), a central thesis of which is that communication requires shared values as much as shared beliefs.

In 1989, Davidson gave the John Dewey Lectures at Columbia, "The Structure and Content of Truth" (Davidson 1990d), echoing the title of the John Locke Lectures delivered almost twenty years before. These provide a comprehensive overview and synthesis of Davidson's work in the theory of meaning and radical interpretation up through the end of the 1980s.

Additions to Davidson's corpus since 1990 mostly follow up themes already present in earlier work. These include a number of papers on interrelated themes in epistemology and thought content – "Epistemology Externalized" (Davidson 2001a [1991]), "Turing's Test" (Davidson 1990e), "Representation and Interpretation" (Davidson 1990c), "Three Varieties of Knowledge" (Davidson 2001b [1991]), "Subjective, Intersubjective, Objective" (Davidson 1996b), "The Second Person" (Davidson 2001 [1992]), "Seeing through Language" (Davidson 1997b), and "Externalisms" (Davidson 2001a). These papers overlap in content. One theme that emerges as new – or at least as newly salient – is a transcendental argument designed to show that it is only in the context of communication that one can have the concept of objective truth and have determinate thoughts about things in one's environment, because only in the context of communication does the concept of error have scope for application, and only in triangulating with another speaker on an object of common discourse can we secure an objectively determinate object of thought. In "Thinking Causes" (Davidson 1993b), Davidson defends his view that action explanations can be causal explanations, while what our beliefs are about is determined in part in terms of what things in the environment typically cause them. Davidson comments on Quine's work and its relation to his own in "Meaning, Truth and Evidence" (Davidson 1990a), "What Is Quine's View of Truth?" (Davidson 1994d), and "Pursuit of the Concept of Truth" (Davidson 1995e). "On Quine's Philosophy" (Davidson 1994a) is an informal comment on Quine's philosophy delivered after a talk by Quine. In "The Social Aspect of Language" (Davidson 1994c), a contribution to a volume on *The Philosophy of Michael Dummett*, Davidson continues a debate with Dummett about the role of conventions in linguistic understanding that had begun in "Communication and Convention" and continued in "A Nice Derangement of Epitaphs." "Locating Literary Language" (Davidson 1993a) is a contribution to a collection of papers entitled *Literary Theory after Davidson* (Dasenbrock 1993), which discusses the interpretation of literature in the light of Davidson's views about interpretation more generally. "Laws and Cause" (Davidson 1995b) offers a Kantian-style argument for an assumption employed, but not defended, in the argument for a token-token identity theory of mental and physical events in "Mental Events" – namely, the nomological character of causality, the principle that any two events related as cause and effect are subsumed by some strict law. Several papers defend a thesis about truth that has been a constant theme of Davidson's work, namely, that it is (a) irreducible to other, more basic concepts, and (b) a substantive concept, in the sense

that no deflationary conception of the concept of truth is correct. These include "The Folly of Trying to Define Truth" (Davidson 1996a), "The Centrality of Truth" (Davidson 1997a), and "Truth Rehabilitated" (Davidson 2000c). (These last two are slightly different versions of the same essay.) Other essays during this period include "Who Is Fooled?" (Davidson 1997c), which returns to the topic of self-deception; "Could There Be a Science of Rationality?" (Davidson 1995a), which discusses the import of the anomalousness of the mental for the prospects of a science of the mind; "Objectivity and Practical Reason" (Davidson 2000a) and "The Objectivity of Values" (Davidson 1995c), which return to the themes of *Expressing Evaluations* and "Judging Interpersonal Interests"; "The Problem of Objectivity" (Davidson 1995d), which reviews the arguments for the necessity of having the concept of truth in order to have thoughts, and for the need for interpersonal communication to have the concept of objective truth; "Interpretation: Hard in Theory, Easy in Practice" (Davidson 1999b) and "Perils and Pleasures of Interpretation" (Davidson 2000b), which are versions of the same paper and summarize Davidson's views on the nature of thought and its relation to interpretation; and two papers on historical figures, "The Socratic Conception of Truth" (Davidson 1992b) and "Spinoza's Causal Theory of the Affects" (Davidson 1999j).

In 2001, a new volume of essays appeared, *Subjective, Intersubjective, Objective* (Davidson 2001b), bringing together a number of papers from 1982 to 1998 on interrelated themes in philosophy of mind and epistemology. This is to be followed by two further volumes of collected papers: *Problems of Rationality*, collecting papers from 1974 to 1999 on values, on the relation of rationality to thought, and on irrationality; and *Truth, Language and History*, bringing together papers from 1986 to 2000 on truth, nonliteral language use and literature, and essays on issues and figures in the history of philosophy.

During this period, a number of collections of essays on Davidson's work have appeared: *Reflecting Davidson: Donald Davidson Responding to an International Forum of Philosophers* (Stoecker 1993); *Language, Mind, and Epistemology: On Donald Davidson's Philosophy* (Preyer 1994); *Literary Theory after Davidson* (Dasenbrock 1993), mentioned earlier; *The Philosophy of Donald Davidson*, the volume on Davidson in the Library of Living Philosophers series (Hahn 1999); and *Interpreting Davidson* (Kotatko, Pagin, and Segal 2001). Davidson replies to the essays in the first, fourth, and fifth of these.

3. THEORY OF MEANING AND NATURAL LANGUAGE SEMANTICS

Davidson's central contribution to natural language semantics, introduced in "Truth and Meaning" (Davidson 1984 [1967]), is the proposal to employ a truth theory, in the sense of a finite axiomatic theory characterizing a truth predicate for a language, in the style of Tarski, to do the work of a compositional meaning theory for a language. The insight that this relies upon is that an axiomatic truth definition that meets Tarski's Convention *T* enables one to read off from the canonical theorems of the theory what sentences of the language mean. Tarski's Convention *T* required that from a correct truth definition, for a context-insensitive language *L*, every sentence of the form (*T*), where 'is *T*' is the truth predicate for *L* in the language of the theory, be derivable, where '*s*' is replaced by a description of an object language sentence in terms of its composition out of its simple meaningful constituents, and '*p*' is replaced by a sentence that translates *s* into the language of the theory. If we know that the sentence that replaces '*p*' translates *s*, then we can replace 'is *T* iff' to obtain (*M*).

(*T*) *s* is *T* iff *p*.

(*M*) *s* means that *p*.

Thus, from axioms that themselves use metalanguage expressions (expressions in the language of the theory), in specifying the contribution of object language expressions to truth conditions, which translate those expressions, we can produce theorems that we can use to interpret object language expressions in the light of our knowledge that the theory meets Convention *T*. Generalizing this to natural languages, which contain context-sensitive elements such as tense, indexicals such as 'I' and 'now', and demonstratives such as 'this', 'that', 'then', 'there', and so on, requires treating the truth predicate either as applying to utterances, or as relativized to at least speaker and utterance time.

Employing a truth theory as the vehicle of a meaning theory enables us to achieve the goal of a meaning theory – provided that we understand this to be met when understanding of the theory *puts one in a position to interpret* utterances of sentences of the language on the basis of their structures and rules showing how the parts contribute to what is expressed by an utterance of the sentence. It does this without appeal to entities such as meanings, properties, relations, or any other abstract objects assigned to words and sentences. At the same time, it provides a framework for investigations of

logical (or semantic) form in natural languages by requiring that a role be assigned to each word or construction in the language that determines its systematic contribution to the truth conditions of any sentence in which it is used.

In "Truth and Meaning," Davidson had proposed that a merely extensionally adequate truth theory for a natural language would also meet a suitable analog of Tarski's Convention *T*. A natural language containing context-sensitive elements, particularly demonstratives, requires axioms that accommodate any potential application of a predicate to any object a speaker might demonstrate, putting greater constraints on a correct truth theory for a context-sensitive language than for one that is not. If any true truth theory met a suitable analog of Convention *T*, then merely showing that a theory for a language was true would enable one to use it, in the fashion just described, to interpret speakers of that language. However, this is not adequate, since replacing one extensionally adequate axiom with another will not disturb the distribution of truth values over sentences, though it may result in a failure to meet (an analog of) Convention *T* (for details, see Chapter 1, §5). Davidson returned to the question of what informative constraints one could place on a truth theory in order for it to be used for interpretation in "Radical Interpretation." (See Chapter 1 for further discussion.)

4. PHILOSOPHY OF ACTION

"Actions, Reasons, and Causes" (Davidson 1980 [1963]) defended the view that reasons – that is, beliefs and desires (or pro attitudes) in the light of which we act – are causes of actions, conceived as bodily movements (broadly construed to include mental acts), and that action explanation is a species of causal explanation. Action explanations cite belief-desire pairs that conjointly cause the action, but that also show what was to be said in favor of it from the point of view of the agent. The desire (or, more generally, pro attitude) specifies an end that the agent has, and the belief links some particular action to some likelihood of achieving the end. Davidson calls action explanations "rationalizations." On this view, the concept of an action is a backward-looking causal concept, in the sense that it is the concept of an event (a bodily movement) that is caused and rationalized by a belief-desire pair. The concepts of belief and desire, on the other hand, are forward-looking causal concepts, in the sense that they are understood as concepts of states with a propensity conjointly to cause

bodily movements. This basic picture of the nature of human action and its relation to our reasons for acting was elaborated, extended, and refined in a series of articles. "How Is Weakness of the Will Possible?" (Davidson 1980b [1970]) takes up a specific puzzle about how irrational behavior is possible, namely, the puzzle of how someone can intentionally do something that he does not believe, all things considered, is the best thing for him to do. Davidson here abandons a view he had held in "Actions, Reasons, and Causes," namely, that the propositions that express a person's reasons for action are deductively related to the proposition that expresses the desirability of the act that they would rationalize; rather, the conclusion drawn about the desirability of the act is conditioned by the specific reasons for it, and not detachable from them. The causal account is deployed in explaining the possibility of weakness of the will by distinguishing between which reasons for action are causally strongest and which reasons provide the best grounds for action. "Agency" (Davidson 1980a [1971]) takes up the question of the relation between an agent and those events that are his actions; this essay defends the view that actions are bodily movements that can be picked out under different descriptions – under some of which an action can be intentional and under others of which it is unintentional – and that an action may be described in terms of its effects – so that a killing, for example, is nothing more than a bodily movement that causes a death, and so occurs before the death does. Davidson despairs of a final analysis in this paper, largely because of the problem of deviant causal chains – that is, the problem of describing how reasons must cause an action or event for it to count as an action done for those reasons (see Chapter 2, §5, for further discussion). "Freedom to Act" (Davidson 1980a [1973]) defends the causal theory against the charge that it allows no room for free action. "Intending" (Davidson 1980 [1978]) returns to, and rejects, a claim made in "Actions, Reasons, and Causes," namely, that acting intentionally is acting with an intention, and that the phrase 'an intention' in 'acting with an intention' is syncategorematic, merely signaling by what follows 'with' another description of the action in terms of its reasons. The paper instead identifies intentions as distinct attitudes that play an important role in mediating reasons and the actions they cause. This revision of the earlier view had already made an appearance in "How Is Weakness of the Will Possible?" where it plays a role in the explanation of how one can form a judgment, all things considered, to do something, and yet not form an all-out or unconditional judgment in favor of it (i.e., an intention to do it) but instead form an all-out judgment in favor of (an intention to do) something judged, all things considered, less favorably. (See Chapter 2, §4 for an extended discussion.)

5. RADICAL INTERPRETATION

The project of radical interpretation is mentioned in "Truth and Meaning," where Davidson takes a truth theory for a language to be an empirical theory, to be confirmed for particular speakers or groups of speakers on the basis of their behavior. It first takes center stage, however, in "Radical Interpretation" (Davidson 1984b [1973]). The project is that of interpreting another speaker without the usual assumptions of commonality of language. The description of the project of radical interpretation aims at illuminating what it is to speak a language by describing how a theory for interpreting a speaker could be confirmed by evidence that did not already presuppose any knowledge of what the speaker means by his words. The guiding idea is expressed in this passage from "Belief and the Basis of Behavior":

> Everyday linguistic and semantic concepts are part of an intuitive theory for organizing more primitive data, so only confusion can result from treating these concepts and their supposed objects as if they had a life of their own. (Davidson 1984a [1974], p. 143)

Specifically, in light of the commitment to using a truth theory as the vehicle of a meaning theory, the data to which Davidson restricts the radical interpreter is knowledge of the speaker's *hold-true attitudes*, that is, his beliefs about what sentences of his language are true, and how he interacts with his environment and with others like him. Though Davidson takes the interpreter to have access to a speaker's hold-true attitudes, these may be presumed to be identifiable ultimately on the basis of more primitive evidence.

In a nutshell, the radical interpreter's procedure involves identifying correlations between hold-true attitudes directed toward sentences, on the one hand, and events and circumstances in the speaker's environment, on the other. The speaker's hold-true attitudes are assumed to be the result of his knowledge of what his sentences mean and what he believes. If he knows that *s* means that *p* and believes that *p*, then he holds true *s*. Then, if we know that, for example, ceteris paribus, the speaker holds true *s* at a time iff there is a white rabbit in his vicinity, and we can assume the speaker is mostly right about his environment, we can with some justification infer that at any time *t*, *s* means that there is at *t* a white rabbit about. The assumption that the speaker is mostly right, both about his environment and in his beliefs generally, as well as by and large rational, Davidson calls the *Principle of Charity*. Davidson takes the Principle of Charity to be, not a contingent

assumption, but constitutive of what it is to be a speaker at all, and so not an option in interpretation.

The Principle of Charity can be separated into two strands, which Davidson has more recently called the Principle of Correspondence and the Principle of Coherence (Davidson 2001b [1991]). The first of these strands is the assumption that a speaker's beliefs, particularly about his environment, are by and large true. This plays a crucial role in bridging the gap between noticing correlations between a speaker's hold-true attitudes and his environment and assigning interpretations to his sentences. The Principle of Correspondence is a solution to the problem of separating out meaning from belief in hold-true attitudes. Which sentences a speaker holds true depends on what he thinks they mean and what he believes. If we knew either, we could solve for the other. Assuming that what the speaker believes is true, in the light of the conditions under which he has his beliefs, enables the interpreter to solve for meaning, and then to assign corresponding belief contents. The Principle of Charity is justified by the assumption that the position of the radical interpreter is the most fundamental position from which to investigate meaning and related matters, and it is needed to make sense of how the interpreter can see, on the basis of his evidence, another as a speaker. This assumption plays a central role in Davidson's epistemology and his arguments for the relational individuation of thought content. This is the view that, generally, what the contents of our thoughts are is a matter at least in part of our relations to things and events in our environments, so that we would not have had, as a matter of the *logic* of our concepts, the thoughts we do if our environments had been very different.

The Principle of Coherence has to do with the principles governing attributions of attitudes to an agent and descriptions of the agent's behavior so as to make the agent out to be by and large rational. It subsumes such principles as that, by and large, an agent's beliefs are consistent and his preferences transitive, and that attitudes are attributed in patterns that both (1) sustain the attribution of particular concepts to the agent by seeing them as fitting into a coherent pattern of beliefs deploying the concepts, and (2) enable us to see the agent's behavior as rationalizable in the light of his beliefs and pro attitudes. The Principle of Coherence is grounded in the analysis of the nature of agency, that is, in a priori principles governing our conception of what it is for anything to be an agent.

This represents an important point of connection between Davidson's work in the philosophy of action and his work in the theory of meaning. It is obvious also that any account of communication must involve the theory of action, since we understand speech, which is a form of action, only against a

background of complex intentions, some of which are directed toward how our words will be understood by others.

The procedure of the radical interpreter represents the attribution of attitudes and assignment of meanings as holistic in two different ways. First, there is an element of holism in the attribution of attitude content that derives from the requirement that attitudes be assigned in patterns that make sense of the speaker as a rational agent, and that make sense of the speaker as possessing the concepts used in characterizing his attitude contents. It is important to note in this regard, however, that Davidson does not think that there is any particular list of additional attitudes that an agent must have in order to have a given one, but only some supporting cast of attitudes appropriately related (Davidson 2001b [1982], p. 98). Second, attitudes and meanings are assigned not one by one, but in the context of a theory of all of a speaker's attitudes and the whole of his language: the criterion for correctness of any given attitude attribution or of any given assignment of meaning to an expression is that it be a part of the overall theory of the speaker's attitudes and language that is a best fit with all of the relevant evidence, the one that makes best sense of the speaker as a rational agent responding appropriately to events in his environment and to other speakers.

In work beginning in the late 1970s, as already noted, Davidson sketched more explicitly how the principles of decision theory can be employed in radical interpretation. In particular, in "A New Basis for Decision Theory," he outlined a procedure for identifying logical constants on the basis of patterns among preferring true attitudes. This subsumes the portion of the procedure of the radical interpreter that Davidson had concentrated on in earlier papers, rather than replacing it. (See Chapter 3 for further discussion.)

6. PHILOSOPHY OF PSYCHOLOGY

Davidson's main contributions to the philosophy of psychology, apart from his contributions to that branch that subsumes the philosophy of action, are (1) his thesis of, and arguments for, anomalous monism; (2) his arguments for a nonreductionist account of the relational individuation of thought content; and (3) his arguments for the necessity of language for thought. Each of these positions in the philosophy of mind is connected, more or less directly, to his reflections on the nature of language as seen from the standpoint of an interpreter of another speaker.

(1) Anomalous monism is the thesis that all mental events (more specifically, mental events involving propositional attitudes) are token (as opposed to type) identical with physical events (monism), and that there are no strict psychophysical laws (anomalousness). This proposal, and Davidson's argument for it, have been very influential in discussions of the relation between mental and physical events and properties. Its interest lies in its embracing materialism while eschewing reduction of mental properties to physical properties, and in the argument for the important claim that there cannot be, even in principle, strict laws connecting psychological and physical vocabularies. This thesis was first advanced (in print) in "Mental Events" (Davidson 1980c [1970]). Davidson there gives an argument for token-token identity of mental with physical events that relies on the thesis that there are no strict psychological laws. The argument, in brief, is as follows:

(1) Mental events causally interact with physical events.
(2) If two events stand in the causal relation, they are subsumed by a strict law.
(3) There are no strict psychological laws, only strict physical laws.
(4) Therefore, every mental event is subsumed by a strict physical law. (1–3)
(5) If an event has a physical description in terms suitable for subsumption by a strict law, it is a physical event.
(6) Therefore, every mental event is also a physical event. (4–5)

A strict physical law is one that figures in a "comprehensive closed system guaranteed to yield a standardized, unique description of every physical event couched in a vocabulary amenable to law" (Davidson 1980c [1970], p. 223–4). The argument for the anomalousness of the mental (3), with which "Mental Events" is primarily concerned, is notoriously difficult. It depends on the idea that different families of concepts are governed by different constitutive principles, and that for laws to be strict they must be couched in terms that are drawn from a single family. The idea, roughly, is that if applications of the predicates of each family are to retain allegiance to the constitutive principles that govern them, evidence in the form of correlations connecting them cannot give us reason to think that such correlations will be projectible to future instances. The constitutive principles for the attribution of attitudes are just those that provide the framework for radical interpretation, which seeks to fit observed behavior into a pattern provided by the constitutive structure of the concepts of the theory of agency and interpretation. (See Chapter 4 for further discussion.)

If the thesis of the anomalousness of the mental is correct, it shows that there are limits to the extent to which psychology may aspire to be a science like physics, since it precludes the possibility of a comprehensive closed system of psychological laws for predicating and explaining behavior.

(2) Davidson's nonreductionist account of the relational individuation of thought content rests on reflections on what assumptions the radical interpreter has to make in order to succeed in fitting the concepts of the theory of interpretation onto behavioral evidence. The radical interpreter interprets another on the basis of evidence that consists in part essentially in what prompts behavior of a speaker that is potentially interpretable as intentional. An idea implicit in the adoption of this position as basic to understanding meaning and the propositional attitudes is that what a specific utterance means, and what a particular thought is about, depends upon how a speaker is embedded in his environment. While this idea is implicit in the basic methodological stance that Davidson takes on meaning and thought, it comes to prominence only in essays of the 1980s and 1990s.

The discussion develops in two phases. In the first, Davidson brings out the reliance of the interpreter on correlating hold-true attitudes with events and conditions in the environment as his first entry into what a speaker believes and what he means by his words. If we can assume that to be a speaker at all requires that he be interpretable in any environment in which we find him, it will follow that what a speaker's thoughts are about will depend on what their pattern of typical causes is, for it is only by linking a speaker's thoughts to their typical causes, as identified by an interpreter, that interpretation from the third-person point of view is possible. This connection is already embodied in the treatment of the Principle of Charity as a constitutive principle of correct interpretation. This makes the concepts of the propositional attitudes also backward-looking causal concepts (if Davidson is right), because their causal history is essential to their individuation. This provides another ground for the thesis of the anomalousness of the mental, since causal concepts do not figure in strict laws.

In the second phase, Davidson emphasizes the importance of communication as a way of narrowing down the choice of relevant causes of a speaker's thoughts. Many causes of any given thought can be isolated for attention by treating different elements of the total physical cause as part of the background, and there are potentially many candidates for what a thought is about along any causal chain leading up to the thought. Which one is the right one? What objective criterion tells us what the thought is about? The suggestion that Davidson makes is that it is the "triangulation" between interpreter, speaker, and a common object of thought. That is, it is

only in the context of interpretation, a context in which a speaker and interpreter are responding to each other's common response to a stimulus in the environment, that we can find an objective determinant of what a thought is about. The object of the thought is where the causal chains leading to each common response intersect.

(Perhaps some additional work is required, for it is not clear that there are not also many common causes of common responses for two communicants in any situation. Imagine two people watching the news on television – there are events at the screen's surface, in the cable, at the cable station, in a satellite in geosynchronous orbit, and at distance trouble spots around the world, which are common causes of their thoughts. It is no different in other situations in which it is more difficult to identify all of the links in the causal chains.)

In some passages, it sounds as if Davidson thinks that it is only if there is an actual interpreter that it is possible to say determinately that a speaker has a thought. "If we consider a single creature by itself, its responses, no matter how complex, cannot show that it is reacting to, or thinking about, events a certain distance away rather than, say, on its skin" ("The Second Person" [Davidson 1992a, p. 263]). But a more plausible interpretation is that we can make sense of what a speaker's thoughts are about only against the background of a pattern of interaction with other speakers.

Importantly, while these connections between our thoughts and our environments are treated as constitutive of them, and as essential for their correct individuation, there is no suggestion in Davidson's arguments that we can offer any conceptual reduction of what it is to have thoughts, or to be a rational agent, to anything else. It is rather an upshot, on Davidson's view, of the character of the irreducible concepts of the theory of interpretation and agency that they organize data that includes the pattern of interaction between speakers and their environments. However, as we have seen, Davidson also argues that the constitutive principles governing the concepts that are thus fit onto behavior preclude even any strict projectible correlations between mental and physical properties.

(3) A third important theme in Davidson's work in philosophical psychology is the thesis, first advanced in "Thought and Talk" (Davidson 1984 [1975]), that language is essential for thought.

(1) One can have any propositional attitudes only if one has beliefs.

(2) One can have beliefs only if one has the concept of belief.

(3) One can have the concept of belief only if one is a speaker.

(4) Therefore, one can have thoughts only if one is a speaker.

The thesis is important in the light of Davidson's other commitments, for the assumption that the position of the interpreter of another speaker is methodologically basic in investigating meaning and thought presupposes that our understanding of thought in general is connected to having a language. An argument for the claim that thought requires language supports this position by fending off a potential objection – namely, that we understand what it is for nonlinguistic animals to have thoughts, so that the emphasis on the role of a truth theory serving as an interpretation theory in understanding thought as well as language is fundamentally misguided.

In addition, the argument for the third premise is also an argument directly for accepting the stance of the interpreter as basic in understanding meaning and the propositional attitudes. The central idea of the argument is that to have the concept of belief we must be able to understand what it is to fall into error. To understand this, there must be a role in our experience for judging that someone has made a mistake. Davidson claims that it is only in the context of communication that there is a role for judging someone to have made an error, as a way of adjusting one's overall picture of an interlocutor in order to make him out to be more rational than otherwise. "Only communication can provide the concept, for to have the concept of objectivity [i.e., of the contrast between truth and falsehood, and, hence, of error] . . . requires that we are aware of the fact that we share thoughts and a world with others" (Davidson 1991a, p. 201). For example, to make sense of an agent's drinking a glass of gasoline, in the light of desires plausibly attributable to him in the light of past behavior, including verbal behavior, we may wish to ascribe to him the false belief that the glass contains gin. Of course, the argument is successful only if there is scope for the application of the concept only in the context of communication. This is not immediately obvious. Prima facie, there is a point to the application of the concept of error wherever mistakes are apt to be made. Even a completely solitary thinker, who has never communicated with others, may have occasion to be surprised by finding that some result he expected does not obtain.

7. EPISTEMOLOGY

The central feature of Davidson's account of our knowledge of our own minds, the external world, and the minds of others is the denial of the epistemic priority of knowledge of our own minds over knowledge of the external world and the minds of others. He argues that each of these three varieties of knowledge is coordinate; none is reducible to the others, but

each is necessary for each of the others. The assumption that the basic standpoint from which to investigate the nature of meaning and the proposition attitudes is that of a radical interpreter of another speaker is the arch stone of the argument.

Forms of skepticism about the external world and other minds assume that we know facts in some domain (our own minds, the behavior of other bodies), that we are faced with the task of constructing an argument from those facts to facts in another domain (the external world, the minds of others), and that there is no a priori route from the one to the other, because propositions about each domain are logically independent of those about the other. Davidson's strategy in responding to skepticism about each of these domains is to deny the assumptions of logical independence that the skeptic relies upon.

In the case of skepticism about the external world, the result falls out of that part of the Principle of Charity that Davidson calls the Principle of Correspondence. In more recent formulations, Davidson has characterized this as the assumption that "the stimuli that cause our most basic verbal responses also determine what those verbal responses mean, and the content of the beliefs that accompany them" (Davidson 2001b [1991], p. 213), so as to guarantee that, in basic cases involving beliefs about our environs, our beliefs are for the most part true and about things, events, and states in the environment that are (or have typically been) prompting them. If this is correct, it immediately undercuts the most radical form of skepticism about the external world, because it guarantees, as a condition for having any beliefs at all, that most of our beliefs about the world are and have been true. (Whether this restores us to our full epistemic innocence is not clear; see Chapter 6 for further discussion.) This assurance that most of our empirical beliefs are true (or, as Davidson has also put it, that massive error in our empirical beliefs is impossible) is not directly an argument for our having knowledge of the external world. But this provides a response to the skeptic in two ways. First, it rejects the assumption of logical independence that the skeptic relies upon. Second, the general guarantee that most of our beliefs are true provides a test of any given belief by the degree to which it is supported by most of our beliefs. A belief that "coheres" well with most of our beliefs (a belief that if false or less likely to be true would entail that most of our beliefs were false or not likely to be true) may be assumed to be true or likely to be true. Thus, the guarantee that most of our beliefs are true provides a way of satisfying what might be thought to be a general requirement for a belief's being justified – namely, that it be, in principle, possible to provide a reason to think it more likely to be true than not.

To use coherence as a test for the likelihood of truth against the background of a guarantee that most of our beliefs are true is not to endorse a coherence theory of truth, which argues that our concept of truth is to be explained in terms of the coherence of beliefs with other beliefs. In Davidson's view, we take the concept of truth as primitive, argue that most of our beliefs are true on the basis of general considerations, and then observe that in this case coherence provides a test of the likelihood of truth.

The connections between knowledge of other minds and knowledge of our own minds and the procedures of the radical interpreter are even more direct than in the case of knowledge of the external world. In the case of our knowledge of other minds, it is obvious that if we insist that it is constitutive of meanings and attitude contents that they are accessible from the standpoint of the radical interpreter, the suggestion that we might never be in a position to be justified in believing that others have minds, or to be justified in our beliefs about what their thoughts are, cannot get off the ground. Taking the standpoint of the radical interpreter as basic is tantamount to saying that facts about meanings and attitudes logically supervene on facts about behavior, where this includes interaction with the environment.[4] Thus, on the assumption that we can have knowledge of our environments, knowledge of other minds is guaranteed to be possible because we are thereby guaranteed access to the facts that fix others' attitudes.[5]

This leaves knowledge of our own minds. Knowledge of our own minds is not based, in the first instance, on inference from behavior, or, apparently, anything else. On the face of it, this poses a problem for the assumption that facts about behavior are all that is relevant to fixing what attitudes a speaker has. For if all the facts that logically fix what our thoughts are and what our words mean are facts about our behavioral dispositions and our interaction with our environments, it is puzzling how we can have noninferential knowledge of them. Why are we not required to investigate our environments in order to discover what we think, in the same way that we learn what others think? Davidson's answer, in brief, is that it is an unavoidable presumption of interpretation that another knows what he means and thinks, while there is no such presumption about the interpreter. That this is an assumption of the procedure Davidson describes can be seen from the role that hold-true attitudes play in the interpreter's procedure. For hold-true attitudes are the result of what the speaker believes and what his words mean. We assume that the speaker holds true a sentence *s* if he believes that *p* and believes that *s* (on the occasion) means that *p*. For the content assigned to the sentence to be read back into the speaker's belief, we must assume that the speaker knows both what he believes and

what his sentences mean. (While this is implicit in Davidson's account of radical interpretation, his actual arguments are somewhat more involved; for a detailed discussion of Davidson's arguments, see Ludwig 1994.) If what must be assumed in order for interpretation of another to be possible is constitutive of him, then we can infer that it is constitutive of speakers (and, on the assumption that in order to think one must be a language user, of thinkers in general) that they know their own thoughts and what they mean by their words. Knowledge of our own thoughts is thus seen as having the same status as the thesis that in basic cases our thought contents are fixed by what causes our beliefs. (See Chapter 6 for additional discussion.)

8. EVENTS

There are two connected strands in Davidson's work on the topic of events. The first is an argument for seeing events as basic to our commonsense ontology. The second is an account of the nature of events as repeatable particulars, though Davidson's account of how events should be individuated has undergone some modifications over the years.

The argument for seeing events as basic to our commonsense ontology rests on two things. The first is that our commonsense ontology is codified in the commitments that we undertake in the range of sentences of our languages that we hold true. The second is that the best systematic account of the truth conditions of action sentences (sentences that involve attributing to someone an action of some sort) involving adverbial modification commits us to quantifying over events. Thus, given that we hold true many action sentences, we are committed to the existence of events.

A simple action sentence such as 'John turned on the light' may be modified by any number of adverbs. Thus, we may add, 'John turned on the light at sunset, on the twenty-third of August, while reading in bed'. From this there follows 'John turned on the light', 'John turned on the light at sunset', 'John turned on the light on the twenty-third of August', 'John turned on the light while reading in bed'. An account of the semantic form of these sentences must make sense of these entailment relations on the basis of a rule that determines the contribution of the adverbs to the truth conditions for the sentence. Briefly (see Chapters 1 and 5 for further discussion), Davidson treats the main verb as introducing a existential quantifier over events, and treats the adverbs as contributing predicates of the event variable implicitly introduced.[6] Thus, 'John turned on the light at sunset, on the twenty-third of August, while reading in bed', might be represented

as sharing logical form with 'There is an event *e* such that *e* is a turning on of a light by John and *e* was at sunset and *e* was on the twenty-third of August and *e* was done while John read in bed'. This neatly provides both an account of the productive character of adverbs on the basis of a familiar rule and explains the entailment relations that such sentences intuitively enter into on the basis of their form. Davidson has also argued influentially that quantification over events is involved in singular causal statements, such as the statement 'The short circuit caused the fire', and that such statements are not about, in particular, anything that could be called necessary or sufficient conditions for the effect as described (Davidson 1980a [1967]).

Davidson holds that events are datable particulars, and, in particular, changes. In his earliest papers, Davidson argued against the view that events could be individuated on the basis of the region of space-time that they occupy, giving the example of a sphere that was spinning and warming up at the same time. Intuitively, the rotation of the sphere and its rise in temperature are distinct events, different changes. He proposed instead that events should be individuated by their causes and effects (Davidson 1980 [1969], p. 179). That is, for any two events $e_1, e_2, e_1 = e_2$ just in case e_1 and e_2 have all the same causes and all the same effects. He later gave this up in response to a criticism from Quine, who pointed out that this fails as a criterion of individuation because the individuation of any given event would presuppose the prior individuation of others (Quine 1985). Davidson adopted Quine's suggestion that events be individuated in terms of the spatiotemporal regions they occupy. For any events $e_1, e_2, e_1 = e_2$ iff e_1 and e_2 occupy exactly the same spatio-temporal region. This is to accept that, after all, a sphere's rise in temperature and its spinning should be identified. This is a position that Davidson considered, without adopting it, even in "The Individuation of Events." There he suggested that one might hold that "the warming of the ball ... is identical with the sum of the motions of the particles that constitute it ... and so is the rotation" (Davidson 1980 [1969], pp. 178–9). (See Chapter 5, §6.)

9. TRUTH

Davidson is often seen as aiming to illuminate the concept of meaning by appeal to the concept of truth. There is certainly an element of this in his general strategy of aiming to place enough formal and empirical constraints on a truth theory to ensure that it meets an analog of Convention *T*. There is no suggestion in this, however, of any direct reduction of the

concept of meaning to that of truth. Rather, the aim is to illuminate the concept of meaning by relating it systematically to other concepts and to evidence that would confirm an interpretation theory for a speaker. That no reduction of the concept of meaning to that of truth is intended can be seen from the fact that Davidson makes free use of the concept of meaning in characterizing the goal of the radical interpreter. His aim is not just to confirm a true truth theory, but an interpretive one.

As far as the use of a truth theory for the purposes of a compositional meaning theory goes, it is important to note that this does not itself presuppose any substantive view about the concept of truth. For this use of a truth theory exploits the recursive structure of the theory, rather than any account of how the truth predicate is to be analyzed, if at all. The only commitment this carries with respect to the analysis of the concept of truth is that whatever the correct analysis or account is, it respect Convention *T*, or the appropriate analog for natural languages.

The concept of truth itself Davidson holds to be one of the most basic concepts that we deploy, and he has opposed attempts to either deflate it or reduce it. "Without grasp of the concept of truth, not only language, but thought itself, is impossible" (Davidson 2000c, p. 72). It is not amenable to analysis in other terms, and it is a concept possession of which is at least coordinate with other concepts that we use in describing our ability to speak and understand others. As Davidson says in one place, "All these concepts [intention, belief, desire] (and more) . . . are essential to thought, and cannot be reduced to anything simpler or more fundamental" (Davidson 2000c, p. 73). If this is right, the best one can do to illuminate the concept of truth is to show how it is systematically related to other concepts and to the evidence on the basis of which we apply it to utterances and thoughts. This is provided for in Davidson's work by the central role he gives to the concept of truth in the theory of interpretation.

Opposing views fall into two camps: those that seek to reduce truth to something else – the correspondence theory, the coherence theory, and pragmatism – and those that do not reduce truth to something else but argue that there is not much to it – the redundancy theory, and varieties of deflationary theories. The coherence and pragmatic theories receive summary dismissal. Against the correspondence theory, Davidson has deployed a variant of the so-called slingshot argument (see Chapter 1, §2), which aims to show that all true sentences correspond to the same fact. Apart from this, he argues, following Strawson (1965), that since the only way we can identify facts is by appeal to true sentences that correspond to them, appeal to facts and correspondence fails to explain the truth of sentences

(see "True to the Facts," Davidson 1984 [1969]). If we try to say what fact makes it the case that 'The source of the Nile lies in the Mountains of the Moon' is true (that is, what fact in virtue of corresponding with which it is true), we can do no better than to say that it is the fact that the source of the Nile lies in the Mountains of the Moon; and it is a fact that the source of the Nile lies in the Mountains of the Moon if and only if the source of the Nile lies in the Mountains of the Moon. We might as well have said: 'The source of the Nile lies in the Mountains of the Moon' is true iff the source of the Nile lies in the Mountains of the Moon.

Davidson has suggested that the varieties of deflationary theories that have arisen in recent times are a reaction against inflated views of what theories of truth can do (viz., deliver the Truth), and misguided attempts to reduce it to something else. The simple redundancy theory, which holds that any sentence of the form '*s* is true' may be replaced without loss of content with *s*, is the precursor to more recent deflationary theories. It suffers the defect that it cannot handle uses of 'is true' in which the sentences said to be true are not given with the sentence, as in 'Everything Aristotle said is true'. More recent deflationary theories argue that the content of the concept of truth is exhausted by our recognition that every instance of a Tarskian *T*-schema is true: '*p*' is true iff *p*. A defect of this approach when it appeals to the truth predicate as applied to sentences is that it does not tell us how to apply it across languages. The schema tells us only what the extension of 'is true' is when the metalanguage contains the object language (and then only when the language does not allow the construction of semantic paradoxes). It is also useless in the case of natural languages that contain context-sensitive expressions. Let us try disquotation with a context-sensitive sentence: 'I am hungry' is true iff I am hungry. Because it is a context-sensitive expression itself, no utterance of this correctly represents the conditions under which an utterance in general of 'I am hungry' is true. Suppose I now assert: 'I am hungry' is true iff I am hungry. Let *t* be the time of my assertion. Then I have said that 'I am hungry' is true iff Ludwig is hungry at *t*. But this does not give the truth conditions for any assertion of 'I am hungry' by anyone else at any time.

These difficulties have motivated a move to a propositional version of the approach, which appeals to the schema, 'The proposition that *p* is true iff *p*' (restricted to instances that do not give rise to paradox). Davidson has argued, against this, that we lack an account of the semantics of 'The proposition that *p*' that can serve the deflationists' purposes. If we are Fregeans, we should take 'The proposition that *x*' to be a functional expression; it is naturally interpreted as mapping a truth value onto itself, and then we take

the sentence '*p*' to denote a truth value; instances of the schema are then trivially true, but they explain nothing about truth, and rather presuppose a grasp of it. If we do not take this route, then we stand in need of an account of the function of the sentence that appears after 'The proposition that'. If we take the sentence to be mentioned, then it must be relativized to a language, for the same sentence may express different propositions in different languages. Then the concept of truth is exhibited as interconnected with the concept of meaning, and this undercuts the deflationist's attempt to show that the concept is trivial and uninteresting. (See, in particular, Davidson 1996a; 2000c.)

10. RATIONALITY AND IRRATIONALITY

Davidson's work on irrational action is a direct outgrowth of his work on the basic nature of human agency. Rationality is a condition on agency. To describe an object as an agent with psychological attitudes capable of performing actions requires finding in its behavior, or in its dispositions to behave, a pattern that can be described and explained in terms of attributions of patterns of interlocking attitudes that motivate and rationalize what the agent does. Irrational behavior emerges only after we have identified something as an agent. Davidson's fundamental thesis about irrational action or thought is that it is to be viewed not as *non*rational behavior, but as a perturbation of rationality, a disturbance in a largely rational pattern of thought and action, since no object can be irrational except insofar as it is an agent and, hence, largely rational. We identify a particular thought or a particular piece of behavior as irrational by its failure to conform fully to the rational pattern of the rest of the agent's attitudes and behavior. Its irrationality is to be located in its relation to the rest of the thoughts and behavior of the individual. The inconsistency thus identified is internal to the agent, and is a matter of the agent's not adhering to norms that he recognizes are constitutive of the attitudes.

Davidson has written persuasively about a range of types of irrational behavior as well as about the nature of irrational behavior more generally. In the case of weakness of the will, Davidson asks how it is possible for someone to judge that a certain thing is the best thing to do, all things considered, and yet do something else, which seems to imply an all-out judgment in favor of it instead. The answer lies in distinguishing between the normative force of someone's attitudes, which guide his deliberations about what to do, and the causal force of various desires; and in distinguishing between the judgment

that the agent makes on the basis of his survey of his preferences and beliefs, and the actual forming of an intention to do the thing that he thinks best, which corresponds to the all-out judgment. This distinction between the judgment that a certain thing is the best thing to do, all things considered, and forming the intention, makes room for the possibility of a desire or belief leading to intention formation at odds with the agent's own judgment about what is best. (We tend to think here of passions overwhelming our better judgment, but as Davidson notes, habit is just as likely to induce behavior at odds with our best judgment.)

The mechanism involved in weakness of the will provides a model for explaining the possibility of irrational behavior and thought in general (obsessive behavior, self-deception, wishful thinking, etc.). It involves, on the one hand, a subset of desires and beliefs in the light of which a given action, or pattern of thought, is seen as desirable and, on the other hand, a backdrop of the rest of the agent's attitudes in the light of which the action or pattern of thought or behavior is not reasonable or best, all things considered. Irrational behavior or thought occurs when some beliefs and desires have causal power that acts independently of what is recommended in the light of all of an agent's attitudes and preferences (when "forms of causality . . . depart from the norms of rationality" [Davidson 1985c, p. 347]). In these cases, the harmony among all of an agent's attitudes that makes for rationality is disrupted and distorted – but not massively so, on pain of the agent's behavior ceasing to be identifiable as intentional behavior or as expressing thought at all. Davidson holds that we cannot make sense of an agent who believes a proposition that is patently inconsistent, such as that he is and is not middle-aged, overweight, balding, and unattractive. Yet an agent may believe that he is middle-aged, overweight, balding, and unattractive, and also believe the opposite. This kind of inconsistency between, rather than within, beliefs is possible. But if so, then when irrationality involves beliefs that are in conflict with each other, they must be compartmentalized: the agent cannot bring them together. All forms of irrationality that involve such inconsistency, then, also involve a form of mental division or causal insulation of different sets of attitudes.

11. NONLITERAL USES OF LANGUAGE

Nonliteral uses of language extend beyond literature, and not all literary texts are nonliteral, but it is in its nonliteral aspect that literature poses a puzzle for accounts of the nature of language that emphasize that the goal

of language is communication and truth telling, or, in the case of uses of imperatives and interrogatives, the issuing of directives and asking of questions. The puzzle is how to fit the nonliteral uses into such a picture, and how to accommodate such a picture to the ubiquity of nonliteral uses of language. While Davidson has not written a great deal directly about literature, he has written generally about the notion of the literal and conventional uses of expressions ("Communication and Convention," "A Nice Derangement of Epitaphs") and has contributed both to the theory of metaphor ("What Metaphors Mean") and to the theory of literature more generally ("Locating Literary Language") against the background of his philosophy of action. (This is the main topic of Chapter 7.)

The central theme of "Communication and Convention" is that conventions, particularly prior learned conventions, are not essential to communication, but that nonetheless something akin to what is usually thought of as literal meaning is essential – namely, what Davidson calls "first meaning," which is the meaning that a speaker intends his auditor to attach to an utterance of his as the basis for further things he intends to accomplish by making the utterance. It is this notion – which, Davidson argues, is not the same as the notion of a meaning learned prior to the communicative context – that is essential to the act of linguistic communication. It plays a central role in Davidson's account of the nonliteral uses of language, for these uses play off of first meaning. Nonliteral uses of language achieve their effects by what first meaning is attached to them. If this were not so, then there would be nothing specially linguistic about how the effect was achieved.

Literary works are intentional productions. They are the products of actions, and their productions are speech acts. It falls out of Davidson's account of the mechanism of communication that the key to interpreting any text is its first meaning, and that essential to its first meaning are the intentions of the author. Consequently, the interpretation of literary texts, considered as speech acts, cannot float free from authorial intentions. Davidson does not hold, of course, that it is only intentions relevant to first meanings that are relevant to the interpretation and significance of literary works, but only that there is no interpretation without first meanings. The effects of literary works are achieved through the combination of first meaning and context (in the broadest sense). Davidson emphasizes also, however, that the meaning of a text is not determined solely by an author's intention. He argues that for an intention to succeed in imbuing a text with a certain significance, "it is also necessary that the intention be reasonable," that is, that the author have a reasonable expectation that his readers could

figure out what he intended. This is an expression again of Davidson's view that matters of meaning must be publicly accessible.

Davidson's study of metaphor illustrates his general thesis. (See Chapter 7, §3, for additional discussion.) He argues that metaphors achieve their effects through their first meanings, and that we need not assign to them any additional metaphorical or nonliteral meanings to see how their effects are achieved. An author intends a metaphor, through the literal meaning of its words, to draw our attention to a comparison of things to one another. This need not be something that can be cashed out in terms of some specific claim. A rich metaphor is open-ended, and the intended effects of metaphors need not be limited to conveying propositional content. "Metaphor is the dreamwork of language," Davidson writes, "and, like all dreamwork, its interpretation reflects as much on the interpreter as on the originator." We understand this, a marvelous illustration and expression of Davidson's thesis, because we understand what the words in their ordinary employment mean; the full meaning (in another sense) of the metaphor is to be sought in what it encourages us to compare the metaphor to, and, like the best metaphors, the comparisons and extensions are open-ended. It also reminds us that success in metaphor is as much a matter of taste as it is of design.

12. ORGANIZATION OF THE VOLUME

This volume begins with chapters on truth and meaning and on the philosophy of action. The first discusses the relation between a theory of truth and the project of giving a compositional meaning theory for a natural language. The second discusses Davidson's contributions to the philosophy of action. These form a background for the discussion of the nexus of Davidson's philosophy, the project of radical interpretation, in Chapter 3. Chapter 4 takes up Davidson's most celebrated thesis in the philosophy of mind, the thesis of anomalous monism, and his difficult argument for it. Chapter 5 discusses Davidson's work on the semantics of action sentences, which argues for there being a commitment to an ontology of events in many of the things that we assert about ourselves and others, and Davidson's views about the nature and individuation of events. Chapter 6 takes up Davidson's account of the grounding of and relations between our knowledge of our own and others' minds, and of the external world, in which reflections on radical interpretation play a central role. Chapter 7 discusses applications of Davidson's work in the theory of action and meaning to understanding the

language and intent of literature, which is part of the more general project of understanding the nonliteral use of language in light of its literal uses. Each chapter may be read independently, but earlier chapters will help to illuminate later chapters; in particular, Chapters 1 and 2 provide background for Chapter 3, and Chapters 1–3 for all subsequent chapters.

Notes

I wish to thank Donald Davidson for reading a draft of this Introduction and correcting a number of mistakes. I am of course responsible for any that remain.

1. In an interview in 1993, Davidson reported that before he took these courses from Quine he "had thought of philosophy as a form of literature, which indeed it is, but as no more open to basic questions of 'Is this right or wrong?' than poetry" (Bergstrom 1994, p. 223). Davidson also emphasizes the decisive influence of Quine in *Inquiries into Truth and Interpretation* (Davidson 1984b), which collects many of Davidson's papers in the philosophy of language up through 1982; it is dedicated to Quine, "without whom not." In the Introduction, he writes:

   > W. V. Quine was my teacher at a crucial stage in my life. He not only started me thinking about language, but he was the first to give me the idea that there is such a thing as being right, or at least wrong, in philosophy, and that it matters which. (p. xx)

2. This is reflected, for example, in Davidson's explanation of his methodology for illuminating concepts in "Radical Interpretation":

   > I have proposed a looser relation between concepts to be illuminated and the relatively more basic. At the centre stands a formal theory, a theory of truth, which imposes a complex structure on sentences containing the primitive notions of truth and satisfaction. The notions are given application by the form of the theory and the nature of the evidence. The result is a partially interpreted theory. (Davidson 1984b [1973], p. 137)

3. Radical interpretation differs from radical translation in several important respects. First, the radical interpreter aims not to formulate a translation manual, but to confirm a theory of truth for a speaker's language under conditions that ensure that it can be used to interpret the speaker's sentences. Second, the radical interpreter focuses not, as Quine's radical translator does, on stimulus at a speaker's sensory surfaces as grounding the meanings of his words, but on those distal stimuli that prompt his thoughts and speech. Both of these differences reflect a difference in aim. In *Word and Object*, Quine's aim is, in part, to give a scientifically respectable reconstruction of the notion of meaning; in contrast, Davidson's aim is to understand the ordinary notion.

4. Facts about *A* logically supervene on facts about *B* iff for any *B* fact, there is an *A* fact, such that it is logically necessary that if a fact of the latter type obtains, a fact of the former type obtains.

5. There is, prima facie, a tension between the claim that facts about attitudes and meanings supervene on facts about behavior and the view that facts about

attitudes and meanings are not reducible to facts about behavior. However, the supervenience thesis does not entail that there is any necessarily true biconditional linking claims about the mental with claims about behavior: there may be no end to the sorts of facts that could provide a supervenience base for any given mental fact. Whether the logical supervenience thesis is compatible with anomalous monism would require further investigation; the problem, on the face of it, is that if there are any necessarily true statements of the form, for any *x*, if *x* is *P* then *x* is *M*, where 'is *P*' is a physical predicate and 'is *M*' is a mental predicate, then if the physical predicate figures in a strict law there will be a corresponding strict psychophysical law.

6. Davidson credits Reichenbach with a similar suggestion in his *Elements of Symbolic Logic* (Reichenbach 1947, sect. 48), but arrived at the idea of introducing a quantifier over events to accommodate adverbial modification independently. Davidson has related in personal communication that when the idea occurred to him he couldn't believe that it was new, and so read everything he could find on adverbial modification. Reichenbach was one of the few who tried to fit adverbs into ordinary logic, and he did introduce an existential quantifier. But his proposal differs significantly from Davidson's – in quantifying over facts rather than events, and in not advocating the quantified sentence as revealing logical form, but rather as being logically equivalent to the adverbially modified sentence. The choice of facts as the domain of quantification also leads to difficulties that Davidson's proposal avoids.

1 Truth and Meaning

ERNEST LEPORE AND KIRK LUDWIG

Donald Davidson's work on the theory of meaning has been enormously influential since the publication of "Truth and Meaning" in 1967. His central proposal was that an understanding of what it is for "words to mean what they do" (Davidson 1984b, p. viii) can be pursued by way of constructing and confirming for a speaker an axiomatic truth theory, modeled on a Tarski-style axiomatic truth definition, for his language. In this chapter, we first discuss the background of Davidson's famous suggestion, initially introduced in "Truth and Meaning." We begin with his arguments for the importance of attending to the compositionality of natural languages in §1, then turn in §2 to his criticisms of traditional approaches to the theory of meaning. In §3, we discuss Davidson's introduction of a truth theory as the vehicle for a compositional meaning theory; in §4, we explore some interpretive issues that arise about his intentions, specifically the question of whether Davidson intended to *replace* the traditional project (providing an account of meaning) with a more tractable one (providing an account of truth conditions), or whether he intended to pursue the traditional project by novel means. We argue that, though Davidson has been widely misunderstood, his intention is clearly the latter, and, specifically, that his goal has always been to give an account of what illuminating constraints a truth theory can meet that would suffice for it to be used to understand any potential utterance of an object language sentence. We discuss the difficulties arising from his initial suggestion, extensional adequacy, in §5, and then, in §6, the role of radical interpretation, which is discussed in greater detail in Chapter 3. In §7, we consider how the project of a compositional meaning theory may be pursued independently of the more ambitious project in which Davidson embeds it, discuss various problem areas, and consider some of Davidson's contributions to natural language semantics. We conclude in §8.

1. LEARNABLE LANGUAGES AND THE COMPOSITIONALITY REQUIREMENT

In "Theories of Meaning and Learnable Languages" (Davidson 1984 [1966]), Davidson identified a requirement on any adequate theory of meaning for a natural language such as Chinese or English, namely, that it "must be possible to give a constructive account of the meaning of sentences in the language" (Davidson 1984 [1966], p. 3), and then argued that a number of then current theories of meaning failed to meet the requirement. What is it to give a constructive account of the meaning of sentences in a language? This presupposes that we understand some expressions – sentences, in particular – on the basis of other, less complex expressions and their mode of combination. A constructive account provides an account of the meaning of a sentence in terms of the meanings of its semantically noncomplex constituents and their modes of combination. A constructive account of what sentences in a language mean we will call a "compositional meaning theory"; a language that admits of a compositional meaning theory we will call "compositional."

What reason do we have to think that natural languages are compositional? Davidson offers a famous argument for this, known as the learnability argument (pp. 8–9). In a nutshell, the argument is that it is only if natural languages are compositional that we can understand how "an infinite aptitude [competence in speaking and understanding] can be encompassed by finite accomplishments" (p. 8). More fully: we are finite beings. We come into the world without language. We become, in a finite amount of time, fully competent speakers of languages that include an infinite number of nonsynonymous sentences. On the assumption that we cannot "intuit the meanings of sentences on no rule at all, and that each new item of vocabulary, or new grammatical rule, takes some finite time to be learned" (p. 9), we can conclude that there are a finite number of semantical primitives, and that we are put in a position to understand the rest of the expressions we are able to understand because their meanings are determinable from our mastery of the semantical primitives contained in them and rules governing how the meanings of complexes are determined by the meanings of the simples and their modes of combination.

Though not entirely uncontroversial (see, e.g., Schiffer 1987, esp. pp. 137–8), this observation, which accords with common sense, is largely accepted in work in semantics for natural languages. It had certainly been presupposed prior to Davidson's emphasizing its importance.[1] Davidson made salient, though, its importance and usefulness as a criterion of adequacy for analyses of logico-semantic form, and for understanding meaning

more generally. An example of an analysis that runs aground on this requirement is Israel Scheffler's analysis of indirect discourse (Scheffler 1954). Scheffler suggested that we analyze a sentence such as 'Tonkin said that snow is white' as 'Tonkin spoke a that-snow-is-white utterance'. The advantage of such an analysis is that it gives an account of what we mean by such sentences that does not commit us to treating 'that'-clauses as referring to propositions, while remaining sensitive to how we understand 'that snow is white' as used by the speaker. In the analysis, 'that-snow-is-white' is to be treated as a *simple unitary predicate* of utterances. It is not treated as further analyzable and, hence, is treated as a semantical primitive. However, as Davidson pointed out, applying this strategy to indirect discourse generally yields the immediate result that English (and other natural languages, presumably) has an infinite number of semantical primitives, since any sentence may feature in the complement of a report of indirect discourse, and there are an infinite number of nonsynonymous sentences in the language: each one, on this analysis, will have a corresponding primitive predicate of utterances. Thus, since this would make it unintelligible how we could learn English, we can conclude that, no matter what its other virtues, this analysis of the logical form of indirect discourse is mistaken.

2. DAVIDSON'S CRITICISMS OF TRADITIONAL APPROACHES

An adequate theory of meaning for a natural language must be compositional. A compositional meaning theory provides, in some as yet unarticulated sense, an "account" of the meaning of every sentence of the language. What form should a meaning theory take? In what sense should it give an account of the meaning of every sentence of the language?

A minimal constraint we should impose is that the theory enable someone who understands it to interpret any potential utterance of a sentence of the language. A natural first suggestion is that this could be accomplished if we had a theory that, from a finite base, enabled us to derive solely on formal considerations, for every sentence of the object language L, a theorem of the form (M) (where initially we will suppose, for simplicity, that we are dealing with a context insensitive language),

(M) s means in L that p,

where 's' is replaced by a description of a sentence of the object language in terms of how it is composed out of its significant parts, and 'p' is replaced by a sentence in the metalanguage (the language of the theory) that translates

the object language sentence described. If we know what a sentence of the form (*M*) expresses, then we know how to interpret the object language sentence it is about.

The question is how to do this, and whether it can be done in what may seem like the most straightforward way. Part of Davidson's motivation for proposing that a compositional meaning theory make use of the structure of an axiomatic truth theory was his pessimism about the prospects for any other way of accomplishing the goal. Traditional approaches to the theory of meaning, stretching back to Frege (Frege 1892; 1997b [1892]), have appealed to meanings – that is, to abstract entities – in trying to give a theory of meaning. Two terms are then said to be synonymous iff their meanings are identical. However, it is not clear how the appeal to meanings (or other such entities) helps.

Davidson observes that assigning a meaning to a subject term and a predicate in a sentence, such as 'Theaetetus flies', does not tell us anything about how their concatenation yields a new meaning (Davidson 1984 [1967], p. 17). If we are treating each term as referring to its meaning, then their concatenation seems simply to give us a list. Treating concatenation itself as significant just gives us one more meaning to combine. Treating, as Frege did, the referent of 'flies' as unsaturated or incomplete (Frege 1997 [1891]; 1997a [1892]), so that putting together (or indicating a relation between) Theaetetus and the referent of 'flies' yields (grasp of) a new kind of entity, "seems to label a difficulty rather than solve it" (p. 17).

Any constructive theory needs somehow to yield information about how the combination of terms contributes to meaning in a way that enables us to understand the complex expression if we understand the theory. It is not clear that associating entities with terms, or with their concatenation, could itself ever yield appropriate knowledge. What is missing is a rule attaching to the combination that yields in the theorizer's language a sentence understanding of which suffices (perhaps with some auxiliary knowledge) to understand the complex object language expression. In "Truth and Meaning," Davidson gives a simple example that does the job right in the theory of reference (p. 18). His example contains a number of general lessons, so it is worth reviewing. Consider a fragment of English, which we will call '*L*', consisting of the functor 'The father of *x*' and some proper names, say, 'Jesse', 'Saul', 'David', and 'Goliath'. We can give outright the referents of the proper names: 'Jesse' refers in *L* to Jesse; 'Saul' refers in *L* to Saul, 'David' to David, 'Goliath' to Goliath. Here we use names in the metalanguage that refer to what the names in the object language refer to (helping ourselves to a metalanguage that embeds the object language). We allow

that any concatenation of 'The father of' with a proper name is also a referring term, and now the task is to give an account of the referents of the rest of the terms in *L*. Do we need to assign to 'The father of *x*' some entity? No, a rule that we can state in our language suffices to give the referent in terms of the referents of the term that replaces '*x*' ('⌢' is the symbol for concatenation):

> For any referring term *r*, the referent in *L* of 'The father of'⌢*r* = the father of the referent in *L* of *r*.

The rule uses, of course, a term synonymous with the term in *L* that we are giving a rule for, but this is compatible with our goal, which was to provide, from a finite base, a theory understanding of which suffices to understand what the referent is of every referring term in *L*. To put this another way, our small theory meets the following constraint: it enables us to prove from a finite base every instance of the schema,

(*R*) *t* refers in *L* to *x*,

where '*t*' is replaced by a name or structural description of a referring term of *L*, and '*x*' is replaced by a term that refers to what *t* does.

Notice that, in fact, the theory generates theorems of the form (*R*) in which the term that replaces '*x*' is *synonymous with* the term that is denoted by what replaces *t*. This is because we chose terms in our theory when assigning referents to simple expression in *L* that were the same in meaning as those expressions, and we chose a functor in the language of the theory that was the same in meaning as that used to form complex referring terms in *L* when giving a rule for assigning referents to such expressions. Knowing this, we can also infer from each theorem what the expression in *L* means, not just what it refers to – that is, we can infer from each instances of (*R*) a corresponding instance of the schema (M_R), as, for example, in (*M′*).

(M_R) *t* means in *L x*

(*M′*) 'The father of'⌢'David' means in *L the father of David*.

That this provides genuine information is made clear when we consider the general case in which the metalanguage is distinct from the object language.

This simple example shows that we can meet our goal of providing an account of the referents of every expression of *L* without assigning a meaning or other object to every expression of *L*. In particular, the contribution of the expression 'the father of', which we use to form complex expressions,

can be made clear without assigning any entity to it in particular, as opposed to the complex expressions that it helps to form. Furthermore, we can see that without some rule like (*R*), which actually tells us how to interpret our functor, we will not be able to understand what complex expressions in *L* refer to from a finite base. Putting these two points together, we can conclude, with respect to this simple example, that assigning entities to every expression of the language is neither necessary nor sufficient to achieve the aims of a theory of reference. Moreover, in light of the fact that this theory can be used for the purposes of a theory of understanding, we can see that the same point carries over to something approaching a theory of meaning for *L*. This tiny theory turns out to foreshadow the basic form of the suggestion for how to pursue a meaning theory by way of a truth theory, which we take up in the next section.

The success of this theory, however, might encourage one to think that it could be extended, on a neo-Fregean account, to sentences, treating them as terms referring to their meanings and treating predicates and quantifiers as functional terms. Davidson offers a famous argument to scotch this proposal, dubbed the slingshot (for slaying the Fregean giant) by Barwise and Perry in a critical discussion (Barwise 1981; Barwise and Perry 1981a; 1981b; 1983).[2] The argument attempts to show that the view that sentences refer to their meanings is untenable, because it leads to the conclusion, on plausible assumptions, that any two sentences that are alike in truth value refer to the same thing and, hence, are synonymous, which is manifestly false. The argument depends on two assumptions: (A) "logically equivalent singular terms have the same reference," and (B) "a singular term does not change its reference if a contained singular term is replaced by another with the same reference" (p. 19). Suppose that '*R*' and '*S*' stand in for sentences alike in truth value. Consider (1)–(4).

(1) R

(2) $\{x\colon x = \text{x} \;\&\; R\} = \{x\colon x = x\}$

(3) $\{x\colon x = x \;\&\; S\} = \{x\colon x = x\}$

(4) S

Given (A) and (B), Davidson argues that we can identify the referent of (1) with that of (2), the referent of (2) with that of (3), and the referent of (3) with that of (4), showing that all sentences alike in truth value corefer. (1) has the same referent as (2), by (A), because they are logically equivalent. (2) has the same referent as (3), by (B), because they differ only in that one referring term, '$\{x\colon x = x \;\&\; R\}$', has been replaced by another, '$\{x\colon x = x$

& $S\}$', which corefers (both refer to the universal set if 'R' and 'S' are true, and to the empty set if each is false), in a containing singular term (namely, the whole of (2)). We infer that (3) and (4) have the same referent using (A) and the fact that they are logically equivalent.[3]

Is the argument successful? Assumption (B) is forced on anyone who takes predicates to be function terms, for replacing an argument term with another term that corefers supplies the same argument for the function, and so cannot change the value it yields. However, it has been objected that (B) has no application in the argument, on the grounds that the term '$\{x: x = x \ \& \ R\}$' is a quantified noun phrase (a definite description: 'the set of all x such that $x = x$ and R') rather than a singular referring term (Hochberg 1975). If we did interpret it as a quantifier phrase, certainly a neo-Fregean would object to the substitution principle.

Even apart from this, it might be doubted whether assumption (A) would be acceptable to the argument's intended audience. Surely anyone who accepted that sentences refer to their meanings would allow logically equivalent sentences to differ in meaning, and so in reference. Yet, if we consider (A), it is not so easy to see how to deny it. (A) says that logically equivalent singular terms corefer. What are logically equivalent singular terms? The notion needs to be extended from its standard application to sentences. It is natural to say that two singular terms are logically equivalent iff they corefer on all reinterpretations of their nonlogical terms. Thus, for example, 'Woody Allen' and 'Allen Steward Konigsberg', though they corefer, are not logically equivalent singular terms, whereas 'The x such that $x =$ Woody Allen' and 'Woody Allen' are logically equivalent singular terms.[4] If we so understand logical equivalence of singular terms, than (A) expresses a definition, and so is trivially true.

However, securing the truth of (A) in this way nonetheless undermines the argument. For it shows that an illicit assumption is made in the application of (A) to secure that the referents of (1) and (2) are the same. (1) and (2) are logically equivalent sentences – that is, they are alike in truth value on all reinterpretations of their nonlogical terms. But this notion of logical equivalence between sentences is not defined in the same way as the notion that applies to singular terms. We can't infer that sentences that are logically equivalent in the standard sense are thereby logically equivalent singular terms without begging the question. Thus, the argument is invalid without the addition of a question-begging assumption.

Despite the failure of the argument, there is not much to be said for treating sentences as referring to meanings. Grammatically, sentences are not referring terms. Assimilating them to referring terms could at best be

seen as a kind of technical maneuver adopted to achieve the happy end of matching object language sentences with metalanguage sentences that are used and which we know are the same in meaning as the object language sentences. This end, however, essential to a meaning theory, can be achieved without the superfluous introduction of meanings and the distortion of the semantical role of sentences, as we will see.[5] If this is right, then, as Davidson says, "Paradoxically, the one thing meanings do not seem to do is oil the wheels of a theory of meaning" (p. 20).

3. A TRUTH THEORY AS THE VEHICLE OF A COMPOSITIONAL MEANING THEORY

The basic difficulty in formulating a compositional meaning theory that allows one to derive formally all true instances of (*M*) is that it requires one to be able to quantify into the sentence on the right-hand side of 'means'. For one must connect axioms that attach to the components of the sentence described on the left with expressions in the metalanguage that are used on the right systematically, so that we can use axioms for the object language expressions to yield a matching, used sentence alike in meaning. If the metalanguage sentence on the right is only mentioned, then one could know the theory without knowing what the object language sentences mean, only that they were alike in meaning with others. Yet this apparently forces us to treat expressions that are not prima facie referring expressions as referring expressions, and this seems, even if we can sidestep the slingshot, an obvious mistake.

Davidson's makes his famous proposal in the following passage, which we quote in full:

> The theory will have done its work if it provides, for every sentence *s* in the language under study, a matching sentence (to replace '*p*' [in '*s* means *p*'] that, in some way yet to be made clear, 'gives the meaning' of *s*. One obvious candidate for matching sentence is just *s* itself, if the object language is contained in the metalanguage; otherwise a translation of *s* in the metalanguage. As a final bold step, let us try treating the position occupied by '*p*' extensionally: to implement this, sweep away the obscure 'means that', provide the sentence that replaces '*p*' with a proper sentential connective, and supply the description that replaces '*s*' with its own predicate. The plausible result is
>
> (*T*) *s* is *T* if and only if *p*.

> What we require of a theory of meaning for a language *L* is that without appeal to any (further) semantical notions it place enough restrictions on the predicate 'is *T*' to entail all sentences got from schema *T* when '*s*' is replaced by a structural description of a sentence of *L* and '*p*' by that sentence. (p. 23)

The central idea in this passage is that the practical value of having a theory that generates theorems of the form (*M*) is to match, in a way that reveals compositional structure, a mentioned object language sentence with a metalanguage sentence in use that means the same. There are difficulties in implementing this idea when the predicate we are dealing with is 'means'. Thus, Davidson suggests that we retain the idea that we want to match a metalanguage sentence in use with a mentioned object language sentence that is the same in meaning, in a way that reveals compositional structure, but that we do this by defining a predicate of the object language sentences in a way that generates instances of (*T*) in which '*p*' is replaced by a sentence the same in meaning as the object language sentence.

Davidson notes that a recursive characterization of a predicate 'is *T*' that meets this requirement (and is otherwise formally satisfactory) will have in its extension all and only the true sentences of the object language. The requirement that a formal characterization of the predicate 'is *T*' in (*T*) entail all instances of (*T*) in which '*s*' is replaced by a structural description of an object language sentence, and '*p*' by a metalanguage sentence that translates it, is in fact Tarski's Convention *T* on an adequate formal definition of a truth predicate for a language. Thus, the suggestion that emerges is that a truth theory that meets Tarski's Convention *T* (and perhaps certain other constraints) provides us with all the information that we wanted out of a compositional meaning theory.

For illustration, we introduce informally a simple recursive axiomatic truth theory. We first consider a theory without quantifiers or context-sensitive elements (we will extend this theory to context-sensitive languages later, and we will sketch the treatment for a language with quantifiers in the appendix to this chapter). Our predicates are '*x* was a Philistine' and '*x* slew *y*'; our proper names are 'David' and 'Goliath'; our connectives are 'and' and 'not'. We use '(' and ')' as grouping elements. We call our language '*L*'. We suppose, excepting tense, that our terms mean what they do in English. Our theory is given in A1–A6. A1 and A2 give the referents for the proper names of *L*. A3 and A4 give the conditions under which sentences formed from predicates and referring terms in *L* are true. A5 and A6 give, recursively, the conditions under which sentences formed using the logical

connectives are true in *L*.

A1. Ref('David') = David

A2. Ref('Goliath') = Goliath

A3. For any referring term *r*, *r*⌢' was a Philistine' is true-in-*L* iff ref(*r*) was a Philistine.

A4. For any referring terms r_1, r_2, r_1⌢' slew '⌢r_2 is true-in-*L* iff ref(r_1) slew ref(r_2).

A5. For any sentence *s*, 'it is not the case that '⌢*s* is true-in-*L* iff it is not the case that *s* is true-in-*L*.

A6. For any sentences s_1, s_2, '('⌢s_1⌢' and '⌢s_2⌢')' is true-in-*L* iff (s_1 is true-in-*L* and s_2 is true-in-*L*).

Allowing, as rules of inference, substitution of coreferring singular terms (counting 'ref(*a*)' as a singular term), universal quantifier instantiation, and a replacement schema (S(*T*) may be inferred from (i) S(*R*) and (ii) *R* iff *T*), we can derive from A1–A6, for example, (1) and (2).

(1) 'David'⌢' slew '⌢'Goliath' is true-in-*L* iff David slew Goliath.

(2) '('⌢'Goliath'⌢' was a Philistine'⌢' and '⌢'it is not the case that '⌢'David was a Philistine'⌢')' is true-in-*L* iff Goliath was a Philistine and it is not the case that David was a Philistine.

And so on. Given the rules of inference that we've introduced, we can see that any *T*-form theorem (any *T*-theorem) will intuitively draw only on the content of the axioms. Clearly, we can derive a *T*-theorem for each object language sentence. Since we start out with axioms in this theory that use metalanguage terms that are synonymous with the object language terms for which they are used to give truth conditions, our *T*-theorems use metalanguage sentences that are synonymous with the object language sentences for which they are used to give truth conditions. We will call such *T*-theorems '*T*-sentences'. Clearly, our theory meets the requirement that we wish to impose on an adequate definition: it entails all instances of the *T*-schema for the theory (T_L) in which '*s*' is replaced by a structural description of an object language sentence, and '*p*' is replaced by a sentence that translates it.

(T_L) *s* is true-in-*L* iff *p*

We can note further, as Davidson himself does at one point (Davidson 1984 [1970], p. 60), that given that the theory meets this requirement, we

can replace 'is true-in-*L* iff' with 'means in *L* that', *salva veritate*. Thus, knowing that the theory meets Convention *T* (excepting formalization), we can rewrite, (1), for example, as (3).

(3) 'David slew Goliath' means in *L* that David slew Goliath.

And so on. A *T*-sentence will be said to be *interpretive*, and to give *interpretive* truth conditions for its object language sentence. The axioms of such a theory, which use terms synonymous with object language terms when giving truth conditions, will be said to be interpretive, and the theory that has such axioms will be an interpretive truth theory. Such a theory may seem trivial, because we use a metalanguage that is the same as the object language, but, as before, we can see from imagining a metalanguage that differs from the object language that we gain real information about the object language from such a theory and what we know about it. These observations show the sense in which the present suggestion is a generalization to a truth theory of the observations that we made about our simple reference theory in the previous section. The lessons carry over straightforwardly. A truth theory, about which we know enough, serves to put us in a position to interpret any sentence in the object language. Moreover, by examining the proof of a *T*-theorem (in the general case, we will designate those proofs that intuitively draw only on the content of the axioms *canonical proofs*, and the theorems they prove *canonical theorems*) we can see how the parts of an object language sentence contribute systematically to its interpretive truth conditions, and thus gain insight into its compositional structure as that relates to the conditions under which it is true in virtue of what its constituents mean. This is the sense in which a truth theory can serve as the vehicle of a compositional meaning theory.[6]

It is important to note that the information that suffices to understand the object language is not all stated by the truth theory itself. Some of it is provided by things that the theorist knows about the truth theory. In the illustration just given, it is crucial that we know that the axioms of the theory are interpretive in order to use it for interpretation. Thus, we should not say that the truth theory itself is a meaning theory. If we identify the meaning theory with what it is we know that suffices to understand each sentence of the object language, then the meaning theory would be given by a statement of what we know about the truth theory that enables us to use it for interpretation (see Ludwig 2002 for further discussion).

In the example we have given, we have considered a language whose sentences have the same meaning on each occasion of use. The form of the theory must be changed if we are to extend it to a language that contains

elements whose contribution to what sentences are used to mean varies from context of use to context of use. Thus, if we understand 'slew' now as in the past tense, as in English, an utterance of 'David slew Goliath' on the morning of David's battle with Goliath means (roughly) that prior to that morning David had slain Goliath. But an utterance of 'David slew Goliath' on the evening of that day means that prior to that evening David had slain Goliath. Thus, the utterance in the morning is false, while that in the evening is true.

There are two basic ways of accommodating context-sensitive sentences in a truth theory. The first is to move from a theory that predicates truth of *sentences* to one that predicates truth of *utterances* (Weinstein 1974). This would require conditionalizing on the use of a sentence in performing a speech act. Thus, we might have a *T*-sentence such as (4).

(4) For any speech act u of assertion performed using 'David'⌢' slew '⌢'Goliath' in English, u is true iff for some time t earlier than u, slew(David, Goliath, t).

(We take 'slew(x, y, t)' to be a context-insensitive verb with an explicit argument place for time.) The second way of modifying a truth theory so as to extend it to a language containing context-sensitive elements (a context-sensitive language) is to add additional argument places to the truth predicate for contextual features relative to which are determined the contributions of context-sensitive elements in sentences. For present purposes, we will suppose that argument places for speaker and time will suffice. The form of a *T*-sentence then would be (5).

(5) For any speaker S, and any time t, 'David'⌢' slew '⌢'Goliath' is true(S, t, English) iff for some time t' earlier than t, slew(David, Goliath, t').

We will read the predicate 's is true(S, t, L)' as 's understood as if spoken by S at t in L is true'.[7] Corresponding to the context-relativized truth predicate we would have a context-relativized meaning predicate: 's means(S, t, L) that', read as 's understood as if spoken by S at t in L means that'. From the appropriate theorems of the theory we could "read off" meaning theorems, as before. The second of these approaches is easier to implement, so we will use it for purposes of illustration. But any workable theory on either approach could be reformulated in the framework of the other. Each of them takes as the basic truth bearer the speech act performed using a sentence. An axiom for a tensed predicate can be given as in (6) (see Lepore and Ludwig

2003 for further discussion of the semantics of tense from the standpoint of truth-theoretic semantics).

(6) For any speaker S, any time t, and any proper names n_1, n_2, n_1⌢‘ slew ’⌢n_2 is true(S, t, English) iff for some time t' earlier than t, slew(ref(n_1), ref(n_2), t').[8]

In addition to context-sensitive features of predicates such as tense, there are also context-sensitive referring terms in natural languages, such as ‘I’, ‘we’, ‘now’, ‘then’, ‘here’, ‘there’, ‘you’, ‘he’, ‘she’, ‘it’, ‘they’, ‘this’, ‘that’, ‘these’, ‘those’, and so on. An indexical term such as ‘I’ may be assigned a rule for determining its referent, as in (7).

(7) For any speaker S, and time t, the referent of ‘I’ as used by S at $t = S$.

Demonstrative terms, of which ‘this’ and ‘that’ are paradigms, introduce some special complications, because they admit of vacuous uses. One may intend to refer to something using a demonstrative but fail to do so – for example, when hallucinating something. To give a reference clause for a demonstrative, we must conditionalize on nonvacuous uses, as illustrated in (8).

(8) For any speaker S, any time t, and any object x, if S demonstrates x using ‘that’ at t, the referent of ‘that’ as used by S at $t = x$.

(Some additional refinements are needed, which propagate through the truth theory, but these are too involved to discuss here: see the appendix to Lepore and Ludwig 2000.) This means that the truth theory will not issue in fully specified truth conditions for sentences containing demonstratives except relative to actual contexts of use.

It remains to say how to generalize our adequacy condition, that is, how to reformulate Convention T for a context-sensitive language. Convention T says that an adequate definition of a truth predicate for a (context-insensitive) language L must be formally correct and must entail all instances of the T-schema in which ‘s’ is replaced by a structural description of a sentence of L and ‘p’ is replaced by a translation of s. Davidson has never provided a precise characterization of the parallel to Convention T for a context-sensitive language. But we can provide one in a straightforward way by noting first that we can *reformulate* Convention T as follows.

> An adequate truth theory for a context insensitive language L must be formally correct and entail for all sentences of the object language a theorem

of the form (*T*), where '*s*' is replaced by a structural description of an object language sentence,

(*T*) *s* is true in *L* iff *p*,

such that the result of replacing 'is true in *L* iff' with 'means in *L* that' yields a true sentence.

The two formulations are equivalent, because 'is true in *L* iff' can be replaced by 'means in *L* that' to yield a true sentence iff what replaces '*p*' translates *s*. We can generalize this to context-sensitive languages by replacing 'is true in *L* iff' and 'means in *L* that' with the corresponding context-relativized semantic predicates 'is true(*S*, *t*, *L*)' and 'means(*S*, *t*, *L*)'. For convenience, let us call this criterion of adequacy Convention *D*. We will want to impose a corresponding requirement on the axioms of the theory to ensure that our starting points are correct.

Again, the meaning theory itself may be considered to be a statement of the knowledge that we need to have about an appropriate truth theory in order to use it to interpret another speaker. The truth theory itself does not state everything we need to know. Davidson himself is explicit about this (see, e.g., Davidson 1984b [1973], p. 139; 1984 [1976], p. 172). Failure to notice this has led to a number of spurious criticisms and misunderstandings of truth theoretic semantics. See, in particular, the exchange in Foster 1976 and Davidson 1984 [1976], and more recent criticisms along the same lines in Richard 1992 and Soames 1989; 1992. See Ludwig 2002 for an explicit formulation of a meaning theory that exploits a truth theory, and for some comments on how this helps to disarm traditional objections to truth-theoretic semantics.

4. INTERPRETIVE ISSUES: REPLACEMENT OR NOVEL PURSUIT OF TRADITIONAL GOALS?

As we have presented it, the point of using a truth theory to pursue the goal of a compositional meaning theory is that, with no more resources than are required for the theory of reference, it enables us to exhibit the compositional structure of a sentence while providing a matching sentence in the language of the theory that, if we know the right things about the truth theory, we will know translates the object language sentence. This puts us in a position to interpret any potential utterance of a sentence of the object language. One thing we could know about the theory, as we've seen,

is that its axioms use appropriately expressions in the metalanguage that translate the object language expressions for which they are used to specify reference and truth conditions. Davidson does not appeal (in any direct way) to this as a constraint on a truth theory for it to "do duty" as a meaning theory. And that he does not – and what he does initially appeal to – has given rise to the view that when he introduces the suggestion that we should pursue a truth theory rather than a theory that assigns meanings to primitive expressions, he is advocating, not a way of working around the traditional bottlenecks, but a replacement of the traditional project. A clear example of this (once) widespread (but not eradicated) misunderstanding can be found in Stich 1976.

Davidson supplies some fuel for this particular flare-up of confusion in a number of passages. For example, he says this about his proposal in "Truth and Meaning":

> [T]he definition works by giving necessary and sufficient conditions for the truth of every sentence, and to give truth conditions is a way of giving the meaning of a sentence. To know the semantic concept of truth for a language is to know what it is for a sentence – any sentence – to be true, and this amounts, in one good sense we can give to the phrase, to understanding the language. . . . Indeed, since a Tarski-type truth definition supplies all we have asked so far of a theory of meaning, it is clear that such a theory falls comfortably within what Quine terms 'the theory of reference' as distinguished from what he terms the 'theory of meaning'. So much to the good for what I call a theory of meaning, and so much, perhaps, against my so calling it. (p. 24)

A reader might be forgiven for thinking, especially in light of the reference to Quine, that the object of introducing a truth theory in place of a meaning theory is to pursue our work where there is light rather than darkness, and that the suggestion is that the theory of meaning, with its obscure entities and logical difficulties, is a kind of confused, proto-scientific theory, a folk science of language that must be relegated to the status of the evil demon theory of disease.

It should be clear, though, in the light of our development of the point of introducing a truth theory, that this is a significant misunderstanding. This is most clearly seen from the emphasis that Davidson gives to the importance of a truth definition's meeting Convention *T*, or our analogue for a context-sensitive language, Convention *D*. This would make no sense if Davidson's aim were to eschew talk of meaning altogether. Numerous passages in essays later than "Truth and Meaning" bear this out.[9]

What has perhaps obscured this more than it ought to have is that in "Truth in Meaning" Davidson makes the suggestion – in cases where we are concerned with a truth theory for a natural language, which must accommodate the contributions of context-sensitive elements – that if a truth theory is extensionally adequate (that is, if it characterizes a predicate with all and only true sentences of the language in its extension), then its canonical theorems can be used to interpret object language sentences. The aim of this is to shed more light on the concept of meaning than could be expected from explaining constraints that the theory must meet in order to satisfy Convention *D* by appealing to, for example, the requirement that *axioms* use terms that *translate* object language terms. Thus, Davidson suggests:

> What appears to the right of the biconditional in sentences of the form '*s* is true if and only if *p*' when such sentences are consequences of a theory of truth plays its role in determining the meaning of *s* not by pretending synonymy but by adding one more brush-stroke to the picture which, taken as a whole, tells what there is to know of the meaning of *s*; this stroke is added by virtue of the fact that the sentence that replaces '*p*' is true if and only if *s* is.

Since it is not perhaps immediately clear how this would suffice for meeting Convention *D*, it may seem natural to suppose that the aim is rather to urge that there is nothing more to the idea of meaning than could be gleaned from a merely extensionally adequate truth theory. In the next section, we explain why Davidson initially hoped that extensional adequacy would prove adequate, and why it is not. In the section following, we will consider the proposal that he subsequently introduces, namely, that the truth theory be confirmable from the standpoint of a radical interpreter.

5. THE EXTENSIONALITY CONSTRAINT

Davidson suggests in "Truth and Meaning" that a merely extensionally adequate truth theory for a natural language would suffice for interpretation. This amounts to the claim that a merely extensionally adequate truth theory for a natural language would, ipso facto, satisfy Convention *D*. For a context-insensitive language, this is obviously inadequate. For example, a truth theory that issued in (9) and (10) as canonical theorems would be extensionally adequate, but its theorems would not satisfy Convention *T*.

(9) '*A* is a triangle' is true-in-*L* iff *A* is a trilateral.

(10) '*A* is a trilateral' is true-in-*L* iff *A* is a triangle.

Indeed, this example shows that even if the theory issues in theorems that are necessarily true (taking a language to be individuated by its syntax and semantics), it is not guaranteed to meet Convention *T*. Why expect that a theory for a context-sensitive language would fare any better?

Davidson's hope was apparently that the context-sensitive elements of the language would provide the needed additional refinement. "Sentences with demonstratives obviously yield a very sensitive test of the correctness of a theory of meaning, and constitute the most direct link between language and the recurrent macroscopic objects of human interest and attention" (Davidson 1984 [1967], p. 35). This is also indicated in a retrospective footnote to "Truth and Meaning" (Davidson 1984 [1967], note 10). For example, could a theory that had to deal with demonstratives issue in a theorem such as (*S*)?

(*S*) 'Snow is white' is true in English iff grass is green.

It would have to also yield correct theorems for 'That is snow', 'That is grass', 'That is white', and 'That is green'. If our axiom for the demonstrative is the one given in §3, then it looks as if a theory that yielded (*S*) would require (11) and (12) as axioms[10] (suppressing quantification over '*S*' and '*t*').

(11) '*x* is white' is true(*S*, *t*, *L*) of something iff it is(*t*) green.

(12) '*x* is snow' is true(*S*, *t*, *L*) of something iff it is(*t*) grass.

But then if someone demonstrates a bit of snow, σ, and calls it snow, the combination of our reference axiom and our predicate axioms will give the wrong result, namely, (13).

(13) 'That is snow' is true(*S*, *t*, *L*) iff σ is(*t*) grass.

Thus, it may appear that the presence of such elements as demonstratives in the language will provide enough additional resolving power to rule out spurious theories.

Reflection shows, however, that this is not enough. As was observed by a number of critics of this initial suggestion, if we replace any predicate axiom with another that uses a predicate in giving truth conditions that is extensionally equivalent to the original but nonsynonymous, we will have a truth theory that is extensionally adequate if the original was (Foster 1976; Loar 1976). But it cannot be that both of the theories are interpretive. In fact, the examples with which we started this section, (9) and (10), show this already. No test involving demonstratives will show either of those axioms

to be inadequate. Moreover, even in the case of predicates such as 'is snow' and 'is white', the test involving demonstratives works only if we pair them with axioms for demonstratives that are themselves interpretive. But if we are testing at the level of *T*-theorems for the adequacy and truth of the theory, then we cannot help ourselves to correct reference axioms. Thus, extensional adequacy fails to ensure that a truth theory satisfies Convention *T*, even for a natural language. This is something that became apparent quickly. But it is important to note, both in order to understand how Davidson's aim could have been misunderstood and in order to understand Davidson's next suggestion.

6. THE ROLE OF RADICAL INTERPRETATION

Davidson's initial suggestion, that a merely extensionally adequate truth theory for a language would ipso facto be interpretive, was incorrect. The aim of that suggestion was to identify constraints that a truth theory could meet that were not couched in terms of meaning, with the aim of shedding light indirectly on what was involved in understanding another speaker. If extensional adequacy is not enough, the question arises what additional (illuminating) constraints must be placed on a truth theory for a natural language for it to be interpretive.

Davidson returned to this question in "Radical Interpretation" (Davidson 1984b [1973]), which he characterized in "Reply to Foster" (Davidson 1984 [1976], p. 171) as an attempt to say better what the relation is between a truth theory and a meaning theory.[11] The project of radical interpretation is treated in some detail in Chapter 3. In this section, our aim is just to explain how it is related to the project of using a truth theory in pursuit of a meaning theory for a speaker's language.

At the beginning of "Radical Interpretation," Davidson poses two questions: (1) What is it that we could know that would enable us to interpret other speakers?[12] (2) How could we come to know it? The second question is to be answered on the basis of evidence that includes knowledge of a speaker's actual and potential behavior (his behavioral dispositions), but which excludes knowledge of the meaning of any of his terms and any detailed knowledge of his propositional attitudes. This is what defines the position of the radical interpretater, and so the project of radical interpretation. The goal of a theoretical description of the procedures of radical interpretation is to shed light on the concepts we use in interpreting other speakers – concepts of meaning, truth, reference, rationality,

of the propositional attitudes, action, preference, and so on – by relating the theoretical structure that contains them to evidence for their application that is described independently of their use (Davidson 1984b [1973], p. 137).

Our present concern is specifically with question (1). For Davidson's suggestion is that what we could know that would enable us to interpret another speaker is a truth theory for his language, and that it met certain constraints. In "Truth and Meaning," the suggestion was that the appropriate constraint was that the theory be simply extensionally adequate (at the level of *T*-theorems). What is substituted for this constraint in "Radical Interpretation"? The suggestion is the following:

> [T]he totality of *T*-sentences[13] should (in the sense described above) optimally fit evidence about sentences held true by native speakers. The present idea is that what Tarski assumed outright for each *T*-sentence can be indirectly elicited by a holistic constraint. If that constraint is adequate, each T-sentence will in fact yield an acceptable interpretation. (p. 139)

For a theory to optimally fit evidence "in the sense described above" is for it to be confirmable using the procedures Davidson outlines for the radical interpreter. Thus, Davidson's constraint appears to be that the truth theory have been confirmed by the procedures of the radical interpreter. This is supposed to impose a stricter requirement on the theory than simply that it be true, because, for example, if the theory is confirmed empirically, and is about, ultimately – as Davidson holds – an individual speaker's idiolect, then its theorems must be lawlike (see in particular the retrospective note 11 in Davidson 1984b [1973]; Davidson 1984 [1976], p. 174; and Davidson 1999f, p. 688).

It is not clear that this added constraint is adequate. There is, first of all, some difficulty in taking this constraint to be one that could serve as an answer to question (1). For if it does, that a theory is confirmed or optimally fits evidence becomes the thing that we want the radical interpreter to confirm; but that is clearly not the intent. But put this aside. What property does a theory that is confirmed by a radical interpreter have that guarantees that it meets Convention *T*? One property that we know it will have is being projectible and lawlike. Davidson sometimes seems to suggest that this is what he has in mind. But this is not, by itself, adequate. The example given at the beginning of §5 shows this. What we need, then, is some reason to think that any theory of truth that optimally fits evidence in the form of a speaker's behavior will ipso facto be interpretive. If we had an a priori guarantee that speakers were interpretable from the standpoint of a

radical interpreter, this would provide the grounding required. Davidson does offer some arguments for this, though discussion of them is beyond the scope of this chapter (Davidson 1984 [1976]; 1989; 1990d; 2001a [1982]; 2001 [1983]; 2001b [1988]; 2001 [1989]; 2001b [1991]; 2001 [1992]; 2001 [1999]). See Chapter 3 for further discussion, as well as Lepore and Ludwig forthcoming; Ludwig 1992; 1994; 1996.

7. WORK IN TRUTH-THEORETIC SEMANTICS

It is important to note that the project of pursuing compositional semantics by way of a truth theory (truth-theoretic semantics) can proceed independently of this more ambitious project in which Davidson embeds it. We have stated a constraint on a truth theory that suffices for it to serve in pursuit of a compositional meaning theory. That is basically that its axioms meet a suitable analogue of Convention *D*. Thus, proposals for the semantic form of natural language constructions can be cast in the form of proposals for axioms in an interpretive truth theory for the language. The theory can then be tested against intuitions about entailment relations based on formal considerations, and about its systematic implications for constructions in which the relevant words and constructions appear.

It is worth noting in this connection that pursuit of a compositional meaning theory through the vehicle of a truth theory, as indicated in §3, is completely neutral on what the proper analysis of the concept of truth is, beyond the requirement that any coherent definition of a truth predicate that honors the core concept must meet Convention *T*, or Convention *D*, as the case may be. In particular, a deflationary account of the concept of truth is not a threat to truth-theoretic semantics. Truth-theoretic semantics uses the concept of truth in a formal structure. It makes no claim to reduce the concept of meaning to the concept of truth and other concepts. Worries about the deployment of the concept of truth could arise only in a context in which one was concerned to illuminate the concept of meaning by relating it to other, presumably independently grasped concepts, or concepts grasp of which is coordinate with that of meaning.

The program of truth-theoretic semantics, taken independently of Davidson's larger philosophical concerns, has been pursued extensively in philosophy and linguistics.[14] In the remainder of this section, we identify some problem areas for the program of truth-theoretic semantics for natural languages, and indicate where work has been done on them, by Davidson and others. Among the outstanding problems that Davidson identified

were the treatment of context-sensitive terms, counterfactual or subjunctive conditionals, mass terms, adverbs, attributive adjectives and adverbs such as 'slow' and 'slowly', the problem of opaque contexts (for example, in indirect discourse sentences), quotation, and the problem of extending the program to those sentences of the language that, prima facie, are neither truth nor false – namely, interrogatives, such as 'What time is it?' and imperatives, such as 'Put on your hat'. Davidson has contributed to work on a number of these problems. Some additional problem areas are restricted quantifiers (e.g., 'Most philosophers are not rich'), so-called branching and cumulative quantifiers (e.g., 'Most men and most women like each other' and 'Ten firms employed twenty engineers', which are alleged to have readings that cannot be captured with linear quantifiers), and the related problem of plurals in English, which are said to require second-order quantification.

In "The Logical Form of Action Sentences" (Davidson 1980b [1967]), Davidson made a very influential suggestion about how to understand adverbial modification of action verbs (a suggestion that generalizes to the treatment of adverbs generally). The suggestion is that we represent action sentences as having an implicit quantifier over events, and adverbial modification as contributing predicates of the event variable thus introduced. Thus, for example, the logical form of (14) is represented as (15).

(14) David slew Goliath with a sling.

(15) There is an event *e* such that *e* is slaying *and e* was by David *and e* was of Goliath *and e* was done with a sling.

This account of adverbial modification is now widely accepted in philosophy and linguistics (see, e.g., Schein 1993; 2002). This contribution, along with related issues connected with the metaphysics of events, is discussed further in Chapter 5.

Davidson's interesting but controversial solution to the problem of indirect discourse is given in "On Saying That" (Davidson 1984 [1968]). It has come to be called the paratactic account,[15] and it has been extended to other contexts in which words cannot be freely intersubstituted solely on the basis of their referents, or extensions. The account, in brief, treats an utterance of what we would write in English as (16) as semantically functioning like the utterance of two sentences, the utterance of the first of which contains a demonstrative reference to the utterance of the second, as in (17).

(16) Galileo said that the Earth moves.

(17) Galileo said that. The Earth moves.

The first sentence is analyzed as on the event analysis of action sentences, and the verb 'said' is treated as relating two utterances: the relation holds between the utterances if and only if they are the same-in-content (or perhaps relevantly-similar-in-content). Introducing 'samesays' to express the relevant relation, the suggestion can be written as in (18).

(18) There is an event *e* such that *e* was an utterance of Galileo's and *e* samesays with that. The Earth moves.

Among the difficulties that this account has been charged with is that it cannot accommodate quantification into the complement clause of an indirect discourse sentence, as in 'Everyone said that he hadn't seen it' (Higginbotham 1986), since it does not treat 'he hadn't seen it' as part of the sentence containing the quantifier 'Everyone', and that it yields the wrong answer to the question how many things Galileo said (for example), since it treats 'says' as relating Galileo not to a proposition, but to utterances that samesay his – of which there are many (McFetridge 1976). Discussion of this seminal proposal and additional references to the literature can be found in Lepore and Loewer 1989 and Ludwig and Ray 1998.

Davidson extended the paratactic approach to two other problems on our earlier list. The first is the problem of quotation. Quotation is a device for referring to expressions by way of a sample of the expression referred to. The difficulty with quotation is that it appears to be a productive device. We can quote any expression in order to refer to it. So there appear to be an infinity of "quotation names." Yet quotation names appear not to have internal semantic structure. For the word that appears inside the quotation name, e.g., 'duck', in " 'duck' ", does not contribute its semantic properties to determining the referent of the term. In "Quotation" (Davidson 1984c [1979]), Davidson proposes a solution similar to the paratactic account of indirect discourse. The account holds that the semantic form of quotation involves a demonstrative description of an expression type that demonstrates a token of the type to be referred to. The contribution of the quotation marks in (19) is represented as in (20).

(19) 'duck' is a four-letter word.

(20) The expression of which this is a token is a four-letter word. duck

One criticism that has been leveled against the paratactic account of quotation is that there is nothing in it that constrains the demonstrative to refer to the word a token of which is displayed between the quotation marks. An alternative account that treats quotation names as simple referring terms,

syntactically productive but not semantically productive, was given by Wallace (1970), who suggested the following simple reference clause.[16]

(21) For any expression *E*, the expression resulting from placing *E* in quotation marks refers to *E*.

One perhaps unsettling feature of this proposal is that the class of expressions in English consisting of quotation marks around expressions is infinite but not recursively specifiable (i.e., it cannot be constructed from a finite number of primitives and rules). The paratactic account of quotation has been elaborated and defended in Cappelen and Lepore 1997; 1999a; 1999b.

The final application of the paratactic strategy is to the problem of nondeclarative sentences. The difficulty for truth-theoretic semantics presented by nondeclaratives is that they do not, on the face of it, have truth conditions, so that the truth theory appears to be an inappropriate vehicle for specifying their meanings in a context. Davidson's response is to try to assimilate nondeclaratives to declaratives with the aid of a paratactic/demonstrative analysis of the nondeclarative mood markers. This idea is suggested by thinking about the explicit application of the paratactic account of indirect discourse to performative sentences, such as 'I hereby command that you put on your hat'. The basic idea can be illustrated by an imperative such as 'Put on your hat', which Davidson suggests we treat *semantically* as two utterances, one of a mood-setter (playing the role of 'I command that'), which says of something that it is an utterance with a certain (illocutionary) force, and the other as what the mood marker is directed toward. Davidson says, "If we were to represent in linear form the utterance of, say, the imperative sentence 'Put on your hat', it would come out as the utterance of a sentence like 'My next utterance is imperatival in force', followed by an utterance of 'You will put on your hat' " (Davidson 1984b [1979], p. 120). This is not quite right, because, as Davidson says, "it gets the syntax wrong" (p. 120). Davidson suggests that we give truth conditions for the mood-setter as follows: "The mood-setter of an utterance of 'Put on your hat' is true if and only if the utterance of the indicative core is imperatival in force" (p. 120). The indicative core is presumably 'You will put on your hat'. But the difficulty with this is that there is no utterance of 'You will put on your hat' when someone utters 'Put on your hat'. It is not clear how this difficulty is to be overcome in a way that is consonant with Davidson's intentions. One might suggest that it be put (roughly) this way: the mood-setter of an utterance of 'Put on your hat' is true iff in uttering 'Put on your hat' the speaker directs that his auditor put on his hat. But

this does not represent 'Put on your hat' semantically as two utterances, and this turns out to be important for Davidson's explanation of why utterances of imperatives are intuitively judged not to have truth values. His explanation is that an utterance of an imperative is semantically not a single sentence but two, and that we would not judge the complex utterance of 'You are tired', 'You are old', for example, as true or false. There are other difficulties with this approach as well. It is not clear that we have the intuition that in uttering 'Put on your hat' we are saying two things that are truth valued. The account would also need to be extended, in some as yet unexplained way, to accommodate interrogatives whose "indicative cores" are open sentences, such as 'What time is it?' The account also has difficulty handling conditional imperatives and interrogatives, such as 'If you see her, say hello'. No simple directive is issued in uttering this conditional sentence, but Davidson's account would require the mood-setter to cover either the whole conditional or the consequent only, and in either case, the idea that a *conditional* requirement has been issued will be lost.

This does not show, however, that the truth-theoretic approach to semantics cannot be extended to cover nondeclaratives. Some recent work on nondeclaratives in the truth-theoretic framework shows how to generalize the approach to a fulfillment theory of sentences that assigns distinctive kinds of fulfillment conditions to declaratives, imperatives, and interrogatives (Ludwig 1997). Imperatives and interrogatives are treated as having compliance conditions, and declaratives as having truth conditions. The compliance conditions of imperatives and interrogatives, while not truth conditions, are nonetheless recursively characterized ultimately in terms of a truth theory for the language. We provide a quick sketch for 'Put on your hat'.

> 'Put on your hat' is fulfilled(S, t, L) iff ref(S, t, 'you') makes it the case that 'You will put on your hat' with the intention of fulfilling the speech act performed in uttering 'Put on your hat'.

There are additional complications for interrogatives. The approach can be extended to handle quantification across mood-setters, as in 'Invest every penny you earn' (for every x such that x is a penny you earn, invest x).

In closing this section, we briefly give some pointers to recent literature treating some of the other problem areas we have mentioned. (This is by no means intended to be exhaustive.) For recent work on tense and demonstratives, see Higginbotham 1995; Lepore and Ludwig 2000; 2003. For recent work on mass terms, see Koslicki 1999. We show in the

appendix to this chapter how to handle restricted quantifiers in the truth-theoretic framework. See Schein 1993 for important work on the semantics of plurals and so-called branching and cumulative quantifiers in natural languages.[17]

8. CONCLUSION

Davidson's work in the theory of meaning has had great influence on contemporary philosophy of language and on natural language semantics. His central proposal was that the task of understanding "what it is for words to mean what they do" can be fruitfully approached by considering how we could confirm a truth theory for a speaker on the basis of evidence that does not initially presuppose any knowledge of the speaker's meanings or the detailed contents of his propositional attitudes. There are two aspects of this proposal, which are separable. The first is the suggestion that a truth theory can serve in pursuit of a compositional meaning theory for a language essentially by providing the recursion needed to generate a used metalanguage sentence, to match with each object language sentence, which translates the object language sentence. This suggestion has launched a program in natural language semantics that takes the interpretive truth theory as its basic vehicle. In pursuing this project for languages that we know, we can make use of our knowledge of the language in formulating axioms. The second feature of the proposal is that, having a vehicle for a compositional meaning theory, we can gain further insight into the concepts of the theory and related concepts by relating them systematically to "neutral" evidence for their application.

APPENDIX: EXTENSION TO LANGUAGES WITH QUANTIFIERS

We sketch briefly in this appendix an extension of the truth theory presented in §3 to a language with quantifiers. Quantifiers require us to introduce variables in argument places of predicates that can be bound by them, as in

(Some *x*)(some *y*)(*x* slew *y*).

The standard practice is to introduce sequences of objects or functions from variables to objects as "satisfiers" of open sentences in which argument places are occupied by variables, and to define truth in terms of satisfaction

by all sequences or functions. A predicate satisfaction clause would go as follows:

> For all functions *f*, variables *v*, *u*, *f* satisfies-in-*L* *v*⌢‘ slew ’⌢*u* iff f(*v*) slew f(*u*).

This mimics our clause for proper names. Indeed, we can think of each function as an extension of the reference function for the language to the variables treated as names. Intuitively, ‘satisfies’ is the inverse of ‘is true of’. The recursive clauses for connectives are treated in the same way as those for truth, as in

> For all functions *f*, for any sentences *s*, *r*, *f* satisfies-in-*L* ‘(’⌢*s*⌢‘ and ’⌢*r*⌢‘)’ iff (*f* satisfies-in-*L* *s* and *f* satisfies-in-*L* *r*).

Clauses for quantifiers look at variants of a given function (or sequence) with respect to the variable bound by the quantifier. For ‘Some *x*’ and ‘Every *x*’, we have:

> For all functions *f*, for all formulas *F*, variables *v*, *f* satisfies-in-*L* ‘(some *x*)’⌢*F* iff some *v*-variant *f* ′ of *f* is such that *f* ′ satisfies-in-*L* *F*.
>
> For all functions *f*, for all formulas *F*, variables *v*, *f* satisfies-in-*L* ‘(Every *x*)’⌢*F* iff every *v*-variant *f* ′ of *f* is such that *f* ′ satisfies-in-*L* *F*.

We define ‘*f* ′ is a *v*-variant of *f* ’, as ‘*f* ′ differs from *f* at most in what it assigns to *v*’. As for other sorts of expressions, we use in the metalanguage quantifiers the same in meaning as the object language quantifiers they are used to give satisfaction conditions for. These clauses are easily extended to restricted quantifiers, such as ‘(Every *x*: *x* is a man)’ (corresponding in English to ‘Every man’):

> For all functions *f*, for all formulas *F*, variables *v*, *f* satisfies-in-*L* ‘(Every *x*: *x* is *G*)’⌢*F* iff every *v*-variant *f* ′ of *f* which satisfies ‘*x* is *G*’ is such that *f* ′ satisfies-in-*L* *F*.

If one function satisfies a formula, every function will. For a closed sentence, we define truth as satisfaction by all functions.

> For any sentence *s*, *s* is true iff for all functions *f*, *f* satisfies *s*.

The truth and satisfaction predicates can then be relativized to contextual parameters for context-sensitive languages.

Notes

1. Frege, for example, clearly presupposes it in his account of meaning, and discusses the importance of this in the context of recognizing the meaning of novel sentences, in particular, in a letter to Jourdain in January 1914 (Beaney 1997, pp. 319–20).
2. Davidson attributes the argument to Frege, and versions of it can be found in Church 1943; 1956 and Gödel 1966. See Neale 1995; Neale and Dever 1997; and Oppy 1997 for recent discussions.
3. Davidson has used versions of this argument against fact-based ontologies in "True to the Facts" (Davidson 1984 [1969], p. 42) and "The Logical Form of Action Sentences" (Davidson 1980b [1967], pp. 117–18), and against treating sentences as referring to events in Davidson 1980 [1969], p. 169.
4. That is, treating the definite article, 'the', and '=' as logical terms, but not the proper name, it is clear that the denotation of the definite description varies with the assignment to the name.
5. Space prevents a full discussion of other methods of achieving the end that might be scouted, including, e.g., appeal to substitutional quantification. For further discussion, see Lepore and Ludwig forthcoming, Chapter 3; Lepore and Ludwig n.d.; Ludwig 2002.
6. Church (1951, p. 102) seems to make essentially the same observation about Tarski's truth definitions, as Wallace (1978, p. 54) notes.
7. In this we go beyond anything Davidson has said. Davidson's only sketch of the form of a context relativized *T*-sentence that we are aware of can be found in "Truth and Meaning" (Davidson 1984 [1967], p. 34). This way of explaining how to understand the relativized truth predicate is intended to avoid the difficulties canvassed by Evans (1985, pp. 359–60).
8. Note that this requires a metalanguage that has for each tensed verb in the object language a corresponding untensed verb in the metalanguage, with an explicit argument place for time intervals. Thus, for a language such as English that does not have such verbs, the truth theory cannot be given in English, but must be given in a language which at least extends English by the introduction of these additional untensed predicates.
9. In "Belief and the Basis of Meaning" (Davidson 1984a [1974]), Davidson says: "A theory of truth will yield interpretations only if its *T*-sentences state truth conditions in terms that may be treated as 'giving the meaning' of object language sentences. Our problem is to find constraints on a theory strong enough to guarantee that it can be used for interpretation" (p. 150). See also Davidson 1984 [1976], pp. 173, 175; and Davidson 1984b [1977], p. 224.
10. Note that these axioms are themselves false. So Davidson's idea must have been not that a true theory could issue in (S), but that testing at the level of *T*-theorems would ensure that the axioms of the theory were true.
11. Even in "Truth and Meaning" (Davidson 1984 [1967], p. 27), Davidson regarded a theory of truth of the sort he was concerned with as an empirical theory, and argued that we gain insight into the terms of the theory by reflecting on how

it would be confirmed from the standpoint of Quine's radical translator (Quine 1960, Chapter 2). What he did not do there was to explicate the constraints a truth theory was to meet in order to serve as a meaning theory by appeal to empirical confirmation.

12. It is important to note the counterfactual element in this question. Davidson has never claimed that our competence in our languages is constituted by propositional knowledge of an explicit compositional meaning theory or truth theory for the language. The theory rather aims to "capture" in its structure the structure of a complex practical ability, the ability to speak and understand.
13. Here Davidson uses '*T*-sentence' in the sense in which we have used '*T*-theorem'.
14. See, e.g., Larson and Segal 1995, which aims to integrate the Davidsonian approach with standard syntactic descriptions of English grammar, though it also departs in some ways from Davidson's program in assigning semantic "values" to expressions besides grammatically referring terms. An important earlier synoptic work is Davies 1981. Of interest also are Davidson and Harman 1972; 1975; Evans and McDowell 1976; Lepore 1986; Lepore and McLaughlin 1985; Platts 1980; 1997.
15. Parataxis is the placing of propositions or clauses one after another, without indicating by connecting words the relations (of coordination or subordination) between them.
16. Unfortunately, Davidson himself does not explicitly address Wallace's suggestion, though he must have been familiar with it at the time of writing "Quotation."
17. The semantic paradoxes in natural languages have also been thought to be a difficulty for truth-theoretic semantics (see Chihara 1976). The *T*-sentence for a Liar sentence (*L*), '*L* is not true', is '"*L* is not true" is true iff *L* is not true', which by substitution (*L* = '*L* is not true') generates a contradiction, '*L* is true iff *L* is not true'. Thus, a truth theory for a natural language that respects Convention *T* (or *D*) is not true. A similar problem arises for vague predicates (such as 'is bald') that introduce truth value gaps in the language when applied to "borderline" cases, for a *T*-sentence for an object language sentence must use a vague predicate in the metalanguage. So the truth theory inherits the truth value gaps of the object language. Davidson addresses this briefly in "Truth and Meaning" (Davidson 1984 [1967], pp. 28–9), though his remarks are cryptic. About the paradoxes, one can point out that excising the semantic terms from a language would still allow us to give a semantics for most of its vocabulary. About both the semantic paradoxes and vagueness, one can remark that for the truth theory to discharge its job of informing us about the compositional structure of natural language sentences and informing us about what sentences in the object language mean, it is enough that we understand it, can prove canonical theorems for each object language sentence on the basis of interpretive axioms, and can trace through those proofs the systematic contribution, in virtue of meaning, of each of the component expressions of a sentence to its truth conditions. The purposes of understanding are served if we can do this: it is not further required that we endorse as true every sentence of the theory. One way to see this is to

notice that once we make the transition from *T*-sentences to *M*-sentences, we can treat the *M*-sentences as true even if the corresponding *T*-sentences are not. In effect, what we know about a truth theory that enables us to use it for understanding another speaker does not require that we know that the theory is in fact true. (See Lepore and Ludwig forthcoming, Chapter 10; and Ludwig 2002 for a fuller discussion of this issue.)

2 Philosophy of Action

ALFRED R. MELE

The basic subject matter of the philosophy of action is a pair of questions: (1) What are actions? (2) How are actions to be explained? The questions call, respectively, for a theory of the nature of action and a theory of the explanation of actions. Donald Davidson has articulated and defended influential answers to both questions. Those answers are the primary focus of this chapter.

1. ACTIONS AND INDIVIDUATION

Actions, as Davidson understands them, are analogous to money and sunburns in one noteworthy respect. The piece of paper with which I just purchased a drink is a genuine U.S. dollar bill partly in virtue of its having been produced (in the right way) by the U.S. Treasury Department. The burn on my back is a sunburn partly in virtue of its having been produced by exposure to the sun's rays. A duplicate bill produced with plates and paper stolen from the Treasury Department is a counterfeit dollar bill, not a genuine one. A burn that looks and feels just like the one on my back is not a sunburn if it was produced by exposure to a heat lamp rather than to the sun. Similarly, on Davidson's view of action, a certain event is my buying a drink – an action – partly in virtue of its having been appropriately produced by reasons that I had for buying one, reasons being understood as complexes of beliefs and desires (Davidson 1980 [1963]; 1980a [1971]; 1980a [1973]; 1987b). An event that someone else covertly produces by remote control – one including visually indistinguishable bodily motions not appropriately produced by relevant reasons of mine – is not a purchasing of a drink by me, even if it feels to me as though I am in charge.

To forestall confusion, it should be noted that Davidson does not identify actions with *non-actional* events appropriately caused by reasons. That would be analogous to identifying genuine U.S. dollar bills with pieces of printed

paper that are not genuine U.S. dollar bills and are produced in the right way by the U.S. Treasury Department, and so identifying genuine U.S. dollar bills would be absurd. To say that an event *E* is an action partly in virtue of its having been appropriately produced by reasons is not to say that *E* is a non-actional event – any more than to say that a piece of printed paper *P* is a genuine U.S. dollar bill partly in virtue of its having been produced in the right way is to say that *P* is not a genuine U.S. dollar bill.

The question "What are actions?" directly raises two others: How are actions different from events that are not actions?[1] How do actions differ from one another? I have just provided a crude sketch of Davidson's answer to the first question. Actions differ from events that are not actions in their causal history. Events that are actions are produced by reasons in a way that I will discuss in §3. Events that are not actions lack a causal history of this kind. Alternative conceptions of action include an "internalist" position, according to which actions differ experientially from other events in a way that is essentially independent of how, or whether, they are caused (Ginet 1990); a conception of actions as composites of non-actional mental events or states (e.g., intentions) and pertinent non-actional effects (e.g., an arm's rising) (Mill 1961; Searle 1983); and views identifying an action with the causing of a suitable non-actional product by appropriate non-actional mental events or states (Dretske 1988) – or, instead, by an agent (O'Connor 1995).

Davidson's answer to the question of how actions differ from one another – the question of act individuation – may be introduced by means of an example of his: "I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home" (Davidson 1980a, p. 4). How many actions has the agent performed? Davidson's answer is one action "of which four descriptions have been given" (Davidson 1980a, p. 4; cf. Anscombe 1957). A fine-grained alternative to Davidson's coarse-grained view treats *A* and *B* as different actions if, in performing them, the agent exemplifies different act-properties (Goldman 1970). On this view, our agent performs at least four actions in this case, since the act-properties at issue are distinct. For example, the property of flipping a switch is distinct from the property of turning on a light, and the property of turning on a light (in a room) is distinct from the property of illuminating a room. One may flip a switch without turning on a light, and vice versa. Similarly, one may turn on a light in a room without illuminating the room (the light may be painted black), and one may illuminate a room without turning on a light (by setting a dark room on fire). Another alternative – a componential one – represents our agent's

illuminating the room as an action having various components, including his moving his arm, his flipping the switch, and the light's going on (Ginet 1990; Thalberg 1977; Thomson 1977). Where proponents of the coarse-grained and fine-grained theories find, respectively, a single action under different descriptions and a collection of intimately related actions, advocates of the various componential views locate a "larger" action having "smaller" actions among its parts.

In the preceding scenario, things seem to happen in a flash. Consider a scenario with a different look. On Monday, the queen secretly pours poison into the king's ear. "She has done her work; it only remains for the poison to do its" (Davidson 1980a [1971], p. 58). The poison's work takes three days. On Davidson's view, the queen performed a single action describable in a variety of ways, including "pouring poison into the king's ear," "killing the king," and "moving her hand." When did that action take place? On Monday, three days before the king died. To some readers, this answer will sound odd, but that is something that Davidson believes we must learn to live with. "Our primitive actions, the ones we do not do by doing something else, mere movements of the body – these are all the actions there are" (p. 59).[2] These actions occur when we move our bodies, and our bodily movements can have consequences, including temporally remote ones, that licence descriptions that may seem to suggest – misleadingly so, as Davidson sees it – that some of our actions are very lengthy events (Davidson 1980a [1971]; 1985a, pp. 236–9). On Davidson's view of act individuation, actions that are intentional under some descriptions are unintentional under others (Davidson 1970; 1980 [1963]; 1980b [1967]; 1980a [1971]; 1980a [1973]; 1980 [1978]; 1985a). For example, his switch flipper performs an action that (presumably) is intentional under the descriptions "flipping the switch," "turning on the light," and "illuminating the room," but unintentional under the description "alerting the prowler." *Every* action is intentional under some description, according to Davidson (cf. Hornsby 1980). He contends that "a man is the agent of an act if what he does can be described under an aspect that makes it intentional" (Davidson 1980a [1971], p. 46) and that "action ... require[s] that what the agent does is intentional under some description" (p. 50). Putting these remarks together, we get the thesis that *x* is an action if and only if *x* is intentional under some description. Davidson expresses the point differently: "a person is the agent of an event if and only if there is a description of what he did that makes true a sentence that says he did it intentionally" (p. 46).

Davidson claims that "it is (logically) impossible to perform an intentional action without some appropriate reason" (Davidson 1980a [1976],

p. 264). A more precise formulation is the following: necessarily, every intentional action is performed for a reason. In conjunction with the thesis that every action is intentional under some description, this entails that every action is performed for a reason. Of course, an action done for a reason under one description need not be done for a reason under all proper descriptions (Davidson 1980 [1963], p. 5). Our switch flipper did not alert the prowler for a reason. I return to the connection between reasons and actions in §3.

2. DAVIDSON'S REBUTTAL OF ANTICAUSALIST ARGUMENTS AND HIS CHALLENGE

According to causal theories of action, an event's being an action depends upon how it was caused. Familiar causal theories of action feature as causes such psychological items as beliefs, desires, and intentions, or associated events – for example, the acquisition of an intention to do something straightaway (Bishop 1989; Brand 1984; Davidson 1980a; Goldman 1970; Mele 1992c). The idea is at least as old as Aristotle: "The origin of action – its efficient, not its final cause – is choice, and that of choice is desire and reasoning with a view to an end" (*Nicomachean Ethics* 1139a31–32). If causal theories are on the right track, they help both with a metaphysical issue and with an explanatory issue in the philosophy of mind. The metaphysical issue is how actions differ from nonactions. The explanatory issue is how actions are to be explained. Davidson contends that they are to be explained – causally – in terms of reasons, that is, complexes of beliefs and desires (see §3). If actions essentially have reasons as causes, Davidson's theory of action explanation has metaphysical underpinnings.

Owing in part to the influence of Wittgenstein and Ryle, causal theories about how actions are to be explained – specifically, theories of this kind framed in terms of the attitudes – fell into philosophical disfavor for a time. The first major source of their revival was Davidson's "Actions, Reasons, and Causes" (Davidson 1980 [1963]). In that paper, Davidson rebuts a collection of arguments against the causal approach. I take up three of these arguments.

The "logical connection argument" hinges on the premise that cause and effect must be "logically distinct." Because there is a logical or conceptual connection between an agent's having a reason (or wanting, intending) to *A* and her *A*-ing, the latter cannot be an effect of the former; or so it was claimed. Two decades after the publication of "Actions, Reasons, and

Causes," Norman Malcolm advanced a version of this argument that features intentions. There is, Malcolm argues,

> a logical connection between *intending* to do something and *doing* it. If doing it is well within the person's powers, and if he has not given up the intention for some reason or other, and if he has not forgotten his intention, and if no countervailing circumstances have arisen, and if he offers no satisfactory explanation for not fulfilling that intention, and so on – then if he doesn't do the thing, we would conclude that he does not really intend to do it. This way of judging the matter is *required by the concept of intention*. (Armstrong and Malcolm 1984, p. 88)

He adds: "the logical bond, the conceptual connection between intending and doing, is a loose one; nevertheless it is strong enough to rule out the possibility of there being a merely contingent connection between intending and doing." Consequently, granting that any causal connection is a "merely contingent" one, there is no causal connection between intending and doing.

Davidson's reply to arguments of this kind is incisive: causation is a relation between events, no matter how we describe them; the logical connections at issue are connections between event *descriptions* (Davidson 1980 [1963]; 1987b). If *x*, the striking of the bell, caused *y*, the bell's tolling, our describing *x* as "the cause of the bell's tolling" (as in, "the cause of the bell's tolling caused the bell's tolling") plainly cannot change the fact that *x* caused *y* – the "logical" connection between subject and predicate notwithstanding.

A second argument against the causal approach runs as follows. Causal explanations are lawlike; reasons explanations are not; so reasons explanations are not causal explanations, and when we explain actions in terms of reasons, we are not explaining them in terms of causes.

Davidson agrees that *x* causes *y* only if "some law covering the events at hand exists" (Davidson 1980 [1963], p. 18). He argues, however, that the law need not be framed in terms of how we describe *x* and *y* in stating this; and he suggests that the causal transactions required for the production of action are lawlike, even though there are no (strict or suitably rigorous) psychophysical or psychological laws. The idea evolves into "anomalous monism" in later essays (Davidson 1980c [1970]; 1980b [1973]; 1980 [1974]; 1987b; 1993b), a view derived from the following three theses: (1) "at least some mental events interact causally with physical events" (*Principle of Causal Interaction*); (2) "when events are related as cause and effect, then there exists a closed and deterministic system of laws into which these events, when appropriately described, fit" (*Principle of the Nomological*

Character of Causality); (3) "there are no precise psychophysical laws" (*Anomalism of the Mental*).[3] The three principles jointly imply "monism," Davidson argues: "If psychological events are causally related to physical events, there must, by [2], be laws that cover them. By [3], the laws are not psychophysical, so they must be purely physical laws. This means that psychological events are describable, taken one by one, in physical terms, that is, they are physical events" (Davidson 1980 [1974], p. 231).

This leaves open some interesting and important questions. If, as Davidson claims, the only "precise" (Davidson 1980 [1974], p. 231) or "strict" (Davidson 1993b, p. 3) laws are physical laws, why should we think that the *mental* features of physical events and states are causally relevant to the production of action? Are strict psychological or psychophysical laws required, after all, by causal theories of action? If there are no such laws, are we saddled with epiphenomenalism (the view that mental properties are not causally relevant to anything), or might laws or generalizations that are less than strict support the causal relevance of the mental? Here we are venturing onto territory common to the philosophy of mind and metaphysics. Davidson's views on these issues are addressed in Chapter 4.

Davidson also considers an objection that he formulates as follows: "reasons consist of attitudes and beliefs, which are states or dispositions, not events; therefore they cannot be causes" (Davidson 1980 [1963], p. 12). As he observes, we often appeal to states and dispositions as causes: "the bridge collapsed because of a structural defect." However, this does not, he says, "meet a closely related point. Mention of a causal condition for an event gives a cause only on the assumption that there was also a preceding event. But what is the preceding event that causes an action?" Davidson replies that the "onslaught of a state or disposition" – for example, the springing up of "a desire to hurt your feelings" – may fill the bill in some cases. Noticing that the time has come to do something that one wants to do may turn the trick in others: "the driver noticed . . . his turn coming up" (Davidson 1980 [1963], p. 12; 1999e, p. 499). In cases in which "we cannot explain . . . why we acted when we did . . ., explanation in terms of primary reasons parallels the explanation of the collapse of the bridge from a structural defect: we are ignorant of the event or sequence of events that led up to (caused) the collapse, but we are sure that there was such an event or sequence of events" (p. 13).[4]

In "Actions, Reasons, and Causes," in addition to rebutting various arguments against the causal approach, Davidson raises the following challenge to noncausalists about action explanation. If you hold that when we act intentionally we act for reasons, provide an account of the reasons *for which*

we act that does not treat (our having) those reasons as figuring in the causation of the relevant behavior (or, one might add, as realized in physical causes of the behavior)! The challenge is particularly acute when an agent has two or more reasons for *A*-ing but *A*-s only for one of them, as in the following example:

> Al has a pair of reasons for mowing his lawn this morning. First, he wants to mow it this week and he believes that this morning is the most convenient time. Second, Al has an urge to repay his neighbor for the rude awakening he suffered recently when she turned on her mower at the crack of dawn and he believes that his mowing his lawn this morning would constitute suitable repayment. As it happens, Al mows his lawn this morning only for one of these reasons. In virtue of what is it true that he mowed his lawn for this reason, and not the other, if not that this reason (or his having it), and not the other, played a suitable causal role in his mowing his lawn? (Mele 1997a, p. 240)

Elsewhere, I have argued that no noncausalist has successfully answered this challenge (Mele 1992c, Chapter 13; 2000; 2003, Chapter 2). Perhaps, as Davidson claims, "failing a satisfactory alternative, the best argument for a [causal] scheme . . . is that it alone promises to give an account of the 'mysterious connection' between reasons and actions" (p. 11).

3. RATIONALIZING

Davidson's "Actions, Reasons, and Causes" helped to revive a causal approach to action explanation not only by advancing telling objections to leading anticausalist arguments but also by offering a way of accommodating in a causal framework the idea, favored by many anticausalists and causalists alike, that intentional actions are explicable in terms of agents' reasons. A central notion in the Davidsonian synthesis is "rationalization," a species of causal explanation designed in part to reveal the point or purpose of the explananda.

The essay opens as follows: "What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did? We may call such explanations *rationalizations*, and say that the reason *rationalizes* the action" (Davidson 1980 [1963], p. 3). Davidson's thesis in that article is that "rationalization is a species of causal explanation." "The primary reason for an action is its cause"; and "a reason rationalizes an action only if it leads us to see something the agent saw, or thought he saw, in his action – some feature, consequence, or aspect

of the action the agent wanted, desired, prized, held dear, thought dutiful, beneficial, obligatory, or agreeable" (p. 4). When a reason is a "rationalizing" cause of an action, it is a reason *for which* the agent performs that action. In a later article, Davidson remarks: "Two ideas are built into the concept of acting on a reason . . . : the idea of cause and the idea of rationality. A reason is a rational cause. One way rationality is built in is transparent: the cause must be a belief and a desire in light of which the action is reasonable" (Davidson 1980 [1963], p. 9; 1980 [1974], p. 233). In "Actions, Reasons, and Causes," he tells us that "[i]n order to understand how a reason of any kind rationalizes an action it is necessary and sufficient that we see, at least in essential outline, how to construct a primary reason" (p. 4), where "*R* is a primary reason why an agent performed an action *A* under the description *d* only if *R* consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property" (p. 5).

Although Davidson states his position in terms of his coarse-grained theory of act individuation, that theory is not essential to the position. Henceforth, readers may treat the action variable '*A*' (or Davidson's occasional '*x*') as a variable either for actions themselves or for actions under *A*-descriptions (or *x*-descriptions), depending upon their preferred mode of act individuation; the same goes for the term 'action'. That having been said, Davidson's basic idea about rationalization, under one interpretation and with a little refinement, may be expressed as follows: a reason's rationalizing an action is a matter of its being a cause of that action that explains the action (partly) by revealing something that the agent was aiming at in performing it, and, therefore, something that makes the action "reasonable" or "agreeable," to some extent, from the agent's point of view. Obviously, the rationality associated with rationalization is understood in a thin and subjective way (cf. Davidson 1980 [1963], p. 9). An agent who pries the lid off a can of paint for a reason constituted by an urge to drink some paint (p. 4) and a belief that he can put himself in a position to drink some by prying off the lid strikes one as crazy. Even so, his action is rationalized by this reason, and from the narrow perspective of the urge and the belief, prying off the lid is an instrumentally rational course of action.

Davidson's notion of rationalizing is a broadly instrumental one. In some cases, the belief component of a reason for *A*-ing represents *A*-ing as a *means* to *E*. In others, the belief represents *A*-ing as an *instance* of *E* (e.g., the belief that going for a swim would be a good way of exercising today in someone who desires to exercise today). In yet others, the belief represents *A*-ing as a *constituent* of *E*. For example, someone who desires to serve an excellent

Thanksgiving meal may think about what would constitute serving such a meal, judge that serving pumpkin pie would be part of so doing, and desire accordingly to serve pumpkin pie. In each case, one can say, the belief component represents *A*-ing as *conducive* to *E*, conduciveness being understood to include each of the three relations just mentioned.

In a wide range of cases, when an agent *A*-s for a reason *R*, what rationalizes his *A*-ing presumably also rationalizes his *desiring* to *A*, and his *intending* to *A*, if he so intends. If Don's flipping the switch is rationalized by his wanting to illuminate the room together with his believing that the best way to do so is to flip the switch, then, presumably, Don *wants* to flip the switch, and his so wanting is rationalized by the reason that rationalizes the flipping. If it is plausible that the Davidsonian reason identified is a cause of Don's flipping the switch that explains the action (partly) by revealing something "in light of which the action is reasonable" or "agreeable," to some extent, from Don's point of view, then it is plausible, as well, that this reason is a cause of Don's desire or intention to flip the switch that helps to explain the emergence of that *desire* or *intention* (partly) by revealing something in light of which what it is a desire or intention to do is, to some extent, reasonable or agreeable from Don's point of view. In many cases, an agent may do *A* in order to *B*, do *B* in order to *C*, do *C* in order to *D*, and so forth. For instance, Don might flip the switch in order to illuminate the room, in order to make it easier to find his car keys, in order to improve his chances of getting to work on time. In these cases, the reasons for which an agent wants to do things that are relatively remote from his *A*-ing in the in-order-to chain may help to rationalize both his *A*-ing and his wanting to *A*. For example, Don's desire to get to work on time and his belief that he needs to find his keys if he is to do that may figure in a detailed rationalization both of his flipping the switch and of his desire to do so.

The rationalizing of an agent's wanting or intending to *A* by a Davidsonian reason that he has for *A*-ing may be counted as *instrumental* rationalizing. The subjective reasonability or agreeability of his *A*-ing lies in its believed conduciveness to something the agent wants, something identified in a reason for which he *A*-s. And the subjective reasonability of one's wanting or intending to *A* often derives from that of *A*-ing. Even if the agent does not represent his desiring or intending to *A* as a means to an end, the desire and intention have as their object something – the agent's *A*-ing – that is represented in the reasons that rationalize these attitudes as conducive to the achievement of the object of a desire that is a constituent of those reasons, and this helps to explain why those desires or intentions emerge.[5]

Wholly intrinsically motivated actions – actions performed only for their own sakes, or as ends – are problematic when we combine Davidson's account of reasons for action with his thesis that all intentional actions are done for reasons. Consider such actions as displaying one's gratitude to a friend – when this is done only for its own sake, from no ulterior motive – or whistling a tune just because one feels like it. If one's displaying one's gratitude to a friend is motivated by a wholly intrinsic desire to do this, a desire for this solely as an end, there seems to be no room for a belief of the sort that Davidson's account of reasons requires in the reason for which the agent so acts. One might suggest that the reason for which the intrinsically motivated action is performed – the action of displaying one's gratitude to one's friend, Bob, or, in Davidsonian terms, the action under the description "displaying one's gratitude to Bob" – is constituted by a wholly intrinsic desire to perform an action with the property of being a display of gratitude to Bob and a belief that displaying one's gratitude to Bob would have that property.[6] But that belief seems otiose; it lacks an evident explanatory function and smacks of being a device whose only function is to save a theory (see Mele 1988). (It might be suggested that the relevant belief is, for example, the belief that buying Bob a bottle of Glenlivet would display one's gratitude. But although that belief may be part of a reason for buying Bob a bottle of Glenlivet, it is not part of a reason for displaying one's gratitude.)

The problem admits of a simple solution. It has been claimed – plausibly, I have argued elsewhere – that although actions of the kind at issue are done for no *further* reason, they are done for a reason.[7] Insofar as it is plausible that (except perhaps in very special cases; see note 5) intentional actions are done for reasons constituted by psychological states of agents, showing one's gratitude to a friend, when one does this from a wholly intrinsic desire so to act, is plausibly regarded as something done for a reason constituted by an *intrinsic desire* to display one's gratitude to the friend, a desire for this as an end. The reason needs no belief component. Similarly, for the purposes of action explanation, feeling like whistling a tune – or, more precisely, an intrinsic desire to do so – may itself plausibly be understood as a reason for whistling a tune. The *general* worry about intrinsically motivated actions can be laid to rest: one can modify Davidson's account of reasons for action by allowing that intrinsic desires to *A* are themselves reasons for *A*-ing (cf. Davidson 1980 [1963], p. 6). However, problems allegedly posed by a certain *species* of intrinsically motivated action merit attention.

Rosalind Hursthouse appeals to a species of intrinsically motivated action – what she terms "arational action" – in an attempt to show that

Davidson's view of action is "fundamentally flawed" (Hursthouse 1991, p. 63), rests on a "false semantic theory" (p. 57), and introduces "mysteries" (p. 64). She focuses on Davidson's thesis that, in her words, "intentional actions are done because the agent has a certain desire/belief pair that explains the action by rationalizing it" (p. 57).[8] Examples of arational actions include striking an inanimate object in anger and gouging out the eyes in a photograph of a hated person. Hursthouse also adduces, but in another category, "actions prompted by odd physical cravings" – for example, licking something furry when "seized by a sudden desire" to do so (pp. 62–3). Such actions, as she observes, often are not done for the sake of some further goal, and they typically seem unreasonable.

It would be a mistake to infer from this that such actions are done for no reason at all. If Davidson's paint drinker can pry the lid off the paint can for a reason, what prevents us from reasonably supposing that he drinks the paint for a reason, too, a reason constituted by a wacky intrinsic desire to drink the paint? The bizarreness of his drinking the paint does not greatly outstrip the weirdness of his prying off the lid *for that purpose*. Nor is the paint drinking any less motivated than the lid prying. The man is motivated by a yen to drink a can of paint. The yen, one may feel confident, is not a *good* reason. But in the absence of a well-motivated theory of reasons for action that precludes the desire's being a reason for which he acted, are there compelling grounds for insisting that it is not a reason at all? (Hursthouse does not offer a theory of reasons.) If our reasons can be every bit as bizarre as our actions, Davidsonians have no special cause for worry.[9]

Explanations of actions in terms of the reasons for which they are done are rarely, if ever, *complete* explanations (Davidson 1999g, p. 639). Often, at least, considerably more information would be required for something approaching a full understanding of the behavior. We may want to know, for example, why an agent wanted the *end* that he pursued, or why he acted for one reason rather than for another reason that he had at the time. Concerning intrinsically motivated actions, a request for information beyond a specification of the desire on which the agent acted is often appropriate: "OK," one may say, "he did it for its own sake; but why did he want to show his gratitude to Bob, what accounted for his yen to drink the paint, what gave rise to Elvis's effective urge to shoot a television?" To produce helpful answers, we need to work a lot harder in some cases than in others. (In no case, if the action is done *only* for its own sake, does a proper answer identify a *further* reason for which the agent *A*-ed.) But Davidson has never claimed that identifying the reason(s) for which an agent acted will, in all

cases, make it plain to us why the agent acted as he did. In fact, he has denied it. In his treatment of akratic action, for example, the topic of the following section, Davidson observes that citing the reason(s) for which the agent acted leaves much unexplained (Davidson 1980b [1970], esp. p. 42; 1982).

It sometimes is claimed that Davidsonian reasons for action really are not reasons at all. T. M. Scanlon, for example, argues that "desires almost never provide reasons for action in the way described by the standard desire model" (Scanlon 1998, p. 43). Now, philosophical work on reasons for action tends to be guided primarily either by a concern with the *explanation* of intentional actions or by a concern with the *evaluation* of intentional actions or their agents. In work dominated by the former concern, reasons for action tend to be understood as states of mind, along broadly Davidsonian lines. Philosophers with the latter concern may be sympathetic or unsympathetic to this construal, depending on their views about standards for evaluating actions or agents. For example, a theorist whose evaluative concern is with *rational* action and who holds that the pertinent notion of rationality is subjective – in the sense that a proper verdict about the rationality or irrationality of an agent's action is to be made from the perspective of the agent's own desires, beliefs, preferences, principles, and the like, rather than from some external, or partly external, perspective – may be quite happy to understand reasons for action as states of mind. A theorist with a more objective conception of rational action or rational agency is likely to have a more objective conception of reasons for action as well. Such a theorist may find it very natural to insist that many or all reasons for action are facts about the agent-external world.

Michael Woods has remarked that "the concept of a reason for action stands at the point of intersection ... between the theory of the explanation of actions and the theory of their justification" (Woods 1972, p. 189). If there are external justificatory reasons for action, it may be that intentional actions are to be relatively directly explained at least partially in terms of Davidsonian reasons, and that when external justificatory reasons contribute to explanations of intentional actions, they do so less directly, by way of a causal contribution made by an agent's apprehending such a reason to his acquiring a Davidsonian reason. An exploration of the possibility of external justificatory reasons and of their compatibility with the existence of Davidsonian reasons quickly takes one well beyond the philosophy of action into moral philosophy and value theory. Further discussion of this topic is beyond the scope of this chapter (but see Mele 2003).

4. WEAKNESS OF WILL

In his introduction to *Essays on Actions and Events*, Davidson remarks that "Causal theories of action are challenged by intentional actions that are contrary to the actor's best judgment. For if reasons are causes, it is natural to suppose that the strongest reasons are the strongest causes" (Davidson 1980a, p. xii). The actions that he has in mind are incontinent or akratic actions, actions that exhibit weakness of will. In "How Is Weakness of the Will Possible?" (Davidson 1980b [1970]), he tackles this challenge, though he does not make it entirely explicit there that he sees causalism hanging in the balance. Davidson attempts to meet the challenge by arguing that the occurrence of akratic actions is compatible with the truth of a pair of principles that "derive their force" from a "very persuasive," *causal* "view of the nature of intentional action and practical reasoning" (p. 31). He formulates the principles in an attempt to give expression to a "doctrine that has an air of self-evidence" (p. 22):

P1. If an agent wants to do *x* more than he wants to do *y* and he believes himself free to do either *x* or *y*, then he will intentionally do *x* if he does either *x* or *y* intentionally.

P2. If an agent judges that it would be better to do *x* than to do *y* then he wants to do *x* more than he wants to do *y*. (p. 23)[10]

The conjunction of *P1* and *P2* entails (cf. p. 23) the following principle connecting judgment and action:

*P**. If an agent judges that it would be better to do *x* than to do *y*, and he believes himself free to do either *x* or *y*, then he will intentionally do *x* if he does either *x* or *y* intentionally.

Now, Davidson characterizes incontinent action as follows:

D. In doing *x* an agent acts incontinently if and only if: (a) the agent does *x* intentionally; (b) the agent believes that there is an alternative action *y* open to him; and (c) the agent judges that, all things considered, it would be better to do *y* than to do *x*. (p. 22)

Because (by definition) the judgment against which the akratic agent acts is an "all things considered" judgment, and therefore, in Davidson's terminology, a "conditional" judgment, the occurrence of akratic actions does not falsify *P**. The judgments with which *P** is concerned are *unconditional* judgments, not conditional ones (p. 39). ("All things considered" judgments

are counted as conditional judgments because reference to the reasons for holding that it would be best to do *y*, or better to do *y* than *x*, is part of the content of the judgment. The form of these judgments is: prima facie (*a* is better than *b*, given *r*), where *r* holds the place of all things considered [p. 39].)

Obviously, the success of Davidson's attempt to reconcile akratic action with *P1* and *P2* depends on the correctness of his characterization of akratic action. In particular, if his *D*, *minus* the phrase "all things considered," adequately describes some cases of akratic action, then some cases of akratic action plainly falsify *P**. In instances described by the modified version of *D*, the agent judges (without qualification, i.e., unconditionally) that it would be better to do *y* than *x* and believes that he may do either *x* or *y* (thereby satisfying the compound antecedent of *P**). Yet, he intentionally does *x*, falsifying *P**'s consequent. Should we believe that we never act against the *unconditional* judgment that it would be better to do *y* than to do *x* in the circumstances described in *D*?[11]

Davidson has little to say about this in Davidson 1980b, but his notion of an unconditional judgment obviously must play an important part in any answer to the present question. In a later article (Davidson 1980 [1978]), he contends that a certain kind of unconditional or "all out" judgment, "a judgment that something I think I can do . . . is desirable not only for one or another reason, but in the light of all my reasons, . . . is an *intention*" (p. 101; italics altered); and in Davidson 1980b [1970], we are told that "every judgment is made in the light of all the reasons in this sense, that it is made in the presence of, and is conditioned by, that totality" (p. 40).

This clarifies matters considerably. In Davidson 1980b [1970], Davidson claims that "[i]ntentional action . . . is geared directly to unconditional judgments like 'It would be better to do *a* than to do *b*'. Reasoning that stops at conditional judgments . . . is practical in its subject, not in its issue" (p. 39). What this reasoning stops short of is the formation of an *intention*. And what happens in cases of incontinent action, on Davidson's account, is that the agent does not intend to do what he judges it best to do, all things considered. His weakness (*akrasia*) is exhibited, not in a failure to act on an unconditional judgment – that is, on an intention – but rather in a failure to intend (and, hence, to act) in accordance with an "all things considered" judgment, a judgment that is conditional in form.

Davidson reaffirms the point in Davidson 1985e:

> I am committed to the view that an agent is incontinent only if he fails to reason from a conditional 'all things considered' judgment that a certain

> course of action is best to the unconditional conclusion that that course of action is best. . . . [S]uch a failure is just what I defined to be a case of incontinence, and what I argued was possible. . . .
>
> I find it strange . . . to think of an incontinent intention or action as an error in belief, since I think of evaluative judgments as conative propositional attitudes. So to fail to reason to the right 'conclusion' means, in practical reasoning, to fail to form attitudes in a rational, coherent way. Among those attitudes are intentions. Failure to form an intention in accord with the principle of continence is, I still think, all too possible. (p. 206)[12]

This tack is helpful in reconciling at least some cases of akratic action with *P1* and *P2*.[13] This pair of principles is concerned with unconditional judgments or intentions; and it does seem that an agent's weakness may sometimes be manifested, not in action that involves the formation (or acquisition) and subsequent abandonment of an intention, but rather in a failure to intend in accordance with an "all things considered" better judgment. It is plausible that an agent may judge it best, all things considered, to quit smoking without intending to quit.

Near the end of Davidson 1980b [1970], Davidson writes:

> There is no paradox in supposing a person sometimes holds that all that he believes and values supports a certain course of action, when at the same time those same beliefs and values cause him to reject that course of action. If *r* is someone's reason for holding that *p*, then his holding that *r* must be, I think, a cause of his holding that *p*. But, and this is what is crucial here, his holding that *r* may cause his holding that *p* without *r* being his reason; indeed, the agent may even think that *r* is a reason to reject *p*. (p. 41)

His claim, in part, is that when an agent judges, in the light of "the sum of his relevant principles, opinions, attitudes, and desires" (p. 40), that, all things considered, *y* is better than *x*, this same sum of principles, opinions, and so on – that is, this sum of reasons (which is itself treated as a reason by Davidson)[14] – may cause him to judge unconditionally that *x* is better than *y*. Thus, an agent may act, and act *intentionally*, contrary to his "all things considered" judgment; for his unconditional judgment *is* an intention. An agent may judge that, all things considered, *y* is better than *x* and yet, due to weakness, both fail to intend to do *y and* intend to do *x*. Having formed or acquired that intention, he may execute it in an akratic action. The above-mentioned sum of reasons (*r*) does not constitute the agent's *reason* for judging unconditionally that *x* is better than *y*. Indeed, the agent thinks that *r* warrants denying that *x* is better than *y*. Nevertheless, *r* causes the agent to make this judgment (i.e., to intend to *x*); and, presumably, some

member of *r* is the agent's reason for doing *x*. (Compare pp. 32–3: "The incontinent man believes it would be better on the whole to do something else, but he has a reason for what he does, for his action is intentional.")

Akratic actions do not falsify the claim that reasons are causes of action; they leave causalism intact (Mele 1987). The agent who akratically does *x*, does *x* for a reason. That he took his reason to do something else instead to be a better reason does not show that his having the reason for which he acted is not a cause of his action. One still wants to ask, however, whether we ever act akratically when our practical reasoning does *not* stop at a conditional judgment in favor of *y*-ing, but issues in an intention to *y*, even an intention to *y* straightaway. Elsewhere, I have argued that we do (Mele 1987, Chapter 3). Arguably, there is conceptual and psychological space for akratic action even after an intention to *y* straightaway has been formed. Arguably, an agent may abandon such an intention without changing his mind about what it would be best (unconditionally) to do. This dispute may be set aside here.[15]

5. INTENTIONS AND CAUSAL DEVIANCE

In "Actions, Reasons, and Causes," intentions were relegated to the sidelines. There, belief and desire shouldered the explanatory load, and such expressions as "the intention with which James went to church" were understood as referring to no "entity, state, disposition, or event" whatsoever (Davidson 1980 [1963], p. 8, cf. p. 13). In "How Is Weakness of the Will Possible?" (Davidson 1980b [1970]), as I have explained, a robust conception of intention began to take shape, one that became explicit in Davidson 1980 [1978]. Davidson's identification of intentions with better (or best) judgments has attracted a lot of criticism (Bratman 1985; Peacocke 1985; Pears 1984, Chapter 9), and he has said that he is "happy to give up the word 'judgement' " (Davidson 1985e, p. 211). However, this leaves much in place. Davidson continues to regard intentions as propositional attitudes that dispose agents to act, as states that can be generated by practical reasoning and are conditioned by an agent's beliefs, as intermediaries between reasons for action and intentional behavior, and as causal contributors to intentional actions (Davidson 1985e, pp. 195–221; 1987b, pp. 39–40).

Davidson contends that the "relation between the reasons an agent has for acting and his intention" is that "reasons cause the intention 'in the right way' "; further, if the intention is effective, then "the intention, along with further events (like noticing that the time has come), causes the action

'in the right way'," or at least is "a causal factor in the development of the action"(Davidson 1985e, p. 221). He maintains as well that "an intention to act (or to refrain from acting) requires . . . a desire or pro-attitude toward outcomes or situations with certain properties, and a belief that acting in a certain way will promote such an outcome or situation" (Davidson 1987b, p. 40): in short, and in part, an intention to *A* requires a reason for *A*-ing. In Davidson's words, his "story about how beliefs and desires cause an action" – a story encompassing these two theses – "is arrived at . . . by reflecting on the nature of beliefs and pro-attitudes on the one hand, and on the nature of action on the other." This is a traditional approach, and a fruitful one.

I have already touched upon the rationalizing role of reasons in producing intentions (§3), but what is "the right way" for an "intention, along with further events," to cause an action? And what connections between an intention to *A* and an action *A* are necessary and sufficient for the agent's having *intentionally A*-ed (or, in Davidson's terms, for the agent's action to be intentional under the description '*A*')? Here one faces challenges posed by deviant causal chains. The alleged problem is that whatever psychological causes are deemed both necessary and sufficient for a resultant event's being an action or for an action's being intentional, cases can be described in which, owing to a deviant causal connection between the favored psychological antecedents (e.g., intentions or events of intention acquisition) and a resultant event, that event is not an action, or a pertinent resultant action is not done intentionally.

The most common examples of deviance divide into two types. One type raises a problem about a relatively direct connection between mental antecedents (e.g., events of intention acquisition) and resultant bodily motion. The other focuses on behavioral consequences of intentional actions and on the connection between these actions and their consequences. The following are, respectively, representative instances of the two types of case:

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold [unintentionally]. (Davidson 1980a [1973], p. 79)

> A man may try to kill someone by shooting at him. Suppose the killer misses his victim by a mile, but the shot stampedes a herd of wild pigs that trample the intended victim to death. (p. 78)

Davidson says that although he is not sure whether the problem posed by cases of the second sort can be overcome, the difficulty posed by cases of the first sort "seems to me insurmountable" (p. 79). He adds: "What I despair of spelling out is the way in which attitudes must cause actions if they are to rationalize the action." He does not, however, abandon his causal theory of action and of action explanation – far from it.

Anticausalists and some causalists regard the combination of Davidson's causalism with this concession as problematic. The problem may be mainly dialectical. In the absence of a knockdown argument that proper reasons explanations of actions are causal explanations, noncausalists will take the absence of an account of the sort that Davidson despairs of providing as evidence that causalism is false. Davidson's challenge to noncausalists, reviewed in §2, is not itself an argument for causalism. However, Davidson may have given up on causal deviance too soon. Causalist attempts to handle problems posed by deviant causal chains include Bishop 1989; Brand 1984; Mele 2000; Mele 2003, Chapter 2; and Mele and Moser 1994.

6. FREEDOM

Davidson touches upon the topic of free action in "Actions, Reasons, and Causes." The final objection to the causal theory that he considers there is that the theory makes people helpless victims of their states of mind. Davidson (1980 [1963], p. 19) quotes from A. I. Melden (1961, pp. 128–98): "It is futile to attempt to explain conduct through the causal efficacy of desire – all *that* can explain is further happenings, not actions performed by agents. The agent confronting the causal nexus in which such happenings occur is a helpless victim of all that occurs in and to him." Davidson's reply fills just fourteen lines. His core idea is succinctly expressed: "Among ... agentless causes are the states and changes of state in persons which, because they are reasons as well as causes, constitute certain events free and intentional actions" (p. 19).

This idea is developed a decade later in Davidson's "Freedom to Act" (Davidson 1980a [1973]). There, he quickly dismisses worries about the incompatibility of causal determinism with freedom (Davidson 1980a [1973], p. 63). The more interesting worry, he says, is about the causal theory's commitment to the idea that "freedom to act [is] a *causal power* of the actor." Davidson does not offer an analysis of this causal power. Indeed, he counts his "search for a causal analysis of '*A* is free to do *x*' a failure" (p. 80), one partially explained by the alleged intractability of the problem posed by

cases of "internal" deviant causal chains, as in the example of the climber. But he does not find this discouraging: "although we cannot hope to define or analyse freedom to act in terms of concepts that fully identify the causal conditions of intentional action, there is no obstacle to the view that freedom to act is a causal power of the agent" (p. 81).

In the course of explaining the tightness of the connection between the concepts of freedom to act and intentional action, Davidson writes: "to say when an agent is free to perform an action intentionally (i.e., with a certain intention) is to state conditions under which he would perform the action; to explain the performance of an action with a certain intention is to say that the conditions are satisfied" (p. 76). If freedom to *x* intentionally is understood in this way, it is very plausible that an agent who is free to *x* intentionally may *x* intentionally without *x*-ing freely – even setting aside, with Davidson, worries about determinism.

Consider an extreme case (Mele 1995, Chapter 9). Until today, Betty was as sweet and harmless as could be. Unfortunately, skilled brainwashers have just induced in her an irresistible desire to make the world a better place by killing the president. Betty is free to kill the president, given the passage just quoted: by hypothesis, there are conditions under which she would perform the action. Suppose that the conditions obtain and she kills the president for a reason constituted by her desire to make the world a better place and her belief that she will do this by killing the president. She shoots him with the intention of killing him and improving the world thereby. In performing an intentional action that, on Davidson's account, she is free to perform – namely, killing the president – Betty certainly seems not to be acting freely. The gap between freedom to *x*, as Davidson understands it, and *x*-ing freely is significantly larger than the gap between freedom to *x* and *x*-ing intentionally, for, if predictable intuitions about Betty can be trusted, not all intentional actions are free.

7. CONCLUDING REMARKS

Davidson's greatest contribution to the philosophy of action is his resurrection of causal theories of action and action explanation. As long as Davidson's challenge to noncausalists remains unmet (see §2), causalism will be the biggest game in town, if not the only one. In the last few decades, considerable progress has been made in the development of causal theories of action explanation (see references in §2), in articulating roles for intentions in the production of sophisticated intentional behavior (Bratman

1987; 1999), and in the project of providing causal analyses of action and of doing something intentionally (see references in §5). Davidson's theory of act individuation has also inspired detailed rival causal theories (see references in §1), and his own theory certainly remains a leading contender.[16]

Notes

1. The thoughts that some omissions are actions and that omissions are not events complicate the issue, as Davidson acknowledges (Davidson 1980 [1963], note 2).
2. Earlier in the essay, Davidson noted that "the idea of a bodily movement" must be interpreted "generously" if this is to work, and "the generosity must be openhanded enough to encompass such 'movements' as standing fast, and mental acts like deciding and computing" (p. 49).
3. Principle (1) is quoted from Davidson 1980c [1970], p. 208, as are the names of the principles. Principles (2) and (3) are quoted from Davidson 1980 [1974], p. 231. For another formulation of the principles, see Davidson 1993b, p. 3.
4. On these issues, see Audi 1993, Chapter 6.
5. Elsewhere, I have argued that there are reasons for intending to *A* that are not themselves reasons for *A*-ing and that such reasons can contribute to rational intention formation (Mele 1992b; 1995). I have argued, as well, that in a very restricted subset of these cases, agents intentionally *A* but do not *A* for a reason (Mele 1992b).
6. Compare this to Davidson's remark (Davidson 1980 [1974], p. 232) that when " 'He just wanted to' " is "given in explanation of why Sam played the piano at midnight, it implies that he wanted to make true a certain proposition, that Sam play the piano at midnight, and he believed that by acting as he did, he would make it true." Incidentally, one may distinguish between theses *S* (for 'stronger') and *W* (for 'weaker') and attempt to protect Davidson against the objection being advanced by claiming that he endorses only the weaker of them:

   (*S*) If there is a description, *D*, of *x* under which *x* is an intentional action, *x* is done for a reason under *D*.

   (*W*) If there is a description of *x* under which *x* is an intentional action, there is a description of *x* under which *x* is done for a reason.

   However, the passage just quoted is evidence that his is the stronger and more interesting thesis.
7. See Locke 1974, p. 172. My argument is advanced in Mele 1988; cf. Mele 1992c, Chapter 6. Aristotle deemed it a necessary condition of being virtuous that an agent perform virtuous actions "for the sake of the acts themselves" (*Nicomachean Ethics* 1144a18–20, cf. 1105a28–33).
8. For a related objection, see Quinn 1993, pp. 236–52.
9. See Davidson 1980a [1976], p. 267: "the looniest action has its reason." For a detailed reply to Hursthouse's objection, see Mele 1992a, pp. 357–63.

10. Davidson writes: "It is easy to interpret *P2* in a way that makes it false, but it is harder to believe that there is not a natural reading that makes it true. For against our tendency to agree that we often believe we ought to do something and yet don't want to, there is also the opposite tendency to say that if someone really (sincerely) believes he ought, then his belief must show itself in his behaviour (and hence, of course, in his inclination to act, or his desire)" (p. 27).
11. On my own view, acting akratically also requires that one not be *compelled* to act as one does (Mele 1987). The absence of compulsion can be incorporated into the question just asked.
12. Davidson suggests that the following "principle of continence" is one that "the rational man will accept in applying practical reasoning": "perform the action judged best on the basis of all available relevant reasons" (p. 41). "What is hard," he continues, "is to acquire the virtue of continence, to make the principle of continence our own. But there is no reason in principle why it is any more difficult to become continent than to become chaste or brave."
13. For counterexamples to *P1* and for a modified version of the principle that circumvents them, see Mele 1992c, Chapter 3. For criticisms of *P2*, see Mele 1987, Chapters 3 and 6; Pears 1982, pp. 40–3; Taylor 1980, pp. 499–505; Watson 1977, pp. 319–21.
14. Davidson's inclusive understanding of reasons is evident in the following claim: "he does x for a reason r, but he has a reason r' that includes r and more, on the basis of which he judges some alternative y to be better than x" (p. 40).
15. Davidson returns to akratic action in Davidson 1982. For criticism of the central argument there, see Mele 1987, Chapter 6.
16. Parts of this chapter derive from Mele 1987; 1992b; 1992c; 1997a; 1997b; and 2003. I am grateful to Brendan O'Sullivan for comments on a draft of this chapter.

3 Radical Interpretation

PIERS RAWLING

You find yourself stranded on a desert island with one fellow castaway. You rapidly discover that you cannot communicate with her; indeed, it is perhaps not clear initially that she *has* a language. You must build your interpretation of her, as agent and speaker, from scratch. How is such *radical* interpretation possible? Insight into the answer to this question is apt to tell us something quite general about what it is to speak a language, and to be interpretable as a speaker – about what it is, in short, to be a linguistic being. Davidson is famous for pursuing this question by exploring the constraints on interpreting the raw data of overt behavior – constraints that we must respect if we are to count as interpreting our fellows as speakers and thinkers. He argues that in order for any interaction to be mutual interpretation, the parties must make assumptions about each other that could not turn out to be false lest their enterprise fail to be interpretation at all. In a sense, then, no interpretation *is* built *entirely* from scratch, and it is this that makes radical interpretation possible.

Davidson's exercise is "conceptual" (Davidson 1990d, p. 325). What emerges is not a manual for the field linguist, but a distinctive – and in many respects compelling – picture of language and the mind. Importantly, Davidson sees as impossible the conceptual reduction of our intentional concepts (those that we use to describe the meanings and minds of others) to the nonintentional. He seeks, rather, to illuminate the former by examining their relations to one another and to the evidence that we use in their attribution.

One of Davidson's central claims is that the constraints on interpretation are not sufficient to force uniqueness: interpretation is inevitably indeterminate. In this exposition, my main critical argument is to the effect that Davidson's view of indeterminacy is at odds with his commitment to the existence of propositional attitude states. Some have worried that Davidson's "anomalous monism" (Davidson 1980c [1970]) has "epiphenomenalist tendencies" (Kim 1993a). My worry is more radical: I claim that Davidson's views on indeterminacy yield the result that there are no propositional

attitude states. Davidson attempts to render indeterminacy benign by drawing an analogy between the attribution of mental states and measurement, but I shall argue that this analogy fails to save the propositional attitudes.

I begin this chapter, however, by considering the influence of the three figures to whom Davidson appeals the most in his work on interpretation; in alphabetical order, they are Quine, Ramsey, and Tarski. I begin with Ramsey.

1. RAMSEY

In his classic "Truth and Probability" (Ramsey 1931), one of Ramsey's main concerns is the notion of degree of belief (intuitively, degree of confidence that something is or will be so) and how this can be measured on a numerical scale (Jeffrey 1983, Chapter 3). He canvasses two possible accounts of degree of belief. We might "suppose that the degree of a belief is something perceptible by its owner; for instance that beliefs differ in the intensity of a feeling by which they are accompanied, which might be called a belief-feeling or feeling of conviction, and that by the degree of a belief we mean the intensity of this feeling" (p. 169). Ramsey rejects this account: "it seems ... observably false, for the beliefs which we hold most strongly are often accompanied by practically no feeling at all; no one feels strongly about things he takes for granted" (p. 169). His favored account is that "the degree of a belief is a causal property of it, which we can express vaguely as the extent to which we are prepared to act on it" (p. 169). This he cashes out in terms of preferences among gambles. There are, however, two determinants of such preferences: in addition to the degrees of belief in the propositions on which the outcomes of the gambles hinge, the degree to which those outcomes are desired is clearly also crucial. Here Ramsey invokes his ingenious technique[1] for determining both utilities and degrees of belief using only information about preferences among gambles.

Ramsey's method, roughly speaking, begins as follows. Call a possible event, E, "ethically neutral" (p. 177) for an agent if, for all outcomes a, the agent is indifferent between $[a \,\&\, E]$ and $[a \,\&\, \neg E]$ (represent the occurrence of an outcome a and E as '$[a \,\&\, E]$', and the occurrence of a and nonoccurrence of E as '$[a \,\&\, \neg E]$'). We must find an ethically neutral possible event E and two outcomes, a and b, such that our agent, Anne, is indifferent between the two gambles G and H,

G $\quad$ a if E; b if $\neg E$

H $\quad$ b if E; a if $\neg E$,

yet strictly prefers a to b ('a if E' means the agent receives a if E occurs, etc.). Now we suppose that if any other outcomes are substituted uniformly for a and b in G and H, Anne is indifferent between the resulting gambles; and we define Anne's degree of belief (p) in E to be one-half. This definition comports with the notion that an agent's preferences are represented by expected utility (EU), which is the sum of the utilities (or values of) the possible outcomes weighted by the agent's degrees of belief that they will occur. We have:

$$\mathrm{EU}(G) = (\mathrm{utility}(a) \times \mathrm{p}(E)) + (\mathrm{utility}(b) \times \mathrm{p}(\neg E))$$
$$\mathrm{EU}(H) = (\mathrm{utility}(b) \times \mathrm{p}(E)) + (\mathrm{utility}(a) \times \mathrm{p}(\neg E))$$

Since $\mathrm{p}(\neg E) = 1 - \mathrm{p}(E)$, and $\mathrm{utility}(a) \neq \mathrm{utility}(b)$, equating $\mathrm{EU}(H)$ and $\mathrm{EU}(G)$ entails that $\mathrm{p}(E) = 1/2$.

From here, Ramsey goes on to show how to measure utilities and degrees of belief generally. He proves a representation theorem: if an agent's preferences satisfy certain axioms, then, given any proposition P, we can attribute to her a unique numerical degree of belief in P (and measure it); and her degrees of belief across the multiplicity of propositions are coherent (i.e., obey Kolmogorov's [Kolmogorov 1956] axioms for probability [see De Finetti 1972, pp. 67–9; axiom VI can be ignored here]). The numerical measure of (expected) utility represents the agent's preference ranking: the greater the (expected) utility, the higher the rank.

For present purposes, the following of Ramsey's insights are key. First, degrees of belief necessarily fit a rational pattern. (We need not insist on perfect coherence in order for an agent to qualify as having degrees of belief, but massive incoherence would indicate that we were tracking the wrong feature.) Second, this rational pattern ensures that degrees of belief are accessible from without: we can determine them from observable behavior. And third, we see the possibility of rejecting their introspectibility. These insights provide a window on Davidson's account of the mind: he applies them to mental states *across the board*.

Davidson poses a transcendental question: what makes interpretation (radical or otherwise) possible? Ramsey, in his account of how we might gain access to another's degrees of belief, invokes structural rationality constraints built into the very notions of preference and degree of belief. This serves as a model for Davidson's more general project: in his account of how we might gain access to another's mind more generally, Davidson invokes structural restrictions built into the very nature of mental states, which in turn constitute restrictions upon what is to count as interpretation. These

restrictions on mental states make interpretation possible by providing a framework into which the interpreter must fit her behavioral data.

2. TARSKI AND TRUTH

Tarski's definition of truth is another formative influence on Davidson's view of radical interpretation. As explained in Chapter 1, Davidson casts the project of providing a compositional meaning theory for a speaker's language in the form of a Tarski-style theory of truth.

In interpretation, Davidson gives pride of place to the propositional attitudes. (He says little about pains, for example.) But among the propositional attitudes he includes the attitude of "meaning that." (He speaks [Davidson 1986c] of the "first meaning" of an utterance: it is the first meaning to which 'means that' here corresponds.) Admittedly, this attitude has an added component in the form of utterances: an agent means that p by uttering S. However, the interdependence of meaning with the other attitudes suffices for its classification with them.

How does this interact with Davidson's project of employing a theory of truth as a theory of meaning? Why does Davidson countenance an attitude of "meaning that" when he is apparently trying to do away with meaning? Because he is not an eliminativist about meaning when it is construed qua propositional attitude (see Chapter 1 for further discussion). In applying Tarski's work on truth, Davidson's goal is to replace schemata of the form

S means in L that p,

where 'S' is to be replaced by a structural description of a sentence of L and 'that p' is to be replaced by the name of a proposition, with schemata of the form,

S is true in L if and only if p,

where 'p' is now to be replaced by a sentence of English.

In formal logic, the object language is specified in purely syntactic terms. We might stipulate that our language (call it 'SL') has three sentential letters – A, B, and C – two connectives – '#' and '$*$' – and left and right parentheses. We define a sentence as follows:[2]

c is a sentence if and only if:
c is a sentential letter; or
there are sentences of SL a, b such that $c = \ulcorner(a*b)\urcorner$; or
there is a sentence a of SL such that $c = \ulcorner\#a\urcorner$.

An interpretation, **I**, for SL is a function that assigns to each of its sentences exactly one of 1, 0. For example:

$\mathbf{I}[A] = 1$; $\mathbf{I}[B] = 0$; $\mathbf{I}[C] = 1$;

For any sentences a, b:

$\mathbf{I}[\ulcorner(a*b)\urcorner] = 1$ if $\mathbf{I}(a) = \mathbf{I}(\mathrm{b}) = 1$; $\mathbf{I}[\ulcorner(\mathrm{a}*\mathrm{b})\urcorner] = 0$ otherwise;

$\mathbf{I}(\ulcorner\#a\urcorner) = 1$ if $\mathbf{I}[a] = 0$; $\mathbf{I}(\ulcorner\#a\urcorner) = 0$ otherwise.

The recursive "trick" enables a finite definition of sentencehood to classify infinitely many objects as sentences, and a finite specification of **I** to assign unique values from {0,1} to all of them. (This is a simple illustration of Tarski's seminal definition of truth for formal languages.)

There are three perspectives one can take on **I**. Tarski supposes that we know that '$*$' "means" 'and', and that '**#**' "means" 'it is not the case that'. Then, supposing a prior grip on the general notion of truth *simpliciter*, we can see **I** as giving truth-under-**I** values for every sentence in the language: $\mathbf{I}(a) = 1$ if and only if a is true-under-**I**.

For Davidson, we do not know in advance the "meanings" of '$*$' and '**#**', and we are not defining **I**, but rather constructing it as an empirical theory. The interpreter is, as it were, collecting evidence to the effect that $\mathbf{I}(a) = 1$ if and only if a is true-in-the-speaker's-language (our understanding of this latter notion supposes, as with Tarski, a prior grip on the general notion of truth *simpliciter*). To say that '$*$' (in the speaker's language) means the same as 'and' (in English) is simply to say that **I** is correct here. Note that **I** displays the compositionality of the speaker's language with respect to truth: how the truth value of a complex sentence depends on the truth values of its parts. In the interpretation of a natural language along these lines, the idea is that compositionality with respect to "meaning" will be revealed.

The third perspective is the one with which I began this exercise, on which the language is defined in purely syntactic terms. On the first two perspectives, the language is specified partly in semantic terms. On the first, the intended interpretation of '$*$' is 'and'; on the second, this interpretation is revealed as evidence is collected. On either of the first two perspectives, but not the third, change the "meaning" of '$*$' and you change the language. Natural languages are, of course, partly specified in semantic terms: change the semantics and you change the language.

Davidson does not claim that actual interpreters construct theories of truth or come to know them. What is their role? In learning logic, students typically grasp the syntax prior to exposure to any definition of

sentencehood. On the whole, they neither construct such a definition nor, in many cases, I suspect (unfortunately), come to know it. The definition serves to state what a sentence is, and it answers the question: what would it suffice a student to know in order to classify entities as sentences of logic? Davidson asks (Davidson 1994b, p. 126): "What would it suffice an interpreter to know in order to understand the speaker of an alien language, and how could he come to know it?" He suggests that "a theory of truth, constructed more or less along the lines of one of Tarski's truth definitions, would go a long way toward answering the first question." His account of radical interpretation is his answer to the second. A theory of truth for the alien *L* in the mother tongue would serve as an account of 'means-in-*L*-that' (or almost, at any rate: I leave aside, for example, discussion of sentences that lack truth values – see Davidson 1984 [1967], p. 36). And revealing 'means-in-*L*-that' is one goal of radical interpretation (the complete goal being the attribution of propositional attitudes generally). But the answers to Davidson's pair of questions might tell us no more about the psychology of actual interpreters than a recursive definition of sentencehood tells us about the psychology of actual students who know a sentence of logic when they see one. The interest of the questions lies elsewhere, according to him:

> ... I did not say speakers or interpreters actually formulate such theories [of truth]. It does seem to me, though, that if we can describe how they *could* we will gain an important insight into the nature of the intentional (including, of course, meaning), in particular into how the intentional supervenes on the observable and the non-intentional. (Davidson 1994b, p. 127)

This partial statement of the project should not be misconstrued as committing an obvious fallacy (a detective can describe how a crime could have come about and yet gain no insight into it whatsoever). Davidson lays out certain *necessary* conditions for interpretation, which include canons that interpreters cannot flout lest they fail to be engaged in interpretation, and restrictions upon the knowledge they acquire. These conditions narrow the field of potential interpretations; and it is this narrowing that makes interpretation possible. But these conditions need reveal little about the psychology of actual interpreters. First, interpreters need not self-consciously obey the canons. Second, the canons do not dictate unique interpretations but merely impose broad restrictions, and nothing is said about the skill required for constructing interpretations. Third, restrictions on knowledge need reveal little about how it is "stored."

The truth-theoretic component of Davidson's project places a restriction on the knowledge of meaning acquired by interpreters – such knowledge must be expressible (not necessarily by the interpreters) as a theory of truth. At the least, pursuing the aim of expressing a theory of meaning as a theory of truth will teach us much about meaning and truth. And if we achieve this aim, we will have a finite way of recursively stating a theory of meaning for a language, thus satisfying, in principle, two crucial conditions – that interpretations be graspable by finite minds (Davidson 1984 [1966], pp. 8–9), and that compositionality be exposed:

> a theory of meaning for a language *L* shows 'how the meanings of sentences depend upon the meanings of words' if it contains a (recursive) definition of truth-in-*L*. (Davidson 1984 [1967], p. 23)

But in seeing how meaning rests upon truth, we shall not have eliminated the former notion. Just as Tarski (1944) explicitly invokes a prior understanding of the concept of truth *simpliciter* in formulating his criterion of "material adequacy" for his definition of truth-in-a-language (the prior understanding enables us to see the correctness of this criterion), so Davidson presupposes prior understandings of both 'true' and 'means that' *simpliciter* in his attempt to use a theory of truth as a theory of meaning.

The success of the latter project is to be measured in terms of the degree of conformity between the output of the theory of truth concerning the sentences in its domain, and our understanding of what the relevant speakers mean by those sentences. The presupposed understanding of 'means that' emerges clearly in the following well-worn problem case:

> 'Es regnet' is true-in-German if and only if: it is raining and $2 + 2 = 4$.

Yet 'Es regnet' does not mean in German that it is raining and $2 + 2 = 4$. From the perspective of a theory of meaning, we might say that the equivalence has arisen via a "deviant" derivation within the truth theory. If a theory of truth is to serve as a theory of meaning, we have to ensure that

> *S* is true in *L* if and only if *p*

holds because, and only because, *S* means in *L* that *p* (Davidson 1984b [1973], p. 138). We can, perhaps, ensure this by placing restrictions on permissible derivations within the truth theory. The details of such restrictions are not my concern here. Rather, I merely want to emphasize that their selection is driven by our prior understanding of 'means that'.

3. QUINE AND THE CONDITIONS OF INTERPRETATION

Quine appeals to Wilson's (1959) "principle of charity" ("We select as designatum that individual which will make the largest possible number of . . . statements true" [Quine 1960, p. 59]) in formulating a key "maxim of translation":

> [A]ssertions startlingly false on the face of them are likely to turn on hidden differences of language. This maxim is strong enough in all of us to swerve us even from the homophonic method that is so fundamental to the very acquisition and use of one's mother tongue.
>
> The common sense behind the maxim is that one's interlocutor's silliness, beyond a certain point, is less likely than bad translation – or, in the domestic case, linguistic divergence. (Quine 1960, p. 59)

A crucial form of silliness for Quine is the flouting of logic: "fair translation preserves logical laws" (Quine 1960, p. 59). To call this "charity" is, of course, a witty misnomer: to be uncharitable is to disengage from the practice of translation.

Davidson "appl[ies] the Principle of Charity across the board. So applied, it counsels us quite generally to prefer theories of interpretation that minimize disagreement" (Davidson 1984b, p. xvii). Davidson sees this as a departure from Quine (Davidson 1984b, p. xvii), and there are others. As Davidson puts it, he "use[s] Quine's inspired method in ways that deviate, sometimes substantially, from his [Quine's]" (Davidson 1990d, p. 319).

In addition to paying heed to Quine's method, Davidson also heeds Quine's strictures. Quine opens the preface to *Word and Object* (Quine 1960, p. ix) thus:

> Language is a social art. In acquiring it we have to depend entirely on intersubjectively available cues as to what to say and when. Hence there is no justification for collating linguistic meanings, unless in terms of men's dispositions to respond overtly to socially observable stimulations. An effect of recognizing this limitation is that the enterprise of translation is found to be involved in a certain systematic indeterminacy . . .

Davidson agrees with Quine (roughly speaking) that there can be nothing more to meaning (and the propositional attitudes more generally) than what can be gleaned from observation (propositional attitudes are evidence-dependent in this sense), and that observation will never determine these attitudes uniquely (the attitudes are indeterminate). But there

are (at least) two provisos. First, Davidson takes a distal view of stimulation, in contrast to Quine's proximal view (see §5). Second, Davidson puts a new spin on indeterminacy by drawing an analogy to measurement theory (see §6).

What makes interpretation possible, according to Davidson, "is the structure the normative character of thought, desire, speech, and action imposes on correct attributions of attitudes to others, and hence on interpretations of their speech and explanations of their actions" (Davidson 1990d, p. 325). Davidson's purpose is to show how it is possible to attribute meanings and other propositional attitudes when observable behavior is our only evidence (and is, furthermore, constitutive evidence). He gives us a sketch of how an idealized interpretation might proceed. Neither the procedure nor the implicit procedural sequence in the sketch is, however, to be taken literally. The sequence in which interpreters gather evidence and formulate hypotheses is a matter of skill and serendipity. Rather, the function of the sketch is to illustrate the structural restrictions on interpretation: interpretation is possible because the interpreter is forced to interpret the behavior of interpretees as conforming to patterns dictated by the Principle of Charity – that is, it is constitutive of the propositional attitudes that they be largely rational, where rationality encompasses, among other norms, those of evidence, preference, desirability, and action. This is not to say that there is no room for irrationality; but it cannot be too pervasive. Just as there is no chaos except against a background of order, so there is no irrationality except against a background of rationality.

Interpretation begins with observable behavior, where this is restricted to bodily movements, nonintentionally described, and their attendant circumstances. Various hypotheses are then mooted concerning such matters as which behaviors are actions, which of those are utterances of sentences, and what certain of the agent's attitudes toward those sentences are. Two key sentential attitudes for Davidson are those of holding a sentence true, and of preferring one sentence true above another. But even access to these attitudes does not eliminate indeterminacy.

Suppose we have worked out that an agent holds true one of her sentences, *S*, under certain circumstances, but holds it false under others. How are we to work out what she means by it? We confront the following difficulty: her holding *S* true is dependent both upon what she means by it, and upon what she believes the circumstances to be. Meaning and belief must be accessed simultaneously. As Davidson notes (Davidson 1990d, pp. 318–19), there is a striking parallel here to Ramsey's problem of moving

from preference to both utilities and degrees of belief – and, as he goes on to say,

> Quine's solution resembles Ramsey's, in principle if not in detail. The crucial step in both cases is to find a way to hold one factor steady in certain situations while determining the other. Quine's key idea is that the correct interpretation of an agent by another cannot intelligibly admit certain kinds and degrees of difference between interpreter and interpreted with respect to belief. (Davidson 1990d, p. 319)

In general, the problem is even more challenging: in interpreting patterns of bodily movement, we must invoke a host of propositional attitudes. Within the bounds of the constraints imposed by charity, there are many ways of attributing the attitudes so as to explain bodily movements. And indeterminacy would remain even if, *per impossibile*, we could determine the totality of an agent's behavioral dispositions. As Quine puts it for the case of vocalization: "translation [is] indeterminate in principle relative to the totality of speech dispositions" (Quine 1960, p. 221).

The more restrictions imposed, however, the less the indeterminacy. As we have seen, Ramsey's decision theory serves as a model for Davidson's account of general interpretation; but decision theory also plays a direct role in, and imposes restrictions on, interpretation. Ramsey's theory, however, is not as well suited to Davidson's enterprise as (a modified [Davidson 1985d] version of) Jeffrey's (1983). This latter Davidson sees as a part of the normative structure imposed by interpretation: qualitative normative constraints (such as transitivity) are imposed on preferences for the truth of (initially uninterpreted) sentences. An agent might have individually weird preferences from our point of view, and he might violate standard canons of preference on occasion, but apparent wholesale unconcern for such canons throughout his pattern of preference would alert us to the fact that we were not tracking *preference*.

Jeffrey employs a uniform ontology of propositions, and sees preference as a matter of preferring some propositions true above others. It can be proven (the proof of the representation theorem is due to Bolker [1967]; see also Jeffrey 1983, pp. 112, 142–9) that, provided an agent's preferences over propositions meet certain qualitative conditions (such as transitivity), there are two mappings of the set of propositions to the real numbers: one yielding degrees of belief, the other yielding conditional expectations of utility. The degrees of belief are coherent, and the conditional expectations of utility represent preference (the higher the proposition on the preference ranking, the greater its conditional expected utility).

Davidson modifies Jeffrey's ontology to quantify over (initially) uninterpreted sentences of a language, rather than propositions, because access to the propositions that sentences express is, of course, a goal, not a starting point, of interpretation.[3] Davidson then applies the theory in two distinct ways (Davidson 1985d; 1990d, pp. 326–328).

First, he supposes that we have formulated hypotheses as to the agent's preferences across uninterpreted sentences, and that we have identified a sentential connective with certain properties as a truth-functional connective. He then shows (first application) that, provided the agent's preferences conform to the Bolker–Jeffrey preference axioms as applied to sentences, the agent's sentential connectives are structurally fixed relative to our hypotheses concerning the agent's sentential preferences.

(In fact, as I show in the appendix to this chapter, decision theory is not required in order for us to get this far. In effect, Davidson is using strict preference as a mark of logical inequivalence: given Davidson's suppositions, an agent's sentential connectives are structurally fixed relative to any indicator of inequivalence.)

Second, having uncovered the truth-functional sentential connectives, the agent is seen to have preferences over a Boolean algebra, and if these conform to the Bolker–Jeffrey qualitative preference axioms, the representation theorem ensures that the sentences carry coherent degrees of belief, and (conditional expected) utilities that measure preference.[4]

Decision theory and traditional logic give, in combination, an account of what it is to be rational in the "thin" sense of being coherent and consistent in one's preferences and beliefs. Charity, however, imposes further requirements: for example, "it cannot happen," according to Davidson (2001a [1991], p. 196), "that most of our plainest beliefs about what exists in the world are false" (see §5). But these strictures notwithstanding, interpretations are not unique. Even in the face of all the relevant evidence (were we able to acquire it), the Principle of Charity leaves many interpretations open: "the evidence on which all these matters depend gives us no way of separating out the contributions of thought, action, desire, and meaning one by one. Total theories are what we must construct, and many theories will do equally well" (Davidson 1984a [1979], p. 241).

Interpretation is possible only if the attitudes we attribute are publicly available; yet all we wear on our sleeves is our behavior, nonintentionally described. From this behavior, the interpreter constructs a theory. But, on pain of violating publicity, any construction that saves the behavioral phenomena and satisfies charity (and there are many) is as good as any other.[5] Hence myriad interpretations will be true of any agent.

It is in making sense of this "myriad interpretations" picture that Davidson employs measurement theory analogically. But before discussing this, I turn to two other aspects of interpretation: holism and the "location of the stimulus" problem. Concerning the first, Davidson is in agreement with Quine; concerning the second, he is not.

4. HOLISM

Holism with respect to the propositional attitudes can take a variety of forms, but, roughly speaking, the holist claims that propositional attitudes can occur only in interdependent patterns. The foregoing considerations speak in favor of the idea that attitudes must meet certain criteria of rationality: for instance, there are strong presumptions that if an agent believes that *P* & *Q*, then she believes that *P*; and if her degree of belief in *P* is high, then her degree of belief in its negation is low. And there are other forms of interdependence. For example:

> In order to believe the cat went up the oak tree I must have many true beliefs about cats and oak trees, this cat and this tree, the place, appearance and habits of cats and trees, and so on; but the same holds if I wonder whether the cat went up the oak tree, fear that it did, hope that it did, wish that it had, or intend to make it do so. (Davidson 2001b [1982], pp. 98–9)

Two questions immediately arise: how rigid is the interdependence, and how large must the pattern be?

Davidson rejects a sharp analytic-synthetic distinction, which is here tantamount to denying that conceptual grasp depends upon a fixed list of particular beliefs:

> [C]an the dog believe of an object that it is a tree? This would seem impossible unless we suppose the dog has many general beliefs about trees: that they are growing things, that they need soil and water, that they have leaves or needles, that they burn. There is no fixed list of things someone with the concept of a tree must believe, but without many general beliefs, there would be no reason to identify a belief as a belief about a tree. . . . (Davidson 2001b [1982], p. 98)

This passage suggests flexibility with respect to both interdependence and pattern size. Yet Davidson makes remarks that can be interpreted as

gainsaying such flexibilities:

> [A] Frege said that only in the context of a sentence does a word have meaning; in the same vein he might have added that only in the context of the language does a sentence (and therefore a word) have meaning. (Davidson 1984 [1967], p. 22)

> [B] Since the identity of a thought cannot be divorced from its place in the logical network of other thoughts, it cannot be relocated in the network without becoming a different thought. (Davidson 2001b [1982], p. 99)

However, Davidson has retreated somewhat from [A] (see note 3 of Davidson 1994b). And [B] misstates his intent (as conveyed verbally) if it is interpreted as attributing such rigidity to the interdependence of the attitudes that an agent cannot move from, say, believing that *P* to believing that *not P* (because interdependence entails that *P* can be no part of the content of the new belief).

This is to misconstrue the nature of the interdependence.[6] There are dependences in which changes in one factor can leave others stable. Consider Ohm's Law: voltage (in certain circuits) is the product of resistance and current. Here we have three mutually interdependent quantities, and certainly a change in one of them must result in a change in *one* of the others; but the third can remain fixed (e.g., by increasing the voltage in a circuit of fixed resistance, we increase the current). In the case of the network of propositional attitudes and utterances, we have a vast number of variables. But the same point applies: changes cause disruptions, but their scope will typically affect only a very small portion of the network. And I can share part of your network without sharing all of it (two circuits can share a potential difference of three volts, while differing in current and resistance).

5. THE DISTAL STIMULUS, CONCEPTUAL SCHEMES, AND PSYCHOPHYSICAL LAWS

The "stimulus" is a notoriously ambiguous construct. For the rat in its Skinner box, is the stimulus the box, or something internal to the rat? Davidson argues for the distal box, whereas Quine gives a proximal account.

Quine sums up (and queries) some of his earlier views on the radical translation of observation sentences thus:

> The linguist tries to match observation sentences of the jungle language with observation sentences of his own that have the same *stimulus meanings*.

> That is to say, assent to the two sentences should be prompted by the same stimulations; likewise dissent. . . . It would seem that this matching of observation sentences hinges on sameness of stimulation of both parties, the linguist and the informant. But an event of stimulation, as I use the term, is the activation of some subset of the subject's sensory receptors. Since the linguist and his informant share no receptors, how can they be said to share a stimulation? We might say rather that they undergo *similar* stimulation, but that would assume still an approximate homology of nerve endings from one individual to another. Surely such anatomical minutiae ought not to matter here. (Quine 1990b, p. 2)

Davidson shares Quine's doubts:

> [L]et us imagine someone who, when a warthog trots by, has just the patterns of stimulation I have when there's a rabbit in view. Let us suppose the one-word sentence the warthog inspires him to assent to is 'Gavagai!' Going by stimulus meaning, I translate his 'Gavagai!' by my 'Lo, a rabbit' though I see only a warthog and no rabbit when he says and believes (according to the proximal theory) that there is a rabbit. (Davidson 1990a, p. 74)

According to Davidson's distal theory 'Gavagai!' is to be interpreted as 'Lo, a warthog'. His distal theory is part of what drives his claim that "it cannot happen that most of our plainest beliefs about what exists in the world are false" (Davidson 2001a [1991], p. 196).

Quine's response to Davidson is to move "to an intermediate point between Don's distal and [his] old proximal position" (Quine 1990a, p. 80), although he "remain[s] unswerved in locating stimulation at the neural input. . . . Unlike Davidson, [he] leave[s] the stimulations at the subject's surface, and private stimulus meaning with them" (Quine 1990b, pp. 3–4). Space precludes details of Quine's intermediate position, beyond noting that "the subject's reification of rabbits and the like is . . . decidedly a part of the plot, not to be passed over as part of the setting" (Quine 1990b, p. 3). The position of reification marks a key distinction between Davidson and Quine on the issue of conceptual schemes.

In Davidson (1984b [1974]), Quine is one of the interlocutors who countenances what Davidson denies: the possibility of differing conceptual schemes – incommensurate ways of carving up the world. Conceptual schemes that are incommensurate at the basic level of objects are a possibility only if "reification . . . is . . . a part of the plot." On Davidson's distal theory, however, reification is "part of the setting": objects such as warthogs are intersubjectively accessible, and it is they that in part determine the content of our speech and thought.

The advocate of a proximal view might well maintain that the propositional attitude states of an agent are caused by matters external to the agent, and cause him to behave in various ways. But, crucially, she would deny that matters external to the agent enter into the *specification* of such states. For Davidson, on the other hand, the contents of propositional attitude states are in part specified by their external causes:

> The contents of beliefs and other mental attitudes are specified by mentioning objects, or kinds of objects, with which the subject of those attitudes must have come into causal contact of one sort or another. (This is naturally not always true, but it must be so in the most basic cases.) (Davidson 1990c, p. 21)

And they are in part specified by what they are disposed to cause the agent to do external to her body:

> [A] belief that a stone is lethal will, when combined with certain desires, cause an intention to kill. . . . (Davidson 1990c, p. 22)

This is not to say that propositional attitude states are not internal states of the agent – it is just that their specification qua contentful propositional attitude requires reference to external objects and events:

> Though our beliefs, intentions, fears, and other feelings are private and subjective if anything is, they cannot be identified or explained except by tying them from the start to external objects and events. (Davidson 1990c, p. 23)

On Davidson's view, a mind must have a history of causal interaction with the world, and dispositions toward it, in order for it to be in states that are about the world; and these two features enter into the standard psychological specification of those states.

In Davidson's view, austere science, when mature, does not specify states in terms of causal history or disposition. 'Sunburn' and 'frangible' will not appear in mature science (Davidson 1990c, p. 22). A mature science will specify, say, the state of sunburned skin in a fashion that makes no reference to the sun. Let '*X*' abbreviate this specification. To learn that *X* can be caused by overexposure to the sun is to make an empirical advance beyond the fact that sunburn is caused by overexposure to the sun. To describe somebody as desiring warthog for supper is to specify their state in part in causal-historical terms, and in part in dispositional terms. Thus, propositional attitude talk will not appear in mature science. And neither will discussion of language, if it is specified partly in semantic terms.

This is one strand in Davidson's argument against psychophysical laws. (Matters are more subtle than Davidson supposes, I think: perhaps 'mass' as it appears in Newtonian physics is a disposition term; perhaps 'entanglement' as it appears in quantum theory is a causal-historical term.) Another strand appeals to the fact that thinkers must be found rational, whereas "[t]he astronomer and physicist are under no compulsion to find black holes or quarks to be rational entities" (Davidson 1990c, p. 25).

Davidson's distal theory, then, helps to undergird his views that: (1) our basic beliefs are largely true (see Chapter 6); (2) thought cannot conceive the world in radically incommensurate ways; and (3) there are no psychophysical laws (see Chapter 4).

Given that conceptual schemes would have to be expressible, (2) receives support from (1) (Davidson 1984b [1974], pp. 194–5). (1) implies that the expression of any conceptual scheme would be a body of sentences that are largely true. Tarski gives us our best understanding of truth, and he ties it, according to Davidson, to "translation into a language we know"[7] (Davidson 1984b [1974], p. 195). Thus we have: a sentence *S* of language *L* is true-in-*L* iff its translation into English is true-in-English. Hence the expression of any conceptual scheme would have to be translatable into English, and no conceptual scheme could be radically incommensurate with that of an English speaker.

What of the claim that conceptual schemes would have to be expressible? This is closely related to Davidson's claim that there is no thought without talk. One consideration here is that, in the absence of language, it is hard to make sense of the attribution of intensionality:

> The dog, we say, knows that its master is home. But does it know that Mr. Smith (who is his master), or that the president of the bank (who is that same master), is home? We have no real idea how to settle, or make sense of, these questions. (Davidson 1984 [1975], p. 63)

A second line of argument runs, in outline, as follows (Davidson 1990a; 2001b [1982]; 2001 [1992]). To have a belief requires cognizance of the possibility of error, which in turn requires cognizance of the distinction between truth and falsity. But this distinction only emerges in the context of communication. Why? Because in order for, say, a predication of redness to have a truth value, the subject of the predication (the distal stimulus) must be fixed. And, Davidson claims, no lone individual can pull off this feat: the "location of the stimulus" problem can be solved only if there are (at least) two communicators "triangulating" upon a mutually salient common stimulus. Otherwise the stimulus is arbitrary, and nothing ensures

that any of one's thoughts are about the external world as opposed to, say, internal neural firings.

Davidson not only argues, then, that beliefs must be largely true and interpretable, but also that they must be (at some point) actually interpret*ed*.

6. THE MEASUREMENT ANALOGY

Quine is skeptical about the reality of the attitudes:

> If we are limning the true and ultimate structure of reality, the canonical scheme for us is the austere scheme that knows no quotation but direct quotation and no propositional attitudes but only the physical constitution and behavior of organisms. (Quine 1960, p. 221)

And, as we have seen, Davidson sees no place for the attitudes in mature science. But he is not an eliminativist. I turn now to the question of whether this can be squared with his take on the notion that interpretation leaves beliefs, meaning, and the other attitudes indeterminate.

Davidson attempts to render this indeterminacy benign by invoking an analogy between the way sentences function in the attribution of propositional attitudes, and the way numbers function in measurement theory: the fact that there are different ways of interpreting our fellows is no more alarming, he thinks, than the fact that, say, length can be represented by different "schemes of measurement" such as feet and meters. If we press this analogy, however, it emerges that indeterminacy is more alarming than Davidson concedes: on his view it transpires that there are no such states of mind as belief states (even externally construed), nor indeed any states corresponding one-to-one with any of the propositional attitudes.

The notion of a representation theorem was introduced in §1. Here is a simple case. The relation 'is at least as long as' conforms to certain qualitative criteria, such as transitivity and connectedness: any concept that fails to conform to these criteria is not the concept of length (Davidson 1980c [1970]). If certain structural conditions are added to the essential qualitative criteria, then the relevant representation theorem can be proven: there is a mapping of the entities to be measured (the possessors of "lengths") to the nonnegative reals that represents the relation 'is at least as long as'. And since in the case of length there is a natural zero, the mapping is unique up to multiplication by a positive real. Note the ingredients: we have the entities to be measured, a qualitative relation on those entities, and a mapping (unique up to some transformation) from the entities to the measuring

domain (the nonnegative reals, in the case of length) that preserves relevant structure (in the case of length, the longer the object, the greater the associated real).

The basic idea of Davidson's analogical strategy is clear enough (Davidson 2001 [1989]). We use sentences to track the propositional attitudes of an agent; we use the reals to track the lengths of objects. The sentences in the former attribution play the role of the real numbers in the latter. We might maintain that there is an "indeterminacy of length": there are infinitely many serviceable schemes for attributing lengths. Similarly, there are many serviceable schemes for attributing meanings and the other attitudes to a given interpretee. And the latter indeterminacy is as benign as the former. But the details of the analogy need spelling out.

Davidson does not deny that there are many disanalogies between the two cases. There is, for example, an algorithm for moving from one scale of length measurement to another (multiply by the relevant positive constant); there is no such algorithm in the case of interpretations. However, we do need at least the following. In the measurement theoretic case, we map to numbers from an underlying framework that is invariant between scales. Thus, in the case of using sentences to attribute attitudes, we need to identify an underlying framework that is invariant across the various schemes of interpretation. The invariant framework must consist of relata and relations between them. In the case of length, we have objects standing in the relation 'is at least as long as'. Infinitely many different mappings of the objects to the reals will serve to track the relation, but the objects and the relation do not vary across the mappings.

In the case of attitude attributions, I shall concern myself only with the relevant relata. It might initially appear that the invariant relata are the propositional attitudes of the interpretee. However, this cannot be. The relata cannot comprise the propositional attitudes themselves, since we *identify* propositional attitudes by the sentences we use to track them. (In measuring length, we use numbers to *measure* the relata, *not* to identify them.) And it is precisely the assignment of these sentences that varies between schemes of interpretation. On Davidson's picture, indeterminacy is simply the fact that we can use different locutions to locate the same node in some pattern. But what are the invariant nodes? They cannot be propositional attitudes: the belief that P, say, under one scheme, will be the belief that Q under another – two different propositional attitudes (both attributions are couched in the one idiolect of the interpreter).

One might wonder whether there is not some "neutral" way of identifying propositional attitudes – so that, despite appearances to the contrary,

the belief that *P* and the belief that *Q* are in fact the same belief. This runs counter to the Davidsonian picture, however. For example, attitudes in one scheme need not map one-to-one onto attitudes in another. Perhaps it is possible that behavior interpreted as a signal could equally well be interpreted, under a different scheme, as a simple scratch. And signals require more propositional attitude states than simple scratches.

Matthews (1994) might appear to provide a way out. In the case of length, the "empirical relational system" to be represented comprises objects and the qualitative relation 'is at least as long as'. And the "representation space" (the space that is to represent the empirical relational system, to which that empirical system is mapped) comprises the real number system. In the case of the attitudes, Matthews supposes that there is a substrate of propositional attitudes to be "measured" (these are the invariant relata), although he (sensibly) leaves open the issue of exactly which relations among the attitudes are to be represented. He argues that the appropriate representation space for representing the propositional attitudes comprises a set of "ordered pairs $<a_i, <s_j, r_k>>$, consisting [of] an attitude-type a_i and what [he calls] a designated proposition $<s_j, r_k>$, where r_k is a Russellian proposition and s_j a sentence-type, a token of which in a particular (unspecified) context serves to designate (express) that Russellian proposition" (Matthews 1994, p. 136).

What is it for there to be two different scales of propositional attitude measurement on Matthews's account? He furnishes the following example:

> Suppose that Smith describes the man that he knows only by the name 'Tully' as being destitute by saying, 'Tully is destitute'. I might explain Smith's remark to someone who knows Tully only as 'Cicero' by saying, 'Smith believes that the wealthy Cicero is destitute', thereby conveying not simply that Smith is mistaken in his belief, but also who his belief is about, namely, the man that my interlocutor knows only as 'Cicero'. In the context, the representations associated with sentence types *Tully is destitute* and *the wealthy Cicero is destitute*, both of which share the same attitude-type and Russellian proposition, represent one and the same belief. In a different explanatory context, e.g., one in which Smith's actual words were important, these same representations might represent different beliefs. (Matthews 1994, p. 143)

As Matthews correctly notes, the context dependence here marks a difference between his account of propositional attitude "measurement" and the measurement of length. Nevertheless, suppose we concede that in the first explanatory context in the cited passage we have one belief mapped to two

different elements of the representation space. This is the analogue of an object being mapped to two different reals in measurement of its length. In the latter case, we have two different scales (feet and meters, say) at work; so perhaps this is so in the former case also (with propositional attitudes being the measured entities that are invariant across the scales).

Matthews's picture is in marked contrast to my construal of Davidson's account, however. If Davidson were to concede that Matthews ascribes one and the same belief to Smith with utterances of 'Smith believes that the wealthy Cicero is destitute' and 'Smith believes that Tully is destitute', this would be, presumably, because (in the contexts of utterance) the two attribution sentences must be so interpreted in any interpretation of Matthews that saves the relevant phenomena. Matthews is not proposing two interpretations of Smith. But for Davidson, the interpretation is the analogue of the scale of measurement; thus we do not have an analogue of two different scales.

What, then, are the invariants across interpretations on Davidson's account? As we have seen, Quine has us tracking "men's dispositions to respond overtly to socially observable stimulations" (Quine 1960, p. ix). And according to Davidson, "a satisfactory theory is one that yields an acceptable explanation of verbal behavior and dispositions" (Davidson 1984a [1979], p. 237). Behavioral dispositions are key; and they are apparently the invariant nodes across different schemes of interpretation – they are what propositional attitude attributions "measure." Just as we attribute numbers in the form of lengths in order to track the relation 'is at least as long as', and numbers in the form of degrees of belief and expected utilities in order to track preference, so we attribute propositional attitudes in order to track behavioral dispositions.

The invariants, then, have the following form: a disposition to behavior *B* under circumstance *C*, where *B* and *C* are specified nonintentionally. A full interpretation of an agent – something unattainable by mere mortals – would account for the totality of the agent's behavioral dispositions. And all adequate interpretations must agree on these dispositions; but that is all that they must agree on. Propositional attitude states are mere posits; they vary from interpretation to interpretation. That we cannot reidentify a belief across interpretative schemes is a consequence of the fact that there is no such state to reidentify.

An instrumentalist in the philosophy of science holds, roughly, that theories must save the phenomena, and must meet criteria of coherence and simplicity (and elegance, perhaps). Theories will (usually) posit entities that underlie the appearances; but, according to the instrumentalist, these

are only posits – constructions designed to save the phenomena. If two theories were to save the phenomena, score equally well in other respects, but disagree as to posits, the instrumentalist would have no difficulty in countenancing both: the posits, being merely that, would not compete.

Davidson's view of the mind, then, appears to be an instrumentalist one. The relevant criteria of coherence are provided by the Principle of Charity; the posits comprise, among other things, propositional attitudes. About some of the other posits, Davidson is quite explicit: "words, meanings of words, reference, and satisfaction are posits we need to implement a theory of truth. They serve this purpose without needing independent confirmation or empirical basis. . . . [Satisfaction and reference are] notions we must treat as theoretical constructs whose function is exhausted in stating the truth conditions for sentences" (Davidson 1984b [1977], pp. 222–3). And: "degree of belief is a construction based on more elementary attitudes" (Davidson 1990d, p. 322). When it comes to the propositional attitudes, however, he seems, more wary.

Davidson does say:

> In thinking and talking of the weights of physical objects we do not need to suppose there are such things as weights for objects to have. Similarly in thinking and talking about the beliefs of people we needn't suppose there are such entities as beliefs. (Davidson 2001 [1989], p. 60)

But in the same paper, he speaks of beliefs as mental states with identity conditions. However, if there were *ontological* indeterminacy (i.e., cases in which two competing interpretations not only save the relevant phenomena, but also are both true), then there would in general be no trans-scheme identity of propositional attitudes, and hence no propositional attitudes. Drawing an analogy to measurement theory fails to render indeterminacy anodyne. The attitudes cannot be both (epistemically) indeterminate (in the sense of outrunning the evidence for their attribution, so that competing interpretations save the relevant phenomena) and evidence-dependent (in the sense that there is nothing more to them than the evidence provides, so that they *cannot* outrun the evidence).

7. IN FAVOR OF ATTITUDES

Davidson emphasizes the anti-Cartesian aspect of his view (Davidson 2001 [1989]). In the *Meditations*, we find a view of the mental on which it makes sense to ask whether "the ideas which are in me are similar . . . to the

things which are outside me" (Descartes 1955 [1643], p. 160). Although Descartes rejects the view that our "ideas" are similar to their causes,[8] it is common to attribute to him a view on which the mind contains ideas that somehow correspond to external objects. There are many objections to this sort of picture. On Davidson's novel alternative, such Cartesian "ideas" are simply rejected – and this is consistent with a reading of Davidson that has him advocating a measurement-theoretic reduction of the propositional attitudes to mere posits. (Note that the Cartesian population is only culled: there remain pains and their ilk – these are not mere posits.)

On this reading perhaps we can maintain our ordinary ways of talking about our fellows: we use the posits to track the dispositions of our neighbors – instrumentalism and realism agree that the phenomena must be saved. But there are other phenomena to be saved, and a host of difficulties associated with abandoning the attitudes – for example, epistemology (and much else, of course) is apparently undercut, since there is no such state as knowing that *P*.[9] Davidson himself is committed to the existence of propositional attitudes. In his monistic metaphysics, he quantifies over mental events. "Anomalous monism" (Davidson 1980c [1970]) is his view that all mental events are also physical, but that the mental and the physical are linked by no strict laws. Coming to have a certain propositional attitude is an event, on this view, but if there are no propositional attitudes there are no such events. There is no occupant for the mental side of the monistic equation.

Propositional attitudes are also intrinsic to Davidson's causal account of action. He famously argues that propositional attitude states of the agent cause actions (Davidson 1980 [1963]; 1980a; see Chapter 2 for discussion). And he challenges those who would deny this to furnish an account of the distinction between merely rationalizing reasons for a particular act, and those rationalizing reasons for which the agent acted. A politician presses for environmental legislation. She believes that pressing the legislation will garner votes in the impending election. She also abhors the destruction of the environment. She claims that she acted only for the latter reason. On Davidson's account, her claim is tantamount to claiming that the latter, but not the former, propositional attitude was part of the cause of her pressing of the legislation. (This is oversimplified: see Davidson 1980a [1973].) What becomes of Davidson's account if there are no propositional attitude states? Of course, there are still behavioral dispositions, but these dispositions can do no rationalizing – they have no propositional content.

Indeed, if a necessary condition for a bodily movement to be an action is that it be caused by intentional states (Davidson 1980a), then, absent the latter, there are no actions. In essence, Davidson claims, following Quine, that there can be nothing more to the intentional than is necessary for communication. And this, I have argued, given indeterminacy, rules out intentional states. But this rules out actions.

Instrumentalism concerning the propositional attitudes might be defended by noting that we attribute propositional attitudes in order to track "real" behavioral dispositions. However, there are strong arguments for realism about the attitudes, one of which is the fruitfulness of the essays collected in Davidson's *Essays on Actions and Events* (1980). Some aspect of Davidson's view has to give way. The conclusion that there are no propositional attitudes results from the combination of indeterminacy and evidence dependence with respect to them. I shall endeavor to counter Davidson's commitment to the latter.[10]

Recall Davidson's remark that, when it comes to interpretation, "a satisfactory theory is one that yields an acceptable explanation of verbal behavior and dispositions" (Davidson 1984a [1979], p. 237). With this remark he seems already to have endorsed evidence transcendence: dispositions can be present without ever being made manifest. Why then insist that propositional attitudes be evidence-dependent, when evidence-transcendent dispositions are countenanced? There appear to be two notions of evidence transcendence at work. Davidson is willing to countenance "potential" evidence (Davidson 1984a [1979], p. 237); so although dispositions are evidence-transcendent in the sense that they can be present without ever being made manifest, perhaps there must be potentially available evidence for their presence. However, this line continues, the indeterminacy of propositional attitude attribution outruns even the totality of potential evidence. (Recall Quine: "translation [is] indeterminate in principle relative to the totality of speech dispositions" [Quine 1960, p. 221].) What distinguishes propositional attitude attribution from disposition attribution, then, is that, unlike the latter, the former outruns all potential evidence – hence the nonexistence of propositional attitude states.

However, such a line will lead to a form of antirealism with respect to many dispositions. Many current dispositions will evaporate without ever being made manifest or even leaving a trace for future generations. They are currently real because they do not now outrun all potential evidence. Yet their claim to have existed will eventually outrun all potential evidence, and will thereby come to languish in the no-man's-land between truth and

falsity. To reject this antirealist line is to acknowledge the truth or falsity of dispositional claims that outrun even all potential evidence. And then there would seem to be no barrier to the ontological admission of the attitudes. Davidson countenances behavioral dispositions; thus he should do the same for propositional attitudes, and see indeterminacy as simple underdetermination of theory by observation.

We have arrived at realism concerning the propositional attitudes by dropping evidence dependence. What appeared to be ontological indeterminacy we now see to be merely epistemic underdetermination. However, another way to save the attitudes would be to establish that there is no indeterminacy.

One major apparent source of indeterminacy is the "inscrutability of reference." Davidson (1984a [1979]) considers two interpretations of an utterance of 'Wilt is tall': on theory one, it is interpreted as meaning that Wilt is tall; on theory two, it is interpreted as meaning that the shadow of Wilt is the shadow of a tall thing. According to Davidson, theory one is satisfactory if and only if theory two is satisfactory, since Wilt is tall if and only if the shadow of Wilt is the shadow of a tall thing. (Thus the reference of 'Wilt' and the extension of 'is tall' are inscrutable.) However, these interpretations are equally satisfactory only if all that counts is the preservation of the truth conditions of the interpretee's sentences. And this is tantamount to letting derivations run rampant within theories of truth that are supposedly serving as theories of meaning. In §1, it was noted that, in the absence of derivational restraint, our theory of meaning for German might yield an unfortunate conclusion: 'Es regnet' means that it is raining and $2 + 2 = 4$. The truth conditions for 'Es regnet' and 'it is raining and $2 + 2 = 4$' are the same. Thus, if truth conditions were all that mattered to meaning, our conclusion would not be unfortunate after all – and neither would the conclusion: 'Wilt is tall' in English means that the shadow of Wilt is the shadow of a tall thing.

Inscrutability of reference is only a source of indeterminacy of meaning if meaning is a matter only of truth conditions. But, then, why bother with derivational restraint within the truth theory? Evidence transcendence is implicated here. On the one hand, it is supposed that we can obtain evidence only about truth conditions, and hence, absent evidence transcendence, there can be no more to meaning than truth conditions. And then we can let derivations within the truth theory run rampant. But, on the other hand, a prior understanding of 'meaning that' is invoked in order to place derivational constraints on the truth theory.

A partial resolution might emerge from considering whether it is possible to obtain evidence that distinguishes between theories of truth that, though equivalent qua theories of *truth*, are not equivalent qua theories of *meaning*. If, for example, we admit the relevant pointing behavior as evidence that 'Wilt' refers to Wilt, then only ' "Wilt is tall" is true if and only if Wilt is tall' will serve as a meaning clause in the theory of truth that provides us with our theory of meaning, even though it remains the case that 'Wilt is tall' is true if and only if the shadow of Wilt is the shadow of a tall thing. Such evidential considerations might decrease indeterminacy, while in no way compromising the use of theories of truth (with derivational constraints) as theories of meaning.

These considerations notwithstanding, however, (epistemic) indeterminacy seems to be ineliminable. I began with Ramsey's two-for-one deal: two theoretical constructs (degrees of belief and utilities) for one piece of behavioral evidence (preferences among gambles). The interpretative task is Ramsey's writ large: a host of theoretical constructions are needed to save all of the behavioral evidence. If we are to be realists about these constructions, we must acknowledge that they are evidence-transcendent.

APPENDIX: IDENTIFYING LOGICAL CONSTANTS

Davidson proffers a demonstration that an agent's preferences structurally fix the truth-functional connectives of her idiolect (Davidson 1985d; 1990d, Appendix). He utilizes a modification of Jeffrey's (1983) version of decision theory – the modification consists in making the objects of preference sentences rather than propositions, so as not to beg the question of interpretation. However, matters are in fact simpler than Davidson supposes: truth-functional connectives are structurally fixed in Davidson's fashion provided that we have some way of determining when two sentences are not truth-functionally equivalent. In essence, Davidson's method uses the fact that if a sentence A is strictly preferred true over B, then A is not equivalent to B. I shall simply assume access to some relation R such that:

If ARB then A is not equivalent to B.

I shall assume with Davidson that '|' is a binary truth-functional connective (Davidson 1985d, p. 95) – note that this is quite a strong assumption.

With relation R replacing preference, Davidson's result (Davidson 1985d, p. 95) is:

> Suppose that there are A, B such that:
>
> (i) $(A|A)R(B|B)$
>
> (ii) $AR(A|A)$
>
> (iii) $(A|(A|A))RA$
>
> (iv) $(A|(A|A))R(A|A)$
>
> (v) $(A|(A|A))RF$ (where F is a necessary falsehood: more on (v) below)
>
> Then the truth table for '|' is:

A	B	$A\|B$
T	T	F
T	F	T
F	T	T
F	F	T

> Proof:
>
> We can think of '|' as a function on {<T, T>, <T, F>, <F, T>, <F, F>} into {T, F} (where '<T, T>' abbreviates 'the ordered pair T, T' etc.). (i) and (ii) jointly imply that T|T = F and F|F = T (if T|T = T and F|F = T, or T|T = F and F|F = F, then $A|A$ would be equivalent to $B|B$; if T|T = T and F|F = F, then A would be equivalent to $A|A$).
>
> Hence the only four possibilities for the truth-table for '|' are:

A	B	(1)	(2)	(3)	(4)
T	T	F	F	F	F
T	F	T	T	F	F
F	T	T	F	T	F
F	F	T	T	T	T

> (iii) rules out (2); (iv) rules out (3); and (v) rules out (4).

There is no assumption comparable to (v) in Davidson's scheme. However, it is required in my scheme to rule out (4) in favor of (1). I suspect that Davidson is relying on Jeffrey's (1983) exclusion of necessary falsehoods

from the preference ranking to yield the result that $(A|(A|A))$ (or, actually, $((t|u)|((t|u)|(t|u)))$ [Davidson 1985d, p. 95]) is a necessary truth rather than a necessary falsehood: provided necessary falsehoods are excluded from the preference ranking, Davidson's account yields (1) as the truth table for '|' – otherwise (4) remains a possibility.[11]

Notes

I have had much help from Donald Davidson, Lawrence H. Davis, John Heil, Siggi Kristinsson, Kirk Ludwig, Bob Maydole, Al Mele, Thad Metz, Greg Ray, and Paul Roth, and from audiences at the 1996 V. Karlovy Vary Symposium on Analytic Philosophy: Interpreting Davidson (Czech Republic), Davidson College, Keele University (England), and the University of St. Andrews (Scotland). Of course, those from whom I have learned do not necessarily agree with me. Financial support was provided by the University of Missouri Research Board and the University of Missouri at St. Louis.

1. In conversation, Davidson described his excitement at independently discovering this technique (in the 1950s), before he became aware of Ramsey's priority.
2. '⌜' and '⌝' are the left and right corner quotes respectively; where '*a*' and '*b*' are metalinguistic variables, '⌜(*a*∗*b*)⌝', for example, is an abbreviation for 'the concatenation of '(' with *a* with '∗' with *b* with ')'.
3. There are technical difficulties here that Davidson does not go into. For example, Bolker's representation theorem requires that the domain of preference be a complete atomless Boolean algebra. I shall ignore such difficulties for present purposes.
4. In theories such as Ramsey's, the qualitative preference axioms fix degrees of belief uniquely. But unless utilities are unbounded above and below, Bolker–Jeffrey degrees of belief are *not* unique (see Jeffrey 1983, Chapters 6–8 and p. 154; Joyce 1999, pp. 122–7). From Davidson's perspective, however, this is merely part of the indeterminacy of interpretation.
5. The publicity criterion does not imply that the propositional attitudes are *reducible* to behavior:

   > Propositional attitudes can be discovered by an observer who witnesses nothing but behavior without the attitudes being in any way reducible to behavior. There are conceptual ties between the attitudes and behavior which are sufficient, given enough information about actual and potential behavior, to allow correct inferences to the attitudes. (Davidson 2001b [1982], p. 100)

6. I am not the first person to note this: Priest, for example, makes the point (Priest 1981, p. 78).
7. Davidson goes beyond Tarski here, and needs to say more. Tarski would, I think, maintain only that for each (formal) object language there is a metalanguage in which the former's truth definition can be given. It does not follow from this

alone that there is one metalanguage that accomplishes the task for all object languages, let alone that English does so.

8. He says the following in the *Meditations*:

   > [A]lthough in approaching fire I feel heat, and in approaching it a little too near I even feel pain, there is at the same time no reason in this which could persuade me that there is in the fire something resembling this heat any more than there is in it something resembling the pain; all that I have any reason to believe from this is, that there is something in it, whatever it may be, which excites in me these sensations of heat or of pain. (Descartes 1955 [1643], pp. 193–4)

9. My thanks to Eve Garrard for this point.
10. I shall not here tackle Quine's argument to the evidence dependence of meaning from considerations of language acquisition.
11. Kirk Ludwig pointed this out to me – I am much indebted to him concerning this appendix and other aspects of this chapter.

4 Philosophy of Mind and Psychology

JAEGWON KIM

Donald Davidson's work on the nature of the mind has had a major impact on contemporary discussions in philosophy of mind; it is fair to say that no other philosopher has been more influential in shaping the basic contours of the field as it exists today. As is true in the case of many other philosophers, both of his generation and of following generations, his reflections on mind and its relation to matter have been carried on with a set of broad physicalist assumptions as a backdrop. Precisely what these assumptions are is not easy to say, and it is not necessary to set them down in exact formulations. For most philosophers who have reflected on the status of mind, including Davidson, the main task has been that of finding for mind a place in an essentially physical world. The world is fundamentally a physical world in that the space-time world is the entire world, and that within space-time we find nothing but bits of matter and increasingly complex aggregates made up of bits of matter.[1] As C. Lloyd Morgan, one of the early emergentists, aptly put it, there is "no insertion of alien forces" (Morgan 1923) when complex physical systems are generated out of simpler ones; there are only material elements arranged in new relationships and structures. These material entities have various physical properties – mass, motion, energy, electric charge, temperature, elasticity, and the like – and they behave in accordance with the laws of physics. What gives structure to this world of matter is the causal relation. One widely, if not universally, shared view is that all causal relations that hold in the world are fixed by the prevailing physical laws; once the distribution of bits of matter in space-time and the laws governing their behavior are fixed, that fixes the causal structure of the entire world. As we will see, it appears that Davidson accepts such a view. But where in such an austerely physical world do minds find a home? Finding an answer to this question has been the mind-body problem for contemporary philosophy of mind.

During the past several decades, nonreductive physicalism has been the dominant orthodoxy on the mind-body problem. And this is in no small measure due to the influence of Davidson's work. The basic idea

is that, although all of the entities and events in the world are physical, some of them – in particular, those with an appropriate degree and kind of organizational complexity – can, and do, have properties and features that are not physical, and that among these nonphysical properties of material systems are mental properties. In what sense are these properties nonphysical? How could material things have nonphysical, or nonmaterial, properties? The standard answer runs something like this: First, these properties are nonphysical in the sense that they do not appear in our basic physics (we do not find meanings, thoughts, and qualitative characters of conscious experience in physical theory) and, further, because they are not reducible to the properties dealt with in physics. Biological properties, like metabolism and reproduction, are not found in basic physics either; but, unlike meanings and thoughts, they are plausibly viewed as reducible to physicochemical properties. Second, although mental properties are irreducible to physical properties, they are nonetheless constrained by, or dependent on, physical properties in that the total physical description of any system will completely determine which of these nonphysical properties it has. That is, the nonphysical properties of an object, like mental properties, are supervenient on its physical properties. To summarize, we can characterize nonreductive materialism in the following three claims:

> *Ontological materialism*: There are only material things and events in this world.
>
> *Property dualism*: Material things and events can, and some of them do, have nonphysical properties and features.
>
> *Supervenience*: The nonphysical properties of an object, or event, are wholly determined by its physical properties (thus, any two objects or events that are physically indiscernible are indiscernible *tout court*).

Like other nonreductive materialists, Davidson accepts these three doctrines.[2] However, his version of nonreductive physicalism – that is, "anomalous monism" – is strikingly different from other standard versions, both in its fundamental motivation and in its philosophical implications. What at bottom sets Davidson's position apart from these other nonreductivist positions – indeed, the fundamental insight that anchors Davidson's anomalous monism – is a doctrine that he calls the "anomalism of the mental." Without a doubt, this is the central idea that has shaped Davidson's perspectives on almost all of the issues concerning mind, matter, and agency

that he has addressed. We will begin with a discussion of this celebrated doctrine.

1. THE ANOMALISM OF THE MENTAL

In "Mental Events" (Davidson 1980c [1970]), one of the most widely discussed and influential works in philosophy of mind during the second half of the twentieth century, Davidson writes: "There are no strict deterministic laws on the basis of which mental events can be predicted and explained." The intended contrast is with the domain of physical events where there are such predictive and explanatory laws. At least a couple of things need to be explained about this statement of mental anomalism (as we will call it, for brevity). First, what does Davidson mean by 'strict deterministic law'? Second, what is included under the rubric 'mental'?

Let us consider the second question first. Davidson makes clear that by 'mental' what he has in mind is the propositional attitudes, like belief, desire, and intention – also commonly called "intentional" or "content-bearing" states – and not sensations and other qualitative or phenomenal states, now often referred to as "qualia." As far as I know, Davidson has shown little interest in the latter class of mental states, which now figure prominently in discussions of consciousness. Rather, Davidson's central focus in philosophy of mind and language has been on intentional/cognitive states, as shown by his innovative works on radical interpretation, the relationship between belief and meaning, belief-desire explanations of actions, weakness of the will, and the like. In any case, phenomenal mental phenomena, like feeling a pain and sensing red, are expressly excluded from the purview of mental anomalism. What interests Davidson is the fact (assuming him to be right) that our beliefs, desires, thoughts, and intentions, unlike physical events, are not governed by strict laws and, therefore, are not explainable or predictable in the way that physical events are. Davidson rightly takes these intentional states as forming the foundations of our speech, cognition, intentionality, and agency.

We now turn to the notion of a "strict deterministic law." In "Mental Events," Davidson gives an explanation of 'law' in a fairly standard way (that is, standard for the mid-twentieth century), although he does not believe that a fully general and noncircular characterization of lawhood is possible. A law, on this account, is a true general statement, of the form 'All *F*s are *G*' in simple cases, that (i) is capable of supporting counterfactuals and subjunctives (like 'If *x* were an *F*, it would be *G*') and (ii) is

projectible in the sense that it is confirmed by observation of "positive" instances (things that are both *F* and *G*). Thus, the statement that metals expand when heated is a law, whereas the true statement that every student in my metaphysics class comes from a state whose name begins with an 'M' is not a law. The passage from Davidson just quoted speaks of "strict deterministic" laws. Deterministic laws are usually contrasted with statistical (or probabilistic, stochastic) laws, and this appears to have been Davidson's intention. It seems, though, that Davidson later dropped 'deterministic'; it is the strictness of laws that has become central. I believe that Davidson came to see that the laws of quantum mechanics, for example, do not allow exceptions, and that they are in that sense "strict," even though they are irreducibly statistical. There is a distinction to be made between statistical laws and laws ("ceteris paribus laws") that tolerate, and are not necessarily refuted by, exceptions or "negative" instances. Given the law that atoms of a certain element decay with a probability of *r* in each second, every atom of this kind without exception has probability *r* of decay in the next second; there can be no exceptions, or else the law is falsified. In contrast, the "law" that shy people blush easily tolerates exceptions; exceptions do not necessarily falsify it, and it is not withdrawn when isolated counterexamples are observed.

What about 'strict' in "strict law"? There has been some controversy about Davidson's intended meaning, or about what the term could usefully mean, but a kind of consensus seems to have emerged on this point. A strict law must, first of all, be an exceptionless law; it does not tolerate exceptions in the sense just noted. Second, it must be a member of a set of laws that together form a comprehensive theory over its domain – "comprehensive" in sense that every event or phenomenon in the domain can be given a description in the vocabulary of the theory and can be explained and predicted under that description (to the extent that it can be explained or predicted at all). Basic physical laws are thought to constitute a comprehensive theory in some such sense, and Davidson has lately observed that strict laws are perhaps found only in "developed physics" (Davidson 1993b). In our discussion, however, this second component of the concept of "strict law" will not play an important role (except in one instance where we will discuss the derivation of psychological anomalism from psychophysical anomalism); it is a rather complex notion that isn't easily managed in most argumentative contexts. It is of interest mainly as a point of contrast, as Davidson perceives it, between physical theory, with its aspiration to "full coverage" of all of reality, and other theories and schemes, including the mental, which have more restricted domains and concerns.

The Davidson quotation cited earlier also speaks of predicting and explaining mental events on the basis of laws. It seems clear that this does not add anything new: if mental events should come under strict laws of the sort we have been discussing, that would mean that they are predictable and explainable in terms of these laws. However, there is one further point that is crucial. For Davidson, laws are statements. So when he speaks of psychological laws, these laws are couched in psychological predicates. Or, if we take an ontological turn and think of laws as regularities in nature, then laws connect properties and kinds, not individual events. Thus, psychological laws will involve psychological kinds and properties, connecting them to other kinds and properties, psychological or otherwise. On Davidson's view of events as concrete spatiotemporal particulars, events, like objects, can fall under different kinds and have different properties. Let us take the following as the statement of Davidson's mental anomalism:

> *The anomalism of the mental*: There are no strict laws involving mental kinds or properties.

This means that mental or psychological kinds or properties do not come under exceptionless laws of any kind, whether these laws connect psychological kinds to other psychological kinds or to physical kinds. We can, therefore, divide mental anomalism into two parts:

> *Psychophysical anomalism*: There are no strict laws connecting psychological kinds to physical kinds. That is, there are no laws correlating mental phenomena with physical phenomena.
>
> *Psychological anomalism*: There are no strict laws connecting psychological kinds to psychological kinds. That is, there are no purely psychological laws, laws that concern only psychological phenomena.

Laws of the first kind might connect intentional/cognitive states to neural states of the subject,[3] or to physical behavior. Laws of the second kind might state ways in which beliefs lead to other beliefs, ways in which beliefs and desires together lead to intentions, and so forth. In "Mental Events" and in two other papers from the same period, "The Material Mind" (Davidson 1980b [1973]) and "Psychology as Philosophy" (Davidson 1980 [1974]), psychophysical anomalism receives the lion's share of attention, and the anomalism of the mental, to all intents and purposes, is equated with the claim that there are no laws connecting mental to physical phenomena. And the bulk of Davidson's argument for the anomalism of the mental focuses

on establishing psychophysical anomalism, with psychological anomalism decidedly taking a back seat.

There has been no unanimity as to what exactly Davidson's argument is for the anomalism of the mental, or more specifically, for psychophysical anomalism (Child 1993; Kim 1985; Lycan 1981). But we should first note the groundbreaking and, indeed, shocking nature of Davidson's anomalism when it was first proposed in the late 1960s. The assumption previously held by almost all philosophers, and also psychologists, going back many years, had been that although mental phenomena might be distinct from physical phenomena, there are intimate lawful connections between the two domains. Hoary mind-body theories – such as the double-aspect theory, parallelism, neutral monism, and the doctrine of preestablished harmony – had been erected on the basis of a belief in a pervasive system of lawful correlations between mind and body; and the early mind-body identity theory, formulated by Herbert Feigl and J. J. C. Smart, had presupposed it. The very idea of "nomological danglers," made popular by these advocates of reductionist physicalism, implies that there are psychophysical laws – "dangling" or otherwise – that we must either eliminate or else take as irreducible basic laws. The supposed existence of these correlation laws was what made it possible for us to think that the mental world and the physical world make up a unified and integrated whole, and that although we as persons have both a mental and a physical aspect, they are closely coordinated, with each constraining and being constrained by the other. Unsurprisingly, Davidson's claim that there are no such laws – indeed, that there *cannot* be such laws – was met with widespread incredulity, not to say hostility, in some quarters, and the difficulty of spelling out Davidson's argument in detail only made the doctrine more controversial – more suggestive and intriguing for some, more unbelievable for others.

I believe, however, that there is a line of understanding Davidson's considerations on which psychophysical anomalism is at least presumptively credible. Let us begin with Davidson's mental holism. Consider someone who believes that Boston is fifty miles to the north of Providence. If he holds this belief, he must also hold a number of other related beliefs, such as the belief that Providence is to the south of Boston, that Boston is more than 10 miles from Providence, that Providence is less than 100 miles south of Boston, that Boston is not to the south of Providence, that Boston is a city, and countless others. This is a consequence of our interpretive practice, the procedure whereby we attribute beliefs and other mental states to others: the system of beliefs we attribute to a person must be maximally

rational, coherent, and true relative to observable physical and behavioral evidence, and, even to be considered a belief system, it must meet a certain standard of minimal rationality, coherence, and truth. An interpretive system that attributes to a person the belief that roses are red but refuses to attribute to her the belief that roses have a color, or the belief that roses are not green – or one that, in addition to the belief that roses are red, also attributes to her the belief that roses are transparent or the belief that roses are green – would fail such a test. The requirement of rationality and coherence also applies across mental states of distinct types: we expect that a person who intends to open the window and who believes that the handle must be turned in order to open the window forms an intention to turn the handle. If she lacks the second intention, we must reconsider her first intention or belief, and if we were unable to retract either, we would find her system of intentions and beliefs unintelligible, and this would throw into doubt the reality of the attributed mental states. Beliefs, intentions, and the rest are possible only as elements of an integrated, "holistic" system, and what give the system intelligible structure are the principles of rationality, consistency, and coherence. For Davidson, the norms of rationality and coherence, which underlie mental holism, are the "constitutive principles" of mentality; they give intentional mentality their distinctive identity as an autonomous domain.

On the physical side, however, considerations of rationality and coherence have no application; in Davidson's words, constraints of rationality, coherence, and consistency, which generate "the holistic character of the cognitive field," have "no echo in physical theory" (Davidson 1980 [1974], p. 231). This isn't to say that the physical domain is not itself holistic; it is to say only that the constitutive principles of physical holism are different from those that constitute mental holism. Davidson does not say much about what these physical principles might be, but we can conjecture that they may be principles governing causality, space-time, measurement of fundamental magnitudes, and the like.

Let M be a mental event, say, the belief that Boston is fifty miles to the north of Providence, and let M^* be the belief that Boston is more than ten miles from Providence. Suppose then that M and M^* are connected to neural events N and N^*, respectively, by strict psychophysical laws of the following biconditional form:

(1) Necessarily, M occurs to person S at t iff N occurs to S at t.

(2) Necessarily, M^* occurs to person S at t iff N^* occurs to S at t.

And we have:

(3) Necessarily, if M occurs at t, M^* occurs at t.

From all this, it follows:

(4) Necessarily, if N occurs at t, N^* occurs at t.

This last statement is solely about two physical states, N and N^*, and apparently carries the nomological force of 'necessarily' attaching to the first two laws (the necessity attaching to the third statement is, or is akin to, logical necessity). So the upshot is that psychophysical laws (1) and (2) would enable us to "read off" a physical law, (4), from a purely psychological entailment, (3), which is sanctioned by norms of rationality governing beliefs. And the transition from one physical state, N, to another, N^*, with which it may be logically or metaphysically unrelated, is apparently supported and sanctioned by the normative principles of rationality and coherence. The supposed psychophysical laws would have the effect of importing the norms of rationality into the physical realm, a domain in which such norms have no application.

Or consider the situation in the opposite direction. Suppose that each mental state is connected by a biconditional psychophysical law to some physical state. These systems of biconditional laws would map physical nomological/causal relations onto the mental domain, and, as a result, the patterns of causal connections and dependencies holding among physical events would be projected into the recipient domain, perhaps in conflict with and overriding the pattern of relationships generated by the principles of rationality and coherence. That is, psychophysical laws would serve as a conduit for the transmission of the constitutive causal principles of the physical domain into the mental domain, thereby compromising the fundamental constitutive principles of mentality. In consequence, the intentional mental domain would be threatened with a loss of its distinctive identity, which is defined by norms of rationality and coherence. In Davidson's words:

> If the case of supposed laws linking the mental and the physical is different, it can only be because to allow the possibility of such laws would amount to changing the subject. By changing the subject I mean here: deciding not to accept the criterion of the mental in terms of the vocabulary of the propositional attitudes. (Davidson 1980c [1970], p. 216)

The intuitive core of this argument is perhaps not wholly novel; it is related to the idea that a pervasive system of psychophysical laws – especially those

of the biconditional form – would result in the reduction of the mental to the physical,[4] thereby depriving the mental of its autonomy. It is also related to the commonsensical idea that if mentality were wholly dependent on physical processes, which these psychophysical laws would underwrite, we would all become captives of physical laws, merging with the stream of physical processes, like anything else in the physical domain. What Davidson has done is to bring out explicitly the defining normative character of the mental domain, and to show how the retention of this character, on which the very existence of the mental domain depends, is inconsistent with, and undermined by, the existence of constraints as strong as strict laws connecting it to the physical domain whose constitutive principles are alien to the mental world. In the foregoing, we have assumed that psychophysical laws take the biconditional form; this is a convenient assumption that makes the point of Davidson's argument especially salient. However, it is not necessary for a general argument: whatever form, or forms, the psychophysical laws may take, they will represent the constraints imposed by the physical side on the mental realm, including, in particular, the patterns of interconnections among mental phenomena.

Thus far we have discussed Davidson's considerations for psychophysical anomalism. How does Davidson proceed from here to the full anomalism of the mental, or to psychological anomalism? There is this passage in "Mental Events":

> It is not plausible that mental concepts alone can provide [a comprehensive framework for the description and law-based prediction and explanation of mental events], simply because the mental does not . . . constitute a closed system. Too much happens to affect the mental that is not itself a systematic part of the mental. But if we combine this observation with the conclusion that no psychophysical statement is, or can be built into, a strict law we have the principle of the Anomalism of the Mental: there are no strict laws on the basis of which we can predict and explain mental phenomena. (Davidson 1980c [1970], p. 224)

Here Davidson seems to be arguing as follows: There can be strict laws only in closed domains. This follows from the definition of a 'strict' law. The domain of mental phenomena is not closed; we know that it is not causally closed – that is, that there are events outside the domain that causally affect events within the domain. In order to sharpen predictive generalizations about mental events, therefore, we must take recourse to concepts and properties that are not mental (e.g., biological, physicochemical). This shows the failure of closure for the mental domain, in contrast to the closed physical

domain; and there can be no theory that gives comprehensive coverage of the mental domain. It follows then that there are no strict laws over the mental domain (recall the definition of "strict law"), and this establishes psychological anomalism. When combined with the psychophysical anomalism earlier argued for, this establishes the full anomalism of the mental.

We can be reasonably confident, I think, that the foregoing was what Davidson had in mind – or at any rate, something quite close to it. One question it raises is where the assumption of the closure of the physical domain comes from – for example, whether it is in some sense an a priori metaphysical principle or an a posteriori scientific discovery. As we will see in the section to follow, one of Davidson's basic premises for anomalous monism asserts that mental events causally interact with physical events. If so, why can't we say, mimicking Davidson, that "too much happens to affect the physical that is not itself a systematic part of the physical"? Wouldn't this undermine the closure of the physical system? I believe that there is here room for Davidson to formulate a reply. But we must set these issues aside and move on.[5]

2. ANOMALOUS MONISM

Anomalous monism, a form of nonreductive materialism, is Davidson's position on the mind-body problem. It is physical monism in that it recognizes only physical objects and events. It includes the claim that the mental is anomalous – that is, as we saw in the previous section, the claim that mental properties and kinds do not fall under strict laws. Davidson's argument for the first component of anomalous monism – namely, the claim that all individual mental events are physical events – is highly ingenious. It can be seen as a new and innovative form of the causal argument for physicalism.

Davidson's argument makes use of three premises:

> *Mind-body causal interaction*: Mental events cause, or are caused by, physical events.
>
> *Nomologicality of causation*: Events related as cause and effect must instantiate a strict law.
>
> *The anomalism of the mental*: There are no strict laws involving mental kinds.

Davidson's argument is striking in its simplicity and ingenuity. According to the first premise, mental events are in causal relations with physical events. Let m be a mental event, and suppose that m causes (or is caused by) physical event p. We want to show that m is a physical event. According

to the second premise, the pair of events, *m* and *p*, must instantiate a strict law, *L*. What kind of law could *L* be? According to the third premise, *L* cannot be a psychophysical law; nor can it be a purely psychological law, since there are no such laws.[6] It follows then that *L* must be a physical law connecting physical kinds to physical kinds. Since the pair, *m* and *p*, instantiates a physical law, *m* must fall under a physical kind (or have a physical description). An event is mental or physical as it falls under a mental kind or a physical kind. Hence, *m* is a physical event. To put the conclusion in more general terms, every mental event that enters into a causal relation with a physical event falls under a physical kind and is therefore a physical event.

Notice the generality of the argument, which may not be apparent at first blush: the first step in the argument, to the effect that *m* causes, or is caused by, a physical event, can be weakened. Obviously, it is enough if *m* is causally related to at least one other event, whether physical or mental. The rest of the argument will run exactly the same way. The only mental events that the argument misses, therefore, are those that are not causally related to any other event, whether mental or physical or of any other kind. It is highly dubious that there could be such causeless and effectless events; and even if there were, their existence would be unexplainable, and they would make no difference to anything else. If Davidson's argument doesn't cover them, that is no genuine limitation to the argument. With the additional premise that every event is either a mental event or a physical event, a more general form of physicalist monism can be deduced, namely, that every event is a physical event. Alternatively, if we have available as a premise Davidson's later conjecture that strict laws are found only in physics, the same conclusion can be derived. In either case, all events, on anomalous monism, are physical events, although some of them are also mental events, biological events, and so forth. What is interesting about the argument, as Davidson has observed, is that the anomalousness of the mental serves as an essential premise, whereas most arguments for the mind-body identity theory, like those advanced by Feigl and Smart, begin with the assumption that there are nomological connections between the two domains. The crucial next step in arguments of this kind is to justify the enhancement of mind-body correlation laws into mind-body identities, usually on the basis of considerations of simplicity and explanatory unity (Block and Stalnaker 1999; Smart 1959).

The nonexistence of laws connecting mental to physical properties, according to Davidson, shows the mental to be nomologically irreducible to the physical. As may be recalled, the standard Nagelian conception of reduction requires "bridge laws" connecting properties to be reduced to properties in the base domain (Nagel 1961). Since there are no psychophysical laws, it follows that there can be no psychophysical bridge laws. Because a

behavioristic reduction of the mental, based on definitions of mental concepts in terms of physical/behavioral concepts, is out of the question, Davidson regards the irreducibility of the mental as established. It is clear, then, that anomalous monism is a form of nonreductive physicalism – it is physicalism because it claims that all mental events are physical events, and it is nonreductive in that it considers mental properties to be physically irreducible. It is also a form of "token" physicalism: each individual event, or "token event," is a physical event. And it excludes type physicalism, the claim that mental properties and kinds are physical properties and kinds. But anomalous monism differs importantly from another influential form of nonreductive physicalism, due to Hilary Putnam, Jerry Fodor, and others, which is based on the functionalist approach to mentality. According to this form of functionalism, mental kinds, though irreducible to physical/biological kinds, are physically "realized" or "implemented." Moreover, a single mental kind – say, pain – can, and usually does, have diverse physical realizers in different biological species and structures (this phenomenon of "multiple realization" is what is supposed to preclude the identification of mental properties with physical properties). This means that, on this view, there *are* type-type connections between the mental and the physical, although not of the kind that would sanction reduction or reductive identification. If pain is realized by *C*-fiber stimulation in human beings (as it is said), by *X*-fiber stimulation in octopi, and by *Z*-fiber stimulation in Martians, then each of these realizers constitutes a nomologically sufficient condition for pain. That is, there will be laws such as 'If *C*-fiber stimulation occurs in a human being at *t*, pain occurs to that human being at *t*', 'If *X*-fiber stimulation occurs in an octopus at *t*, pain occurs to it at *t*', and so on. What would enable reduction, but what according to these functionalists we cannot have, are "bridge laws" of the biconditional form, for example, '*C*-fiber stimulation occurs in *x* at *t if and only if* pain occurs to *x* at *t*'. Of course, Davidson's anomalism is not supposed to extend to sensory events such as pains, but the example could easily be couched in terms of belief, desire, or some other intentional state. Davidson's anomalism of the mental disallows even one-way laws like those connecting mental states to their physical realizers.

3. ANOMALOUS MONISM AS A THEORY OF MIND

Anomalous monism was welcome news to many philosophers who had a deep commitment to a physicalist world view but who, like Davidson, were

reluctant to embrace reductionism or eliminativism. In its assertion that all individual mental events are physical events, it attempts to preserve physicalism, and at the same time it promises to protect the autonomy of the mind with its rejection of the physical reducibility of the mental. The mental autonomy that Davidson wants is not the kind of autonomy that philosophers like Jerry Fodor have sought for psychology and other special sciences (Fodor 1974). Rather, it is the autonomy of agency and the will, of the kind that Kant famously sought. In the closing paragraph of "Mental Events," Davidson writes:

> We explain a man's free actions, for example, by appeal to his desires, habits, knowledge and perceptions. Such accounts of intentional behavior operate in a conceptual framework removed from the direct reach of physical law by describing both cause and effect, reason and action, as aspects of a portrait of a human agent. The anomalism of the mental is thus a necessary condition for viewing action as autonomous. (Davidson 1980c [1970], p. 224)

Davidson concludes the paper with a quotation from Kant in which Kant describes the philosophical problem of reconciling the two seemingly contradictory views that we hold about ourselves – first, that we are free and autonomous agents, and second, that we are part of the natural world and subject to its laws. Anomalous monism is Davidson's response to the Kantian challenge. It is not simply a technical thesis about the irreducibility of psychology as a special science; it has a deeper philosophical aim, namely, that of providing a solution to the metaphysical and moral conundrum arising out of our dual nature as agents and natural objects.

All this makes it evident that the anomalism of the mental is the centerpiece of Davidson's philosophy of mind; his entire picture of mentality, its relation to the world of matter and cause, and our status as free agents flow out of it. Mental phenomena, qua mental phenomena (or under mental descriptions), do not come under predictive/explanatory laws of the kind that we have come to know from the physical sciences. Mental events do enter into causal relations, on Davidson's view, but here too the laws that ground these causal relations are physical laws, laws connecting physical kinds to physical kinds. Unsurprisingly, the anomalous character of the mental, especially given Davidson's commitment to the nomologicality of causality, has been a focus of debate over the adequacy of anomalous monism as a theory of mind.

Consider, first, the question exactly what anomalous monism says about the mind-body relation. What do the three premises and the conclusion of

the argument for anomalous monism say about how our mentality is related to our bodily nature? The answer, I believe, is: not much. What it says is mainly negative, namely, that there are no significant relationships between mental types/properties and physical types/properties. As far as anomalous monism goes, the kind of mental creature that we are has nothing to do with the kind of physical being that we are. True, anomalous monism says that each individual mental event is a physical event. But that means only that every event that falls under a mental kind, or that has a psychological description, falls under a physical kind, or has a physical description. And the anomalous element of anomalous monism emphatically reminds us not to expect any significant relationship between the mental kind that a given event is and the physical kind that it is. It is also true that one of the premises says that some mental events stand in causal relation to some physical events. But Davidson's argument for anomalous monism shows that any causal relation involving a mental event and a physical event holds only because a strict physical law subsumes the two events under physical kinds or descriptions. The fact that the mental event is a mental event, or that it is the kind of mental event that it is, appears to be entirely immaterial to the causal relation.

This point has been the basis of a well-known and much-debated criticism to the effect that anomalous monism leads to epiphenomenalism. Individual mental events, under anomalous monism, do have causal efficacy, but only because they fall under physical kinds, and the mental kinds that they are have nothing to say about what physical kinds they fall under, and therefore nothing to say about what causal relations they enter into. The causal structure of the world is wholly determined by the physical kinds and properties instantiated by events of this world.

Consider how this scenario is played out in Davidson's influential and important causal theory of rationalizing ("belief-desire") explanations of actions (Davidson 1980 [1963]). According to this account, reasons for actions explain them in virtue of being their causes, and rationalizing explanations are a species of causal explanations. If I am climbing a ladder because I want to help my cat down from the roof and believe that this is the best way to get the cat down, my want and belief explain my climbing by being its cause. But causal relations must instantiate laws. What law, then, does my want-belief pair and my climbing instantiate? Here is what Davidson says:

> The laws whose existence is required if reasons are causes of actions do not, we may be sure, deal in the concepts in which rationalizations must deal.

> If the causes of a class of events (actions) fall in a certain class (reasons) and there is a law to back each singular causal statement, it does not follow that there is any law connecting events classified as reasons with events classified as actions – the classifications may even be neurological, chemical, or physical. (Davidson 1980 [1963], p. 17)

Davidson is saying that just because my want and belief pair causes my climbing, we should not expect to find laws couched in terms of kinds such as wants, beliefs, and climbings. The subsumptive causal laws may be – according to Davidson's mental anomalism, *must* be – physical laws that connect physical kinds to physical kinds. It would seem that the fact that my wanting is the kind of mental event that it is, and has the kind of content that it has, and the fact that my belief has the content that it has, play no role in the fact that the want and belief together cause, and thereby causally explain, my climbing. Doesn't this mean that mental kinds and properties, and hence reasons, are causally irrelevant?

A number of critics have advanced this objection, apparently independently of one another (Honderich 1982; Kim 1984c; Sosa 1984; Stoutland 1980). For helpful discussion and development of considerations along these and related lines, see Antony 1989; 1994. Also, there have been defenders of Davidson against the epiphenomenalist charge; see, for example, Horgan 1989; Lepore and Loewer 1987; and McLaughlin 1989; see also Marras 1999. In the next section, we will consider Davidson's own response as presented in his "Thinking Causes" (Davidson 1993b).

4. NONSTRICT LAWS AND MIND-BODY SUPERVENIENCE

Davidson has employed three strategies to meet the charge that anomalous monism has epiphenomenalist implications. First, he has insisted that, far from being a form of epiphenomenalism, his anomalous monism explicitly states, in its very first premise ("Mind-Body Causal Interaction"), that mental events cause, and are caused by, physical events. Moreover, he has given an account of how singular causal statements, affirming causal relations between individual events, are to be understood. Causation is a binary extensional relation that connects a pair of individual events – two concrete particulars – no matter how these events are described or conceived. In this sense, individual mental events can cause, and often do cause, other events, and that should suffice to refute the accusation of epiphenomenalism. His critics, Davidson argues, have surreptitiously introduced

questionable nonextensional, multigrade relations like 'event *c* qua *X* causes event *e* qua *Y*' (e.g., 'My wanting qua a wanting caused my climbing qua a climbing'), 'event *c* under description *D* causes event *e* under description *D**', and the like. It isn't clear that this is an effective response, however. Davidson's critics have replied that the extensionality of the causal relation is not an issue here, and that the real issue is the causal efficacy, or relevance, of the *properties* of events, or of the fact that a given mental event is the *kind* of mental event that it is (Kim 1993a; McLaughlin 1989). It appears that Davidson tacitly accepts this point, for in his second and third replies he tries to show that anomalous monism does not render mental properties causally impotent, and that mental properties are indeed causally relevant and efficacious.

In his second reply, Davidson argues that although there are no strict psychological or psychophysical laws, there are nonstrict laws about mental kinds/properties, and that these suffice to ground singular causal statements involving mental events. That is, a mental event can be a cause of a physical event in virtue of the fact that the mental event is of mental kind *M*, the physical event is of physical kind *P*, and there is a nonstrict law connecting *M* to *P*. For example, although it is not a strict law that embarrassment causes blushing, it is a law nonetheless, and this nonstrict law suffices to ground the singular causal statement that Fred's embarrassment at the party yesterday caused him to blush. This means that the cause being an instance of embarrassment and the effect being an instance of blushing are relevant to their being related as cause to effect. If the cause had been an episode of anxiety attack, it would not have caused the effect, as there is no law, strict or nonstrict, connecting anxiety attacks to blushing.

This way of meeting the threat of epiphenomenalism, although it may be intrinsically plausible, is problematic for Davidson. First, adopting the strategy throws in doubt the rationale of Davidson's principle of the nomologicality of causality, which requires that strict laws underwrite causal relations. If nonstrict laws can do the job, what is the reason for the strict law requirement? It may be that whenever there is a nonstrict law subsuming a cause-effect pair, there also is a strict law that subsumes it. Still, this does not fully explain why strict laws are required for causal relations; we need an argument here. Perhaps considerations of the primacy of physical causation might provide such an argument (McLaughlin 1989). Second, if nonstrict laws can ground causal relations, why can't they underwrite nomological reductions as well? Davidson's argument for the irreducibility of the mental depends on his psychophysical anomalism – namely, that there are no "bridge laws" that could be used to support the derivation of

psychological laws from physical laws.[7] If Davidson is right in his claim (Davidson 1993b) that there are strict laws only in "developed physics" (if anywhere), and if strict bridge laws are required for reduction, then there could be no reduction anywhere except within fully developed physics. So why not allow nonstrict psychophysical laws to ground the physical reduction of the mental? If there is no plausible reply to this question, Davidson's case for the autonomy of the mental, which is so crucial to his Kantian project of reconciling free agency with physical determinism, is in serious jeopardy. Most importantly, the appeal to nonstrict laws in order to back singular causal relations apparently threatens his celebrated argument for anomalous monism. Recall that the crucial step in this argument is from the fact that mental event *m* causes physical event *p* to the conclusion that a strict physical law must subsume *m* and *p*, there being no strict psychological or psychophysical law that could do the job. But if nonstrict psychophysical laws suffice to back the singular causal claim, Davidson's reasoning to the conclusion that mental event *m* must fall under a physical kind and is therefore a physical event is rendered fallacious – unless supplemented by defensible extra premises. For these reasons among others (see Kim 1993a; McLaughlin 1993), the strategy of invoking nonstrict laws in order to vindicate the causal efficacy of mental properties appears to be ill-advised.

Finally, Davidson appeals to mind-body supervenience in order to deflect the epiphenomenalist charge. In a well-known passage in "Mental Events" (Davidson 1980c [1970], p. 214), Davidson wrote:

> Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect.

This passage, seemingly tossed off as an afterthought, introduced, or reintroduced,[8] the concept of supervenience into discussions of the mind-body problem. It should be noted that in this paragraph Davidson doesn't say that he accepts mind-body supervenience; he says only that it is "consistent" with his anomalous monism.

The core idea of supervenience is dependence; and mind-body supervenience, as Davidson explains, means that the mental character of a thing, or an event, is dependent on, or is determined by, its physical character. The idea is also standardly explained by saying that things that are alike in all physical respects must be alike in mental respects as well – or, as is

sometimes put, no difference in mental respects without some difference in a physical respect. Suppose that an event falls under a mental kind *M*. If it were not of kind *M*, then it would not be the kind of physical event that it is. Since physical kinds are causally efficacious, and different physical kinds must be expected to differ in causal powers, this means that if the event were not an *M*-type event, it would not have the effects that it has. This, Davidson argues, shows that the fact that the event falls under mental kind *M* makes a causal difference.

Clearly, mind-body supervenience is a net addition to anomalous monism. By adopting it, Davidson has substantially strengthened his position on the mind-body problem. Two questions arise regarding this strategy as a way to stave off the threat of epiphenomenalism. First, does it work? Second, is mind-body supervenience consistent with Davidson's anomalism of the mental?

First, it isn't clear that this solution to the mental causation problem is sufficient for Davidson's purposes. Appeal to supervenience perhaps shows that the mental properties of an event are *causally relevant* – in the sense that if the event didn't have those properties, it would have had different causal effects. But does that show mental properties to be *causally efficacious*? Consider the epiphenomenalist: he could agree with Davidson that mental properties are indeed causally relevant. If a given mental event had not occurred, that would mean that the physical cause of that event had not occurred, and the physical effects of this event would have not occurred either. But these counterfactuals do not support the claim, the epiphenomenalist will argue, that the mental event was causally efficacious or productive. Or consider this: the aesthetic properties of a work of art supervene on its physical properties, but would that show that aesthetic properties have causal effects in the physical world? And it would seem that Davidson needs causal efficacy, not mere causal relevance, for his theory of action explanation. In order for reasons to generate causal explanations of action, they must be causally efficacious in the production of action, not merely causally relevant in some weaker sense. Symptoms of a disease may be causally relevant to the subsequent development of the underlying pathology, but that would hardly justify the claim that they causally explain the course of the disease. In any case, it is not clear that mind-body supervenience can do the job that Davidson needs to get done.[9]

Turning now to the second question, there evidently is a prima facie tension between psychophysical supervenience and psychophysical anomalism. The main thrust of the supervenience thesis is to bring mental and physical properties closer together, in a dependency relation, whereas the

point of psychophysical anomalism is to push the two apart, so that neither constrains the other. In any case, the answer to the question of consistency depends in part – but crucially – on what one means by supervenience. It is by now fairly standard to distinguish among three notions of supervenience – 'weak', 'strong', and 'global' (Kim 1984a). As global supervenience is not relevant to our present concerns, let us briefly consider the other two supervenience relations. Mental properties can be said to 'strongly supervene' on physical properties if the following condition holds:

> For any mental property *M*, if something *x* has *M*, then there is a physical property *P* such that *x* has *P*, and *necessarily* anything that has *P* has *M*.

Mental properties 'weakly supervene' on physical properties if:

> For any mental property *M*, if something *x* has *M*, then there is a physical property *P* such that *x* has *P*, and anything that has *P* has *M*.

As will be noticed, the only difference between weak and strong supervenience is the presence, in strong supervenience but not in weak supervenience, of the italicized modal term 'necessarily' qualifying the generalization 'Anything that has *P* has *M*'. This means that under strong mind-body supervenience, any mental property *M* that is instantiated has a physical subvenient base property *P* such that, necessarily, if anything has *P*, then it has *M*. Doesn't this make 'If anything has *P*, then it has *M*' a psychophysical law? If so, why can't it also be a "strict" law? One might argue that these "supervenience laws" aren't the kind of psychophysical laws that figure in Davidson's anomalous monism. However, this reply is not plausible in the context of Davidson's overall position on the mind-body problem. For, as we have seen, Davidson's project is the Kantian project of reconciling free agency with physical determinism, and his solution consists in decoupling the mental realm from the physical realm. Is there any reason to think that these supervenience psychophysical necessitations would do any less damage to the autonomy of the mental than strict psychophysical laws?

Davidson himself has said that the supervenience he has in mind is weak supervenience (Davidson 1993b). If weak supervenience is all that is involved, it would generate only de facto psychophysical generations of the form 'If anything has *P*, then it has *M*', and these statements, lacking any modal force, would not count as laws, strict or nonstrict, and would not threaten the autonomy of the mental. Thus, weak supervenience seems to

be fully compatible with anomalous monism. The only question is whether weak supervenience could support even the causal relevance of mental properties. The reason is that in order to be useful for Davidson, the supervenience must generate counterfactuals of the form 'If event *e* had not had mental property *M*, it would not have had physical property *P*'. But weak supervenience gives us only the de facto, nonmodalized conditional 'If anything has *P*, then it has *M*', and it is generally thought that conditionals that hold at least with nomological necessity are needed to support counterfactuals. And without appropriate counterfactuals, it would not be possible to claim even the causal relevance of mental properties and kinds.[10]

Thus, the question is open as to how anomalous monism can, by itself or with suitable strengthening, cope with the problem of mental causation, and there is some doubt whether this can be done. This is not surprising, because Davidson's strict law requirement on causation, combined with his view that strict laws can be found only in basic physics, appears to give the physical domain a special role in shaping the causal structure of the world. In fact, it can be interpreted, or perhaps misinterpreted, as implying that physical causation is the only causation that exists. At least, it seems to have the consequence that physical causation is the foundation of all causal relations. Finding for the mental a useful causal role in this picture, we can be sure, is not going to be an easy matter.

5. FURTHER ISSUES

One of Davidson's trio of papers from the early 1970s whose principal topic is the anomalism of the mental bears the intriguing title "Psychology as Philosophy." The contrast that Davidson presumably had in mind is with psychology as a science. In his early career, Davidson worked on decision theory, coauthoring a monograph entitled *Decision Making: An Experimental Approach* (Davidson and Suppes 1957). Apparently Davidson found this work frustrating, and he came to be convinced that it was not possible to develop a theory of rational decision as a purely descriptive empirical theory testable by the usual method of prediction and observation, a method that is clearly applicable in some areas of psychology (for example, theories of psychophysics, such as Fechner's law). 'Psychology' in the title "Psychology as Philosophy," as Davidson has made clear in various places, refers not to sensory events and processes but to intentional cognitive states, states carrying propositional contents, like belief, desire, and intention. It is these intentional states, states that have a central place in our role as agents and

cognizers, that Davidson has claimed to be beyond the reach of the kind of scientific study familiar from the physical sciences.

As may be recalled, Davidson states the anomalism of the mental in these words: "There are no strict deterministic laws on the basis of which mental events can be predicted and explained" (Davidson 1980c [1970], p. 208). We saw that the qualifier 'deterministic' could be dropped from this statement; what is important is Davidson's denial that there are strict laws – a system of exceptionless laws that impose a complete and comprehensive system of predictive and explanatory connections over the domain of beliefs, desires, intentions, and actions. But does that mean that psychology, as a systematic study of belief, desire, and action, cannot be a "science"? Davidson has later elaborated on this issue in "Could There Be a Science of Rationality?" (Davidson 1995a). As he views it, the three intimately intertwined aspects of mentality – its holism, externalism, and normativity – are the reasons for the irreducibility of psychology to physical theory, and they may suffice to show that intentional psychology cannot aspire to become a closed system of strict laws. But, in "Thinking Causes" and elsewhere, Davidson has argued that there can be, and in fact are, lawlike, though not strict, generalizations about intentional/cognitive phenomena, and it seems that unless we tie our idea of a science to the paradigm of physics, there is no reason not to think that a "scientific" study of such phenomena is possible. In fact, it is doubtful that strict laws can exist in biology, either; as we noted, Davidson himself has said that strict laws perhaps exist only in developed basic physics. And the domain of biological phenomena is no more closed than that of mental phenomena. There seems to be no good reason to deny the appellation "science" to biology or to intentional psychology just because these fields cannot uncover strict laws in their domains. However, intentional psychology may differ from biology in a way that can make a difference: normativity is central and essential to intentional psychology but has no place in biology. What difference does this make to the scientific character of intentional psychology? This is a question worth pondering.

One issue that may turn out to complicate this issue concerns mental irrealism and relativism – more specifically, whether or not Davidson's doctrines might have irrealist and/or relativist consequences concerning the intentional. This concern arises from two sources. One is Davidson's interpretation theory; the second is Davidson's normative view of the intentional. Let me briefly indicate why such a concern might arise. According to Davidson, a person has the belief that *p* if and only if he is interpretable to have the belief that *p*. And to say the latter is to say that there is – in fact, or perhaps in principle – an interpreter who constructs an interpretive system

that makes the best sense of the totality of the subject's speech and behavior, and that this interpretive system attributes to the person the belief that *p*. This raises two questions. The first is whether this leads to a serious relativity in the attribution of beliefs, since which of the possible interpretive systems makes the best sense would seem to depend on who the interpreter is – what kind of cognitive being he is. Remember that the charity principle requires that the interpretive process yield a belief/desire system that closely resembles the interpreter's belief/desire system. Waiving the question where the interpreter gets *his* beliefs and desires, is there any reason to think that the interpreter of whom Davidson speaks is some kind of idealization that in the end cancels itself out? Can we show that any two interpreters will come up with the same interpretive system for the given subject? Second, what if two distinct interpretive systems do equally well in making sense of the totality of our subject's speech and behavior but differ in that one attributes to the subject the belief that *p* and the other makes no such attribution? If that were the case, could there be, according to the account on offer, a "fact of the matter" as to whether or not the subject believes that *p*?

The second source of worry is Davidson's view that the essential function of intentional discourse is normative (Antony 1994). The introduction of the intentional idioms of belief, desire, and intention is mandated by the demands of practical reason, that is, our need to make sense of our status as agents and to rationalize our actions and those of our fellow humans. The function of intentional discourse is not descriptive – predictive or causal/explanatory – but normative. We are apt to take our descriptive idioms to represent our commitments about what is real. True, the normativity of the intentional language does not entail its irrealism, but it does raise questions. What exactly is the descriptive content of this language, and how are its normative aspects related to its descriptive aspects? If beliefs are essentially normative and are posited because of our normative requirement, are there beliefs in the same sense in which there are physical objects and events, like trees and explosions?

Undoubtedly, some of these questions arise from an overly simplistic reading of Davidson's subtle, complex, and in many ways deeply insightful views on mind and intentionality. His contributions in this area are especially noteworthy because they present to us an integrated system of views that illuminates not only the metaphysics of the mind and its relation to the world of matter but also the normative dimensions of mentality that are essential to an understanding of our nature as agents in a natural world.

Notes

1. We will set aside the question whether or not events, phenomena, and the like must be recognized *in addition* to things and their properties and relations. There is a plausible reading of Davidson's account according to which individual events form a basic ontological category, along with material objects. It isn't exactly clear how, in Davidson's ontology, objects and events are related to each other, or whether Davidson himself cares much about such issues. (One natural view is to hold that events, if they exist, supervene on objects and their properties and relations, namely, on facts concerning what objects have what properties and enter into what relations with what other objects.) Just how these metaphysical issues affect issues concerning the mind-body problem is also an unsettled question. My own feeling is that substantive issues about the mind, or about anything else, should not depend on such ontological details and should be invariant across all reasonably rich and flexible ontological schemes.
2. A caveat must be entered here. As some philosophers have noted, Davidson is unlikely to feel comfortable with unrestrained talk of properties; he would perhaps prefer to talk of predicates instead, although, like anyone else, he makes unself-conscious use of expressions like 'property', 'feature', and 'characteristic'. Davidson interpreters may need to be careful about such issues. However, we will carry on our discussion in terms of properties and other reified entities. The reason is that property talk has been the norm, and it is important to appreciate some robustly metaphysical versions of Davidson's many central doctrines.
3. Perhaps including references to the subject's physical environment, physical history, etc. (if content externalism is true).
4. In "Mental Events," Davidson explicitly associates the anomalism of the mental with the failure of nomological reduction of the mental. See Davidson 1980c [1970], p. 216.
5. For analysis and critical discussion of the anomalism of the mental, see Child 1993; Honderich 1981; 1988; Kim 1985; Latham 1999; Lycan 1981; and Yalowitz 1997; 1998.
6. This is why we need the full anomalism of the mental, not just psychophysical anomalism, for this argument, something that isn't evident in Davidson's "Mental Events." If there were purely psychological laws, one of these might subsume the pair *m* and *p*, from which we could infer that *p* must be a mental event.
7. As required by the model of nomological reduction developed by Ernest Nagel (1961). Davidson has never said that he accepts this model of reduction; however, it provides a plausible interpretation of why Davidson invokes psychophysical anomalism to support his claim of the nomological irreducibility of the mental (Davidson 1980c [1970]).
8. Because, in Morgan 1923, C. Lloyd Morgan was already using the term 'supervenience', roughly in its current philosophical sense, to describe the relationship between higher-level phenomena such as mentality and lower-level physical conditions. It is generally believed that the concept originated in ethical theory, more specifically in G. E. Moore's writings (see Moore 1922), although

Moore apparently never used the term 'supervenience'. But this seems not to be entirely correct, in view of Morgan's use of the term (and the concept) in the early 1920s.

9. For further discussion of Davidson's invocation of mind supervenience in this context, see McLaughlin 1993 and Sosa 1993. For earlier attempts to use mind-body supervenience to explain mental causation, see Sosa 1984 and Kim 1984b. For further discussion, see Marras 1999.
10. See Lepore and Loewer 1987 for a sustained effort to use counterfactuals to defend Davidson on this issue of mental causation. Another philosopher who invokes counterfactuals to vindicate mental causation is Horgan (1989).

5 Semantics and Metaphysics of Events

PAUL PIETROSKI

Donald Davidson has deeply influenced contemporary work on the semantics of action sentences and adverbial modifiers, mainly through his seminal papers on event analyses together with his related discussions of actions, events, and causation. This chapter will trace these contributions, beginning with Davidson's semantic proposals and then turning to his metaphysical views.

An "action sentence" such as (1), in which the grammatical subject seems to be associated with a notion of agency, can be paraphrased as in (2).

(1) Brutus stabbed Caesar.

(2) There was a stabbing of Caesar by Brutus.

Intuitively, (2) says that something *happened*: there was an event of a certain sort; in particular, there was a stabbing – a stabbing of Caesar by Brutus. So we can, with a little awkwardness, paraphrase (1) as

(3) For some event *e*, *e* was a stabbing of Caesar by Brutus.

Starting with this observation, Davidson (1980b [1967]) argued that theorists can and should specify the meaning of a sentence such as (1) with an overtly quantificational construction along the lines of (3). If this is correct, it has wide-ranging implications for the study of meaning. It also raises the question, which Davidson (1980 [1969]) addressed, of what events *are*: what kind of thing does one quantify over when one (covertly) quantifies over events?

Sections one and two summarize Davidson's (1980b [1967]) core proposal and his main argument for it. Sections three and four briefly review more recent arguments for an event semantics, along with some reasons for modifying the original proposal in a way that Davidson (1985a) has endorsed. Sections five and six are devoted to ontological questions and Davidson's proposed answers to them, which reveal the unity of his work in semantics, metaphysics, and action theory.

1. ACTION SENTENCES AS EVENT REPORTS

At least for now, let's adopt an ordinary conception of events, according to which an event is something that *happens* at a certain place and time. Suppose that Nora snapped her fingers at noon on January 1, 2001. Then something happened at that time in the region of space then occupied by Nora's fingers. Depending on how precisely we specify the time and place, it may be that many things happened then and there. For example, some blood pulsed through Nora's fingers while she snapped them, although the finger snapping produced a characteristic sound that the blood pulsing did not. This highlights another important feature of events: at least typically, they have *effects*. But in any case, each event takes place somewhere and sometime. And while there may be instantaneous events, most of the events we talk about have a duration. This is especially clear with regard to events like the first transatlantic flight or World War II. As the last example illustrates, our ordinary conception also allows for a "single" complex (and potentially scattered) event with "subparts" – such as the invasion at Normandy – that are themselves events with further subparts.

I'll return to these points and some related questions concerning *actions*.[1] But it should be uncontroversial that

(1) Brutus stabbed Caesar.

is true if and only if history includes an event of the right sort. We can make this point about (1) explicit by inventing a predicate, 'Stacaebrutish', that is satisfied by an event *e* if and only if (iff) *e* was a stabbing of Caesar by Brutus. Then (1) is true iff a Stacaebrutish event occurred. Or more formally, using a sentence of the predicate calculus, we can write this as follows:

'Brutus stabbed Caesar' is true iff $(\exists e)$(Stacaebrutish(e)).[2]

Similarly, we can invent a predicate, 'Stadesothish', that is satisfied by an event *e* iff *e* was a stabbing of Desdemona by Othello. Then

(4) Othello stabbed Desdemona.

is true iff $(\exists e)$(Stadesothish(e)).[3] But these biconditional claims fail to reflect an important fact emphasized by Davidson (1984 [1967]): the truth conditions of declarative sentences depend, in ways we want to know more about, on their constituent words and how those words are arranged.

The word 'stabbed', which appears in both (1) and (4), makes a certain semantic contribution to the sentences in which it appears. This is presumably why (1) and (4) are semantically similar – both imply that

there was a *stabbing* of someone by someone – in a way that (1) and (5) are not.

(5) Othello kissed Desdemona.

Likewise, (4) and (5) are semantically similar – both imply that there was an event in which Othello did something and something happened to Desdemona – in a way that (1) and (5) are not. Appeal to (unary) predicates such as 'Stacaebrutish' and 'Stadesothish' obscures these facts. So while it is an old idea that action sentences somehow describe actions, and that actions are among the things that happen, it is not immediately obvious how to square this idea with semantic *compositionality*.

Moreover, the most obvious way of accounting for compositionality is not by appeal to events. A standard formal representation of (1), familiar to any student of logic, would be

(1′) $\text{Stabbed}_2(\text{Brutus, Caesar})$,

where 'Stabbed_2' is, by stipulation, a binary predicate satisfied by an ordered *pair* $<x, y>$ iff x stabbed y. One can say, plausibly enough, that

'Brutus stabbed Caesar' is true iff $\text{Stabbed}_2(\text{Brutus, Caesar})$.

In which case, (1) is true iff the ordered pair <Brutus, Caesar> satisfies 'Stabbed_2'. Similarly, (4) is true iff <Othello, Desdemona> satisfies 'Stabbed_2'; and (5) is true iff <Othello, Desdemona> satisfies 'Kissed_2', which is satisfied by an ordered pair $<x, y>$ iff x kissed y. This at least begins to show how the truth conditions of action sentences can be determined compositionally; and spelling out the further details is not difficult, given the work of Frege (1960 [1891]), Tarski (1944), and others. (See Larson and Segal 1995 for a clear discussion that takes account of recent work on the syntax of natural languages.)

Davidson (1980b [1967]) showed how to extend this familiar suggestion, about how to account for certain aspects of semantic compositionality, in a way that treats action sentences as reports of event occurrences. We can replace (1′) with

(1a) $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e))$,

where 'Stabbed_3' is, by stipulation, a *ternary* predicate satisfied by an ordered *triple* $<x, y, e>$ iff e was a stabbing of y by x. Then we can say that

'Brutus stabbed Caesar' is true iff $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e))$.

That is, (1) is true iff for some event e, <Brutus, Caesar, e> satisfies 'Stabbed_3'. Similarly, (4) is true iff for some event e, <Othello, Desdemona,

e> satisfies 'Stabbed$_3$'; and (5) is true iff for some event *e*, <Othello, Desdemona, *e*> satisfies 'Kissed$_3$', which is satisfied by an ordered triple <*x*, *y*, *e*> iff *e* was a kissing of *y* by *x*. This suggests an alternative way of showing how the truth conditions of action sentences could be determined compositionally; and spelling out the details is no harder than it would be on the more traditional approach. (Again, see Larson and Segal 1995.)

Still, while introduction of the "event variable" is clever, this doesn't yet constitute an argument that semanticists *should* represent the truth condition for (1) with (1a) as opposed to (1′). Or put another way, it hasn't yet been shown that (1a) specifies the *meaning* of (1). Why prefer (1a) to (1′), especially since (1′) seems simpler and is still compositional? As indicated in previous chapters (see Chapters 1 and 3), Davidson's leading idea is that semanticists specify the meaning of a declarative sentence by correctly specifying its truth condition in terms of a finitely statable truth theory that meets certain formal and empirical constraints. Minimally, the theory must assign correct truth conditions to all of the declarative sentences of the object language (see the Introduction and Chapter 1).

So we need to see why, in light of such demands, (1a) is the better way to represent the truth condition for (1). As Davidson recognized, it is not enough to note that (1) can be paraphrased as

(2) There was a stabbing of Caesar by Brutus.

For paraphrase – intuitive similarity of meaning – is a two-way street. While (2) may well capture the gist of (1), in a way that (5) clearly does not, (1) also captures the gist of (2). So perhaps semanticists should specify the truth condition for (2) as

(1′) Stabbed$_2$(Brutus, Caesar).

Or perhaps semanticists should describe the truth conditions of (1) and (2) differently, even though speakers can use these sentences interchangeably in many situations. In short, even if (1) is true iff $(\exists e)$(Stabbed$_3$(Brutus, Caesar, *e*)), it needs to be demonstrated that this fact about (1) is a *semantic* truth that follows from facts about what the words in (1) mean and how those words are arranged.

Davidson (1980b [1967]) provided an argument, reviewed in the next section, that others have since amplified and supplemented. This is what made the idea of an "event variable" really interesting. But let me end this section by highlighting a respect in which Davidson's hypothesis might initially seem implausible. For he posited a certain mismatch between semantic structure and standard grammatical classifications of verbs. The

English verb 'stabbed' is a transitive verb that combines with *two* (overt) grammatical arguments – a subject and a direct object – to form a sentence. By contrast, an intransitive verb such as 'fell' takes only a subject, while a ditransitive verb such as 'gave' takes three grammatical objects (including an indirect object), as in (6) and (7).

(6) Caesar fell.

(7) Caesar gave Brutus a coin.

So one might have expected a semantic theory for English to treat 'stabbed' as a *binary* predicate, thus specifying the truth condition for (1) as (1′). Similarly, one might have expected 'fell' and 'gave' to be treated as unary and ternary predicates, respectively, as in (6′) and (7′).

(6′) Fell_1(Caesar)

(7′) Gave_3(Caesar, Brutus, a coin)

Here, 'Fell_1' is satisfied by an individual x iff x fell, and 'Gave_3' is satisfied by an ordered triple $<x, y, z>$ iff x gave z to y. But Davidson (1980b [1967]) urges us to think of the English verb 'stabbed' as a semantically *ternary* predicate – the kind of predicate that 'gave' seems to be – with 'fell' corresponding to a binary semantic predicate and 'gave' corresponding to a four-place semantic predicate.[4] On this view, there is more to the semantic structure of an action sentence than one might have expected, and not just because of the covert existential quantifier. The idea, which requires defense, is that action verbs themselves are semantically more complex than initial appearances suggest.

2. ADDING ADVERBS

Davidson thinks that compositionality considerations tell in favor of an event semantics for action sentences, once we consider slightly more complex variants of (1), repeated here.

(1) Brutus stabbed Caesar.

Note that (1) can be extended indefinitely by means of adverbs, or adverbial phrases, as in (8)–(14).

(8) Brutus stabbed Caesar violently.

(9) Brutus stabbed Caesar with a knife.

(10) Brutus stabbed Caesar on the Ides of March.

(11) Brutus stabbed Caesar violently with a knife.

(12) Brutus stabbed Caesar with a knife violently.

(13) Brutus stabbed Caesar violently on the Ides of March.

(14) Brutus stabbed Caesar violently with a knife on the Ides of March.

Sentence (8) is evidently true iff there was a *violent* stabbing of Brutus by Caesar. This suggests that (8) is true iff for some event *e*, *e* was a stabbing of Brutus by Caesar, *and e* was violent. This truth condition is indicated in

(8a) $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e) \,\&\, \text{Violent}_1(e))$,

where 'Violent_1' is a unary predicate satisfied by an event *e* iff *e* was violent. One can think of 'violent' as specifying a *manner* in which some events are done. Similarly, Davidson would say that (9) is true iff

(9a) $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e) \,\&\, \text{With}_2(\text{a knife, } e))$

is true; where to a first approximation, 'With_2' is a binary predicate satisfied by an ordered pair $<x, e>$ *iff* e *was done using* x *as an instrument*. Likewise, he would represent the truth condition for (10) with

(10a) $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e) \,\&\, \text{On}_2(\text{the Ides of March, } e))$,

where to a first approximation, 'On_2' is a binary predicate satisfied by an ordered pair $<x, e>$ iff *e* occurred within the temporal interval *x*. A phrase such as 'in Rome' would specify a *place* where some events occurred, and so on.

Accounting for multiple adverbial phrases presents no difficulty on this view. One can say, plausibly, that (11) is true iff a violent stabbing of Caesar by Brutus was done with a knife; that is, (11) is true iff (11a) is true.

(11a) $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e) \,\&\, \text{Violent}_1(e) \,\&\, \text{With}_2(\text{a knife, } e))$

Intuitively, (12) has the same truth condition as (11). And since the *order* of conjoined predicates is truth conditionally irrelevant, (11a) is truth conditionally equivalent to (12a).

(12a) $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e) \,\&\, \text{With}_2(\text{a knife, } e) \,\&\, \text{Violent}_1(e))$

Likewise, (13) is true iff some violent stabbing of Caesar occurred on the Ides of March; and (14) is true iff some such event was done with a knife. More formally, letting 'IM' stand for the Ides of March, we can represent

the truth conditions of (13) and (14) as in (13a) and (14a).

(13a) $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e) \& \text{Violent}_1(e) \& \text{On}_2(\text{IM, } e))$

(14a) $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e) \& \text{Violent}_1(e) \& \text{With}_2(e, \text{a knife}) \& \text{On}_2(\text{IM, } e))$

This begins, at least, to show how the truth conditions of sentences such as (8)–(14) can be determined compositionally; for details, see Higginbotham 1985; Larson and Segal 1995; Parsons 1990; Taylor 1985.

Davidson's event representations also capture the intuitively compelling *inferential* relations that hold among networks of sentences such as those just discussed. If Brutus stabbed Caesar violently, then it *follows* that Brutus stabbed Caesar. Necessarily, (1) is true if (8) is true; and this seems to be the kind of necessity that attaches to *logical* truths (cf. 'Someone stabbed Caesar if Brutus stabbed Caesar'). It seems that (8) is "inferentially linked" to (1) in a way that (8) is not linked to

(15) Caesar bled.

– even though the truth of (8) makes the truth of (15) overwhelmingly likely. This suggests, though it does not prove, that the inference from (8) to (1) is compelling by virtue of its *form*. So, other things being equal, one would like representations of truth conditions that reveal the inference from (8) to (1) as an instance of some intuitively compelling and truth-preserving inferential form. Thus, it is a virtue of Davidson's proposal that the inference from (8a) to (1a) – $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e) \& \text{Violent}_1(e))$, so $(\exists e)(\text{Stabbed}_3(\text{Brutus, Caesar, } e))$ – is an instance of the form '$(\exists e)(\Phi(e) \& \Psi(e))$, so $(\exists e)(\Phi(e))$'.[5]

In this respect, the inference from (8a) to (1a) is like the inference 'Something is red and shiny, so something is red'. By representing the truth conditions of (1) and (8) in terms of quantification over events, Davidson provides an account of *why* (8) implies (1): (8) says that there was an event with *two* properties – it was a stabbing of Caesar by Brutus, and it was violent; (1) says that there was an event with *one* of those properties – it was a stabbing of Caesar by Brutus. Similar remarks apply to (9)–(14), all of which imply (1).

Moreover, (11) implies both of (8) and (9); and, as noted earlier, each of (11) and (12) implies the other. Yet none of (8)–(10) imply any of the others; and while (13) implies (8) and (10), (13) does not imply (9), since Brutus could have stabbed Caesar violently on the Ides of March with a fork. But (14) implies (8)–(13). Davidson can account for all of these facts.

For example, if (14a) is true, it follows by conjunction reduction that (13a) is true; but the truth of (13a) does not ensure the truth of (9a).

It is worth stressing that on Davidson's view, (8) is true iff *some* stabbing of Caesar by Brutus was violent. Similarly, (9) is true iff *some* stabbing of Caesar by Brutus was done with a knife; and (10) is true iff *some* stabbing of Caesar by Brutus took place on the Ides of March. If (14) is true, then there was an event that "verifies" all four sentences; at least one stabbing of Caesar by Brutus was violent *and* done with a knife *and* on the Ides of March. But (8)–(10) could be true even if (14) were false. Brutus might have stabbed Caesar on three occasions: once violently in January, once with a knife in February, and once with a fork in March. (By analogy, if something is red and shiny, something is red and something is shiny; but the converse does not hold.) This is the key feature of an example, due to Gareth Evans and discussed by Taylor (1985), that highlights the attractions of Davidson's approach.

Suppose that Shem poked Shaun *twice* at the same time – once with a red stick and once with a blue stick. Suppose that these were the *only* pokings of Shaun by Shem, and that the poking with the red stick was forceful, while the poking with the blue stick was gentle. Then (16)–(22) are true:

(16) Shem poked Shaun.

(17) Shem poked Shaun forcefully.

(18) Shem poked Shaun gently.

(19) Shem poked Shaun with a red stick.

(20) Shem poked Shaun with a blue stick.

(21) Shem poked Shaun forcefully with a red stick.

(22) Shem poked Shaun gently with a blue stick.

Assuming that no poking is both forceful and gentle, (23) and (24) are false.

(23) Shem poked Shaun gently with a red stick.

(24) Shem poked Shaun forcefully with a blue stick.

Each of (17)–(24) has characteristic implications that Davidson can explain, as with (8)–(14). Davidson can also account for why (17)–(22) are true while (23) and (24) are false. For let e_1 be the forceful poking of Shaun by Shem with a red stick, and let e_2 be the gentle poking of Shaun by Shem with a blue stick. Then e_1 verifies (17), (19), and (21), while e_2 verifies (18), (20), and (22). But since there was no gentle poking of Shaun by Shem with a

red stick, (23) is false; and since there was no forceful poking of Shaun by Shem with a blue stick, (24) is false.

By contrast, suppose that we represent the truth condition for (16) with

(16′) $\text{Poked}_2(\text{Shem, Shaun})$,

thus treating the English verb 'poked' as a binary predicate satisfied by ordered pairs $<x, y>$ such that x poked y. This notation seems inadequate to the example. One *wants* to say that <Shem, Shaun> satisfies 'poked' twice over; but it makes no sense to say that a single ordered pair satisfies a predicate twice. Whereas if 'poked' is satisfied by ordered triples $<x, y, e>$, we can say that <Shem, Shaun, e_1> satisfies 'poked' *and* that <Shem, Shaun, e_2> satisfies 'poked'.

The advocate of (16′) does, however, have a potential reply. While <Shem, Shaun> does not satisfy 'poked' twice over, this ordered pair does satisfy the *complex* predicates 'poked forcefully' and 'poked gently'. The idea would be that for many predicates **P** – such as 'poked', 'stabbed', and many others – the complex predicate formed by combining **P** with 'forcefully' is satisfied by ordered pairs $<x, y>$ such that x **P** y forcefully. Similarly, one could say that <Shem, Shaun> satisfies the following complex predicates: 'poked with a red stick', 'poked with a blue stick', 'poked forcefully with a red stick', and 'poked gently with a blue stick'. Then one could represent the truth conditions of (17)–(22) as in (17′)–(22′).

(17′) $\text{Poked-Forcefully}_2(\text{Shem, Shaun})$

(18′) $\text{Poked-Gently}_2(\text{Shem, Shaun})$

(19′) $\text{Poked-<With a red stick>}_2(\text{Shem, Shaun})$

(20′) $\text{Poked-<With a blue stick>}_2(\text{Shem, Shaun})$

(21′) $\text{Poked-Forcefully-<With a red stick>}_2(\text{Shem, Shaun})$

(22′) $\text{Poked-Gently-<With a blue stick>}_2(\text{Shem, Shaun})$

And one could say that <Shem, Shaun> does *not* satisfy the following complex predicates: 'poked gently with a red stick' and 'poked forcefully with a blue stick', thus glossing the false (23) and (24) as in (23′) and (24′).

(23′) $\text{Poked-Gently-<With a red stick>}_2(\text{Shem, Shaun})$

(24′) $\text{Poked-Forcefully-<With a blue stick>}_2(\text{Shem, Shaun})$

While this is more cumbersome than the Davidsonian alternative, it is not obviously wrong.

It does, however, fail to provide any explanation for the network of implications. One wants to know why, for example, an ordered pair that satisfies 'Poked-Forcefully$_2$' also satisfies 'Poked$_2$'. The advocate of (16′) and (17′) can say that 'forcefully' is a *restrictive* modifier. But this comes very close to restating the fact to be explained – viz., that anything that satisfies a complex predicate formed by combining **P** with 'forcefully' also satisfies **P**. By contrast, Davidson can account for the implications as instances of conjunction reduction.[6] Moreover, (24) is truth conditionally equivalent to (25).

(25) Shem poked Shaun with a blue stick forcefully.

So without appeal to events, not only must one say that 'forcefully' and 'with a blue stick' are both restrictive modifiers, one must add that the *order* of restriction is truth conditionally irrelevant (at least with respect to verbs such as 'poked'). This extra assumption might seem natural *if* (24) were equivalent to (26).

(26) Shem poked Shaun forcefully, and Shem poked Shaun with a blue stick.

But as the example makes clear, (26) can be true while (24) is false.

This shows that one cannot think of predicate restriction in the following simple way: 'poked' specifies a certain set, S, of ordered pairs; a complex predicate such as 'poked forcefully' or 'poked with a blue stick' specifies a certain subset of S; and a doubly complex predicate such as 'poked forcefully with a blue stick' or 'poked with a blue stick forcefully' specifies the *intersection* of the relevant subsets. On this view, which would ensure the "order invariance" of modifiers, an ordered pair $<x, y>$ satisfies 'poked forcefully with a blue stick' iff $<x, y>$ satisfies both 'poked forcefully' and 'poked with a blue stick'. But this predicts, incorrectly, that (24) and (26) are truth conditionally equivalent. This raises the question of how we *are* supposed to think of predicate restriction, and of *why* sentences with restricted predicates exhibit the implication patterns just discussed. Insofar as such questions are left unanswered on nonevent views, Davidson has an argument for his proposed answer in terms of quantification over events.

3. OTHER EVIDENCE

There are also other reasons for adopting Davidson's hypothesis about the semantics of action sentences. This is not the place to review the extensive

literature on this topic; see Larson and Segal 1995; Parsons 1990; Pianesi and Varzi 2000; Taylor 1985. Still, a few examples may give the reader a feel for the kinds of considerations – some suggested by Davidson (1985a) – that have been brought to bear on the issue.

We can specify the truth condition for (27) with (27a).

(27) Brutus fled after he stabbed Caesar.

(27a) $(\exists e)(\text{Fled}_2(\text{Brutus}, e) \ \& \ (\exists f)(\text{After}_2(e, f) \ \& \ \text{Stabbed}_3(\text{he}, \text{Caesar}, f)))$

That is, (27) is true iff an event of Brutus's fleeing occurred after an event of his stabbing Caesar. Thus, a Davidsonian gloss of (27) lets us treat 'after' as a binary predicate satisfied by ordered pairs of events – things that happen in space-time. Similarly, (28) has the truth condition indicated by (28a).

(28) Brutus dropped the knife before Caesar died.

(28a) $(\exists e)(\text{Dropped}_3(\text{Brutus}, \text{the knife}, e) \ \& \ (\exists f)(\text{Before}_2(e, f) \ \& \ \text{Died}_2(\text{Caesar}, f)))$

One can also paraphrase (28) by using overt event nominals, as in (29).

(29) *A dropping of the knife by Brutus* occurred before *a death of Caesar*.[7]

This bolsters the idea that 'Brutus dropped the knife' and 'Caesar died' covertly involve the kind of quantification over events that is overt in (28a). For the italicized expressions in (29) seem to be event descriptions; and given the presence of words such as 'before' and 'occurred', it is hard to see how one can deny these appearances. Likewise, (30) and (31) are nearly synonymous:

(30) The prince won the race, and then a court jester stole the prize.

(31) There was a winning of the race by the prince, and then there was a stealing of the prize by a court jester.

This suggests that (30) and (31) both involve quantification over events.

Another source of evidence stems from sentences used to make perceptual reports. Notice that (32) differs in various ways from (33).

(32) Cassius saw Brutus flee.

(33) Cassius saw that Brutus fled.

In (32), 'flee' is untensed, and substituting coreferential expressions for 'Brutus' preserves truth. If Brutus was the noblest Roman, the truth of (32) ensures that Cassius saw the noblest Roman flee; but the truth of (33)

does not ensure that Cassius saw that the noblest Roman fled. So following Higginbotham (1983) and Vlach (1983), one might render (32) as

(32a) $(\exists e)(\exists f)(\text{Saw}_3(\text{Cassius}, f, e) \ \& \ \text{Flee}_2(\text{Brutus}, f))$,

which is true iff there was a seeing (e) by Cassius of a fleeing (f) by Brutus. Similar remarks apply to (34) and (35).

(34) Nora heard Nick shout.

(35) Nora heard Nick shout in the hallway.

This provides a basis for explaining the ambiguity of (35), which can mean either that Nora was in the hallway when she heard Nick's shouting, or that she heard Nick's shouting, which was coming from the hallway. This suggests that 'in the hallway' can be predicated of the e-position event (the hearing) *or* of the f-position event (the shouting), as indicated in (35a) and (35a′).

(35a) $(\exists e)(\exists f)(\text{Heard}_3(\text{Nora}, f, e) \ \& \ \text{Shout}_2(\text{Nick}, f) \ \& \ \text{In}_2(\text{the hallway}, e))$

(35a′) $(\exists e)(\exists f)(\text{Heard}_3(\text{Nora}, f, e) \ \& \ \text{Shout}_2(\text{Nick}, f) \ \& \ \text{In}_2(\text{the hallway}, f))$

Moreover, these various considerations interact, as suggested by (36).

(36) Nora heard Nick shout loudly in the hallway before seeing him leave quickly.

Those who wish to avoid appeal to events need to provide theoretically perspicuous semantic representations that help to explain the relevant implications (and paraphrases) for each reading of (36).

4. A POSSIBLE MODIFICATION

While Davidson (1980b [1967]) held that 'Stabbed_2(Brutus, Caesar)' is an inadequate semantic representation of (1), he did not deny that (1) is true iff Stabbed_2(Brutus, Caesar).

(1) Brutus stabbed Caesar.

Indeed, as Parsons (1990) notes, one can view Davidson as *supplementing* the traditional proposal with the following thesis:

$$(\forall x)(\forall y)(\text{Stabbed}_2(x, y) \text{ iff } (\exists e)(\text{Stabbed}_3(x, y, e)))$$

This raises the question of whether further supplementation is desirable. For as Castañeda (1967) pointed out in response to Davidson's original paper, we can specify the truth condition for (1) with

(1b) $(\exists e)(\text{Agent}_2(\text{Brutus}, e) \,\&\, \text{Stabbed}_1(e) \,\&\, \text{Patient}_2(\text{Caesar}, e))$,

where 'Stabbed$_1$' is a unary predicate of events, and to a first approximation: 'Agent$_2$' is a binary predicate satisfied by ordered pairs $<x, e>$ such that x is the individual who (typically by acting) brought e about; while 'Patient$_2$' is a binary predicate satisfied by ordered pairs $<y, e>$ such that y is the individual saliently affected by the end of e. This is to treat 'Agent$_2$' and 'Patient$_2$' as labels for so-called thematic relations, which correspond to ways in which individuals can *participate* in events. And plausibly, there was a stabbing of y by x iff there was a stabbing, and it was done by x, and it was done to y. That is,

> $(\forall x)(\forall y)((\exists e)(\text{Stabbed}_3(x, y, e))$ iff $(\exists e)(\text{Agent}_2(x, e) \,\&\, \text{Stabbed}_1(e) \,\&\, \text{Patient}_2(y, e)))$.

This further supplementation, which Davidson (1985a) adopted, makes it easy to explain why it follows from (1) that Brutus did something *and* that there was a stabbing *and* that something happened to Caesar. It also preserves a sense in which the transitive verb 'stabbed' is semantically binary. For while verbs are treated as unary predicates of events on this view, 'stabbed' is associated with *two* thematic roles. By contrast, an intransitive verb such as 'fell' is associated with one role, and a ditransitive verb such as 'give' is associated with three thematic roles. For example,

(7) Caesar gave Brutus a coin.

is arguably true iff $(\exists e)(\text{Agent}_2(\text{Caesar}, e) \,\&\, \text{Gave}_1(e) \,\&\, \text{Recipient}_2(\text{Brutus}, e) \,\&\, \text{Patient}_2(\text{a coin}, e))$.[8] Many theorists have also held that slightly more complex sentences tell in favor of "thematically elaborated" event analyses. Again, this is not the place for a review of details; see Parsons 1990; Pietroski forthcoming-b; Schein 1993). But let me mention one kind of example, discussed by contemporary Davidsonians, that raises interesting issues about how to represent truth conditions.

Consider the collective-subject reading of the plural subject in (37).

(37) Two linguists pushed five pianos.

On this reading, two linguists *together* pushed five pianos. (This contrasts with a distributive reading according to which *each* of two linguists pushed

five pianos, for a total of ten piano-pushings.) Potential semantic representations of this reading include (37′), (37a), and (37b).

(37′) Pushed_2(two linguists, five pianos)

(37a) $(\exists e)(\text{Pushed}_3$(two linguists, five pianos, $e))$

(37b) $(\exists e)(\text{Agent}_2$(two linguists, $e)$ & $\text{Pushed}_1(e)$ & Patient_2(five pianos, $e))$

On the assumption that 'Pushed_2' is satisfied by ordered pairs $<x, y>$ such that x pushed y, where 'x' and 'y' range over individuals, advocates of (37′) are committed to saying that 'two linguists' and 'five pianos' can be used to describe "plural" individuals. A standard expression of this idea is that (37) is true on its collective reading iff: for some collection λ^2 consisting of two linguists, and some collection Π^5 consisting of five pianos, the ordered pair $<\lambda^2, \Pi^5>$ satisfies 'Pushed_2'. Perhaps one can extend this idea to an event analysis by saying that (37) is true iff for some event e, the ordered triple $<\lambda^2, \Pi^5, e>$ satisfies 'Pushed_3' – or that (37) is true iff $(\exists e)(\text{Agent}(\lambda^2, e)$ & $\text{Pushed}_1(e)$ & $\text{Patient}(\Pi^5, e))$. But one might wonder if plural objects can be event *participants* of the right sort. (Can λ^2 really *bring about* an event of pushing, even if λ^2 is not itself a person with a mind of its own? And if not, can there really be a pushing of Π^5 *by* λ^2?) In any case, advocates of a plural-object semantics have often eschewed events; see Landman 1996 for a review.

One can, however, adopt (37b) without supposing that plural objects are agents. Schein (1993) argues that one should interpret 'Agent_2(two linguists, e)' as follows: two linguists were *the agents* of e; each of two linguists was *an* agent of e – each linguist contributed to bringing about e – and e had only two agents. (Like Schein, I assume that if two linguists and eight philosophers each contributed to the pushing of five pianos, then (37) is not true.) This raises the question of what it is for an event to have multiple agents, as opposed to having a plural object as its unique agent. But recall that our intuitive conception of events allows for complex events with subparts; think of a banquet at which 100 guests eat 30 chickens. So perhaps an event e has (exactly) two linguists as its agents and (exactly) five pianos as its patients iff: each of two linguists is the agent of a subpart of e, and no one else is an agent of a subpart of e; and each of five pianos is the patient of a subpart of e, and nothing else is a patient of a subpart of e. The idea would be that e – an event in which two linguists did some pushing and five pianos got pushed – can be divided into two subparts along its agent dimension, while e can be divided into five subparts along its patient dimension; see also Carlson 1984.

A simple way to formalize this idea is to quantify over both individuals and sets (or collections) of those individuals. One might say that (37) is true iff for some event *e*: there is a two-membered set *S* of linguists, such that *x* is a member of *S* iff *x* is the agent of a subpart of *e*; *and e* was a complex event of pushing; *and* there is a five-membered set *S′* of pianos, such that *y* is a member of *S′* iff *y* is the patient of a subpart of *e*. Alternatively, following Boolos (1985), one can avoid explicit quantification over sets by employing a formal metalanguage that allows for (quantificational) second-order predicates, where such predicates are understood in terms of English plural constructions. Either way, this idea is at odds with Davidson's preference for more traditional first-order representations of truth conditions; see Davidson 1984b, pp. xv–xvi. But the theoretical motivations for this preference are unclear. And the preference seems indefensible, as Boolos notes, given sentences such as 'There are some horses that are faster than Zev and also faster than the sire of any horse that is slower than them'. (See Boolos 1984; 1985 for penetrating discussion and further references; see also Rescher 1962; Wiggins 1980.)

Moreover, Schein (1993) argues that appealing to plural objects in the semantics of plural constructions makes it hard to explain the truistic character of conditionals such as (38) and (39).

(38) If the boys sang, then there is a boy.

(39) If every one of the boys sang, then every boy sang.

Prima facie, one would expect a semantic theory to account for the intuitively compelling inferential relations exhibited by such conditionals (cf. 'If Brutus killed Caesar violently, then Brutus killed Caesar'). Given a plural object theory, such an account will have to be based on principles that *ensure* that if the boys sang, then some plural object – the boys – has at least one element, and each element of it is a boy. It is not enough just to *say* that 'the boys' describes a plural entity such that some boy is "one of" it and each "one of" it is a boy. An *explanation* requires principles that exclude the following logical possibilities, which would falsify (38) and (39): there is a plural object designated by 'the boys', and it satisfies 'sang', but nothing is a boy; or there is no plural object designated by 'the boys', in which case the antecedent of (39) is vacuously true, but there is a single nonsinging boy.

The plural object theorist can try to formulate axioms (about the "one of" relation that allegedly holds between a plural object and its elements) that ensure the truth of conditionals such as (38) and (39). But Schein, drawing on Boolos, argues that this cannot be done consistently. One cannot

appeal to a general principle such as (40).

(40) If there is a Φ, there is something – the Φs – such that every Φ is one of it, and every one of it is a Φ.

Given a boy, it follows from (40) that there is something – the boys – such that every boy is one of it, and every one of it is a boy. But likewise, given a thing that is not one of itself, it follows from (40) that there is something *X* such that every thing that is not one of itself is one of *X*, and every one of *X* is a thing that is not one of itself. Thus, (40) leads to a paradox: if *X* is not one of itself, then *X* is one of *X*; yet if *X* is one of itself, and so *X* is one of *X*, then *X* is not one of itself. Hence, theorists cannot base explanations on (40).

This does not prove that plural object theorists cannot explain the truistic character of (38) and (39) in some other way. Perhaps a correct resolution of the semantic paradoxes will point the way to a true and explanatory variant of (40). But as things stand, this is just speculation. By contrast, on a Schein-style semantics: if $(\exists e)(\text{Sang}_1(e)$ & $\text{Agent}_2(e$, the boys)), then each boy is the agent of some subpart of *e*; in which case, there is a boy. Similarly, if every one of the boys is the agent of a singing – for every *x* such that *x* is a boy, $(\exists e)(\text{Agent}_2(x, e)$ & $\text{Sang}_1(e))$ – then every boy sang. So we have a prima facie reason for adopting a thematically elaborated event analysis, even if this requires construing '$\text{Agent}_2(e$, the boys)' as a second-order predicate of events (without appeal to plural objects).

Schein (1993) also argues for this conclusion via constructions such as

(41) Three linguists taught four philosophers five theories.

While (41) is multiply ambiguous, we can focus on the following reading: three linguists (collectively) taught each of four philosophers five theories (for a total of twenty episodes of theory-teaching). This suggests a complex event of teaching in which three linguists were the tutors, four philosophers were the pupils, and each pupil learned five theories. Or put another way, (41) is true on the relevant reading iff for some event *e*: three linguists were the agents of *e*; *and e* was a complex event of teaching; *and* four philosophers were the recipients of *e*; *and* each recipient of *e* was involved in five subevents of learning a theory.[9] Spelling out this argument for "thematic separation" takes care; but Schein argues that no plural object analysis will correctly capture the relevant reading of (41). Arguments of this sort suggest that Castañeda (1967) was right to propose semantic representations such as (1b).

(1b) $(\exists e)(\text{Agent}_2(\text{Brutus}, e)$ & $\text{Stabbed}_1(e)$ & $\text{Patient}_2(\text{Caesar}, e))$

5. ACTIONS AND THE ACCORDION EFFECT

Let us now return to the simple action sentence

(1) Brutus stabbed Caesar.

and consider a question not yet explicitly addressed: assuming that actions are events of some kind, *which* events are actions? If we represent the truth condition for (1) with (1b), then an obvious thought is that events with agents – events of stabbing, pushing, poking, and so on – are the actions. This fits with the idea that verbs such as 'stab' ('push', etc.) are action predicates. But one can still wonder whether actions are mental events ("volitions" that often cause the motions of limbs), bodily movements, processes that often include certain effects of bodily movement, some mixture of these, or something else entirely. No answer is without difficulties, if only because of some initially puzzling facts about the *number* of things that agents do. But Davidson provided an influential defense of the claim that actions are bodily movements.

As many theorists have noted, a person often does one thing by doing another; and in such cases, it seems that the person does two things by performing one action.[10] For example, Brutus killed Caesar by stabbing him. So (1) is true, as is (42).

(42) Brutus killed Caesar.

But assuming that Brutus stabbed Caesar exactly once, it seems wrong to say that (1) and (42) report different actions. Brutus did not perform two actions; he stabbed *and thereby* killed Caesar. Brutus did not have to do anything *other than* stab Caesar in order to kill him. Nonetheless, (1) and (42) have different truth conditions. Not all events of stabbing are events of killing (or vice versa). Indeed, even with respect to Brutus and Caesar, there is at least some reason to deny that the event of stabbing *was* the event of killing. For presumably, Caesar lingered on a bit after being stabbed. Thus, (43) is true. By contrast, (44) seems wrong.

(43) Caesar died after Brutus stabbed Caesar.

(44) Caesar died after Brutus killed Caesar.

And if Caesar's death occurred after the stabbing but not after the killing, then the stabbing was not the same event as the killing. Any account of what actions are must come to grips with such facts.

Consider, in this regard, one of Davidson's (1980 [1963]) examples. If you illuminate a room (by turning on the light) by flipping a switch (by moving your finger), there is a sense in which you do just one thing. Yet in another sense, you do several things. While there is pressure to identify the event of illuminating the room with the event of flipping the switch, 'illuminate' and 'flip' clearly have different extensions; and even in the case at hand, it seems that the event of illuminating the room somehow involves bulbs and wires – and thus has slightly different spatiotemporal properties than the event of flipping the switch. Davidson's strategy for dealing with these facts is to say that there is just one action – your bodily movement. But this action can be described in many ways, and you "do several things" in the sense of bringing about several effects.

According to Davidson, "We never do more than move our bodies: the rest is up to nature" (Davidson 1980a [1971], p. 59). On this view, Brutus's action of killing Caesar *was* the relevant event of Brutus moving his hand (which held the knife). While killing someone differs from stabbing someone, it does not follow that no stabbing is a killing. By analogy, being a cousin differs from being a son, but it doesn't follow that no cousin is a son. As Davidson notes, the difference between killing someone by stabbing him and stabbing someone without killing him lies with the *stabbee*: did he die from the wound or not? And even if the wound is fatal, the *stabber* makes his causal contribution prior to the death. Davidson holds that similar remarks apply to anything that happens *as a result of* an agent's bodily movements. If you shoot an arrow, your causal contribution is over before the arrow reaches its target; and even if you move a hand that holds a knife, intending to thrust the knife into something else, the knife reaches the intended target only if it does not fly out of your hand (and the target is not armored).

Given this "agent-located" conception of action, Davidson needs to say how events of stabbing (killing, pushing, etc.) can be actions. His claim is that each such event is indeed a movement of the relevant agent's body – letting mental actions such as plotting revenge (or adding numbers "in one's head") count as "internal" movements of the body. On this view, a killing is not an event that somehow includes a death; a killing is a bodily movement that is suitably related to a death. Similarly, a stabbing is a bodily movement with a certain kind of effect; and a bodily movement can have such an effect *and also* be suitably related to a death. In the case of Brutus, one and the same bodily movement was both his action of stabbing Caesar and his action of killing Caesar; hence, the killing *was* the stabbing, which *was* a certain movement of Brutus's body.

The idea is that we often *describe* actions (and other events) in terms of their salient effects, much as we describe some things (such as sunburn) in terms of their salient causes. We *call* certain actions killings, as opposed to stabbings or hand movements, in order to highlight their causal consequences. But on Davidson's view, to say that an action is a killing is to ascribe a *relational* property to a bodily movement. This is, in many respects, an attractive view. But it still has the uncomfortable – and perhaps false – consequence that (44) is true. For Caesar's death occurred after Brutus's action. Similarly, if killings are actions, Davidson is committed to the truth of (45).

(45) The killing of Caesar caused the death of Caesar.

Davidson tries to make this more palatable by noting that one can say (somewhat theatrically) 'You have killed him' when describing the action of someone who has inflicted (what will turn out to be) a fatal wound, even before the victim has died. And one can link this point to a proposal to be (discussed shortly) about the semantics of '*x* killed *y*'. But the felt wrongness of (44) and (45) remain. And there are ways to avoid the most counterintuitive aspects of Davidson's view.

Feinberg (1965), drawing on Anscombe (1957), spoke of the "accordion effect" that attends action. We can say what Brutus did by focusing narrowly on the movement of his hand, by extending our focus a little to include the knife and its entry into Caesar, or by extending our focus still farther to include the subsequent death of Caesar. Davidson (1980a [1971]) says that this is "an important feature of *the language we use to describe* actions" (p. 53, my italics). But perhaps our language correctly reflects the events described. We can speak of a complex event – or a *process* – that is the toppling of ten dominoes, where the first toppling causes the second, and so on. Similarly, many theorists have held that an event of killing is a complex event/process that *starts* with a bodily movement and *ends* with a death; see, for example, Thomson 1977. One can add that killings are actions, albeit "nonbasic" actions that are always done by moving one's body; or one can retain Davidson's view that all actions are bodily movements, while saying that killings start with actions (instead of saying that killings are actions). While these are partly terminological matters, there is a substantive issue concerning how a Davidsonian semantics of action sentences should be combined with a Davidsonian metaphysics of (basic) actions.

One might – and Davidson (1985a) does – try to defend the claim that killings cause deaths by reflecting on the fact that (42) implies (46), much as (47) implies (48), where the subscripts indicate transitive and intransitive

forms of 'boiled':

(42) Brutus killed Caesar.

(46) Caesar died.

(47) Pat boiled$_T$ the soup.

(48) The soup boiled$_I$.

Many verbs exhibit this pattern, in which instances of 'x V_T y so y V_I' are truistic; consider 'freeze', 'break', 'shatter', and so on. Such verbs are often called *causatives*, because (speakers take it to be obvious that) x caused y to V_I if x V_T y. The converse does not hold; there are cases in which x caused y to V_I (in some indirect way) but x did not V_T y.[11] As a first approximation, though, one might say that (47) is true iff

$(\exists e)(\text{Agent}_2(\text{Pat}, e)$ &
$(\exists f)(\text{Caused}_2(e, f)$ & $\text{Boiled}_I(f)$ & $\text{Patient}_2(\text{the soup}, f)))$.

That is, Pat was the agent of some event that caused a boiling$_I$ of the soup. Similarly, one might say that (42) is true iff Brutus was the agent of some event that caused a death of Caesar:

$(\exists e)(\text{Agent}_2(\text{Brutus}, e)$ &
$(\exists f)(\text{Caused}_2(e, f)$ & $\text{Died}_I(f)$ & $\text{Patient}_2(\text{Caesar}, f)))$.

And if one assumes that x is the agent of e iff e is a basic action performed by x, one might well conclude that the killing of Caesar by Brutus was an event that caused Caesar's death. But if x can be the agent of complex accordion-style events, each of which *starts with* a basic action performed by x, one can say that (42) is true iff

$(\exists e)(\text{Agent}_2(\text{Brutus}, e)$ &
$(\exists f)(\text{Ended-with}_2(e, f)$ & $\text{Died}_I(f))$ & $\text{Patient}_2(\text{Caesar}, e))$.[12]

The difference between these views is relevant to a case due to John Wallace and discussed by Parsons (1990). Suppose that Slim sinks the seven ball in the side pocket and the eight ball in the corner, in succession, with a single movement of his body (while holding a pool cue). At least initially, it seems that on Davidson's view: Slim's action is both a sinking of the seven and a sinking of the eight; in which case, the sinking of the seven *is* the sinking of the eight, despite the apparent differences. Davidson replies (1985a), plausibly enough, that phrases such as 'into the corner pocket' are not predicates of actions. But he concludes that we should analyze 'x hit the eight ball' along the following lines, where it is understood that one ball

can hit another: *x* did something that caused a hitting of the eight ball. Yet even granting that this analysis is correct (and extendable to other cases), it is hard to see why a hitting of the eight ball by the cue ball should count as an "into-the-corner-pocket" event, if Slim's action does not. Appeal to accordion-style events lets one say that Slim's sinking of the seven (which was over when the seven-ball sank) was an event distinct from Slim's sinking of the eight (which was not over until the eight ball sank), while *also* saying that both complex events *started with* the same basic action (which was Slim's action of striking the cue ball). The question is whether such appeal to such events – in addition to complex event descriptions – is warranted, all things considered.

6. COUNTING BASIC EVENTS

As we have seen, one might allow for various kinds of complex events: historic episodes such as World War II; "accordion-style" events, such as the toppling of ten dominoes; and "collective" events, with multiple agents and/or patients, of the sort discussed in connection with plurality. But such events seem to be composed, in one way or another, of simpler events. And one wants to know what the "basic" events are.[13] There are, however, at least two questions here: which events are the basic ones, and what kind of thing is an event? With regard to the first question, Davidson holds that all events – or at any rate, all events that enter into causal relations – are *physical* events (in a fairly demanding sense of 'physical'). More about this in a moment. With regard to the second question, one might say that events are metaphysically primitive; and some of Davidson's remarks suggest that this is his view. On the other hand, the crucial features of events seem to be their spatiotemporal and causal properties; and this suggests two more substantive answers to the second question, both of which Davidson has explored.

One might think that events are individuated by regions of space-time. On this view, which is associated with Quine (1960),

$$(\forall e)(\forall f)(e = f \text{ iff } e \text{ has the same spatiotemporal properties as } f).$$

This is an analog of a familiar claim about objects: there cannot be two objects in the same place at the same time. But this claim is tendentious.[14] And it is not obvious that two events never occupy the same region of space-time. Consider the much discussed example of a ball that is spinning while it is also warming up. The event of the ball spinning and the event of the ball

warming up – from, say, 1:00 p.m. to 1:01 p.m. – evidently occupy the same region of space-time. Yet it seems that the spinning can have effects (and causes) that *differ* from those of the warming. In which case, the spinning and the warming are distinct events.

Motivated by such considerations, one might follow Davidson (1980a [1967]) in saying that events are individuated by their causes and effects. That is,

$$(\forall e)(\forall f)(e = f \text{ iff } e \text{ has the same causes and effects as } f).$$

A counterexample to this claim would be a pair of distinct events with all the same causes and effects. But if we set aside the possibility of events with no causes or effects – if only because, by hypothesis, such events never have an impact on our experience – it is hard to see how spatiotemporally distinct events could have the same causes and effects; and it is hard to see how there could be any meaningful distinction between spatiotemporally identical events that are also causally indistinguishable. But since causes and effects are themselves events, on Davidson's view, his proposed "identity criterion" does not reduce events to anything else. In this sense, he is willing to adopt events as metaphysically primitive.[15]

Davidson does, however, have something to say about causation; see, for example, his Davidson 1980a [1967]. Unsurprisingly, he takes it to be a relation that holds between events, as suggested by sentences such as

(49) The event of Brutus stabbing Caesar caused Caesar's death.

With respect to such sentences, substitution of coreferential terms evidently preserves truth value: such substitutions transform true sentences into true sentences, and false into false. For example, if Brutus was the noblest Roman and Caesar was the greatest emperor, then (49) is true iff (50) is true.

(50) The event of the noblest Roman stabbing the greatest emperor caused Caesar's death.

By contrast, while *the fact that* Brutus stabbed Caesar may well *explain why* Caesar died, *the fact that* the noblest Roman stabbed the greatest emperor does not yet *explain why* Caesar died. This bolsters Davidson's claim that causation is a relation between spatiotemporal particulars, which can be described in many ways, and that events stand in causal relations without any relativization to descriptions of (or facts concerning) those events. By contrast, if we speak of explanation as a relation between events, we must speak of it as a description-relative relation. If e_1 is the event of Brutus stabbing Caesar, and e_2 is Caesar's death, then e_1 explains e_2 relative to the

event descriptions in (49) but not relative to those in (50). But for Davidson, causation is not relative to a description; e_1 caused e_2, period. Correlatively, e_1 and e_2 themselves satisfy the descriptions in both (49) and (50); see also Child 1994; Pietroski 2000; Strawson 1985.

Kim (1993b) and others have challenged this conception of events on the basis of examples such as the following: a bridge collapses because a bolt suddenly gave way; but the bridge would not have collapsed if the bolt had given way less suddenly. According to Davidson, the event of the bolt giving way *was* the event of the bolt giving way suddenly; for the event of the bolt giving way was, in fact, sudden. Thus, Davidson is committed to both of the following:

(51) The event of the bolt giving way caused the collapse of the bridge.

(52) The event of the bolt giving way suddenly caused the collapse of the bridge.

Kim would say that only (52) is really true. Davidson's (1980a [1967]) reply would be that (51) is true, but that we must take care to distinguish causation from explanation. For in the case described, (53) is false, while (54) is true.

(53) The fact that the bolt gave way explains why the bridge collapsed.

(54) The fact that the bolt gave way suddenly explains why the bridge collapsed.

And this may well account for our inclination to *use* (52) rather than (51).

Given this kind of response to Kim, and Davidson's emphasis on the idea that events are spatiotemporal particulars (describable in many ways) that exhibit description-insensitive causal relations, one might wonder why he rejected the spatiotemporal criterion of event individuation. Why not say that the *event* of the ball spinning (from 1:00 p.m. to 1:01 p.m.) was indeed the *event* of the ball warming up, but *the fact that* the ball was spinning explains facts not explained by *the fact that* the ball was warming up? Perhaps all that differs here are the descriptions. In response to Quine, Davidson (1985f) revisited these questions and tentatively concluded that a spatiotemporal criterion might be acceptable after all. The intuition remains that the spinning and the warming have different causes/effects; but perhaps this intuition (along with the intuition that a killing of *y* cannot precede the death of *y*) can be explained away by attention to the difference between events and event descriptions. Davidson did not address this issue again. But it is closely related to the question of what causation is and the question of which events are the basic ones.

So let me conclude by noting that Davidson does have views, most famously expressed in his Davidson 1980c [1970], on these large questions. He holds that all causes – or better, all *basic* causes – satisfy predicates that figure in systems of "strict" laws (i.e., laws without "ceteris paribus" clauses that cover factors not captured by the relevant system of laws). Davidson also claims, plausibly enough, that only predicates from the physical sciences figure in strict laws; and so he concludes that all causes are physical causes. Spelling out this view, and assessing its plausibility, is a topic for another chapter.[16] For these purposes, the important point is that Davidson's claims about events are embedded in an overall conception of causal relations that includes claims about facts, explanations, laws of nature, and mind-body relations; consider, in this regard, the remarkably broad yet integrated range of ideas found in Davidson 1980a [1967]; 1980b [1967]; 1980 [1969]; 1980c [1970]; 1980a [1971]; 1984 [1967]; 1984 [1968]. The methodological suggestion is that claims about "what events are" need to be evaluated in light of (i) our best theories of meaning for the languages we speak, and (ii) our best conceptions of the characteristics – intrinsic and relational – that events appear to exhibit. For ordinary claims have implications about events; and claims about events are in turn crucially related to how we think about causation, space-time, ourselves, and how we are related to the physical world that we often talk about and occasionally comprehend. Davidson thus shows how apparently narrow and technical questions about the semantics of natural language sentences can bear on the more traditional questions of philosophy.

Notes

1. Davidson (1980 [1963]) argued that the actions of persons, such as Nora's action of snapping her fingers, are events with mental (and rationalizing) causes (see Chapter 2). Yet even granting that actions are things that happen, it remains an open question whether Nora's action *was* the relevant motion of her fingers, or a mental episode of "willing" her fingers to move, or something else (perhaps a complex event with subparts).
2. For present purposes, I ignore tense and relativization of the truth predicate to a language. This will not affect the substance of the following discussion.
3. See Chapter 1 for a discussion of Davidson's program of truth-theoretic semantics and the importance of representing the compositional structure of natural language sentences. For these purposes, I won't worry about the distinction between *representing* the truth conditions of natural language sentences with sentences of a formal language (whose predicates are not vague) and *regimenting* the natural language sentence. See Pietroski forthcoming-a for a discussion of the possibility that a semantic theory will not associate (contextualized utterances

of) unregimented declarative sentences with compositionally determined *truth* conditions.

4. I will come back to sentences such as (6) and (7). Davidson (1980b [1967]) would represent their truth conditions as in (6a) and (7a).

   (6a) $(\exists e)(\text{Fell}_2(\text{Caesar}, e))$

   (7a) $(\exists e)(\text{Gave}_4(\text{Caesar}, \text{Brutus}, \text{a coin}, e))$

   Davidson 1985a adopts a modified proposal to be discussed in §4.

5. Think of instantiating 'Φ' with 'stabbing of Caesar by Brutus' – i.e., 'Stacaebrutish'. And note that any instance of conjunction reduction in the scope of an existential quantifier is formally valid. This is not to say that all adverbs can be viewed as predicates of events. Davidson (1980b [1967]) suggests that we gloss 'Brutus intentionally stabbed Caesar' along the lines of 'It was intentional of Brutus that he stabbed Caesar' (or 'Brutus intended to stab Caesar and did so'), with 'intend' taking scope over an entire embedded clause; cf. 'It was believed by Brutus that he stabbed Caesar'. Parsons (1990) provides an event accounts of 'She opened the door halfway', which does not mean than an event of opening was halfway; see also Parsons's account of progressive constructions such as 'He was crossing the street when he was struck by a bus'.

6. Cf. Montague 1974. One could say that (19) is true iff Poked-With_3(Shem, Shaun, a red stick); see Kenny 1963. But as Davidson (1980b [1967]) noted, pursuing this strategy requires an open-ended number of complex predicates that take differing numbers of arguments: 'Shem poked Shaun with a red stick at noon' is true iff Poked-With-At_4(Shem, Shaun, a red stick, noon), etc. And without meaning postulates that effectively encode the relevant implications, the truth of 'Poked-With-At_4(Shem, Shaun, a red stick, noon)' does not ensure the truth of 'Poked-With_3(Shem, Shaun, a red stick)' or 'Poked_2(Shem, Shaun)'. See Taylor (1985), who provides another argument for an event semantics, based on the ambiguity of 'Henry gracefully ate all the crisps' – which could mean either that *each* eating of a crisp by Henry was graceful, or that the complex event of eating all the crisps was graceful. See also the following discussion of plurality.

7. In order to see that it should be 'a death of Caesar' rather than 'Caesar's death', we only have to notice that (28) could be true even if reincarnation were possible.

8. Though it is sometimes hard to say *which* thematic roles a given verb has; see Larson and Segal 1995 for an introduction to this issue. One could represent the truth condition for 'there was a stabbing' with $(\exists x)(\exists y)(\exists e)(\text{Stabbed}_3(x, y, e))$. But as Parsons (1990) notes, this is to assume (tendentiously) that if there was a stabbing, it *follows* that someone stabbed someone.

9. Note that 'for four philosophers x, $(\exists e)$(Taught(three linguists, x, five theories, e))' fails to require that the same three linguists did all of the teaching. One can impose this requirement as follows: for some plurality x of three linguists, there are four individual philosophers y, such that $(\exists e)(\text{Taught}(x, y, \text{five theories}, e))$. But then it follows that the whole plurality of linguists taught each student five theories; whereas the intended reading leaves open the possibility that each philosopher was taught five theories by a single linguist.

10. See Goldman 1970; O'Shaughnessy 1973; Thalberg 1972; 1977; Thomson 1971; 1977. See also Costa 1987; Francken and Lombard 1992; Ginet 1990; Lombard 1985; O'Shaughnessy 1980; Parsons 1990; and Wilson 1985, among others.
11. See Fodor 1970. If an arsonist burned down a house that happened to contain a pot of soup, the arsonist may well have caused the soup to boil without boiling the soup. And not every verb with both transitive and intransitive forms is a causative. If you counted the marbles, it doesn't follow that the marbles counted.
12. For elaboration, see Pietroski forthcoming-b; for related proposals, see note 10. As Hornsby (1980) and others have noted, one also needs to think about what movements are in light of inferences such as the following: Brutus moved the knife, so the knife moved; and Brutus moved his hand, so his hand moved. If the former suggests that Brutus's action caused the motion of the knife, perhaps we should also say that his action caused the motion of his hand – or that Brutus's action of moving his hand was a "short" accordion-style event that ended with the motion of his hand; see also Dretske 1988.
13. Schein (2002) provides a detailed discussion of how appeal to thematic relations interacts with issues concerning event individuation.
14. It is not obvious that a bronze statue (call it '*S*') has all the same properties as the lump of bronze (call it '*L*') that constitutes the statue. Perhaps *L* can survive changes that *S* cannot. If so, then *S* and *L* have different "persistence" properties despite having the same spatiotemporal properties. And if $x = y$, then x has *all* the same properties as y. Similarly, one might think that persons and their bodies have different persistence properties, either because (i) a person's body can continue to exist after the person has died and ceased to exist, or (ii) a person can continue to exist after her body has ceased to exist. These are notoriously complex matters.
15. Davidson (1987a, p. 452) also appears to embrace the following (far more tendentious) modal thesis: had there been an event f with different causes and/or effects than some actual event e, then it is not true that e would have been f. This is at odds with the following kind of claim: a physical event in my head that was actually caused by the presence of Oscar could have been caused by the presence of someone *else* who is (from my perspective) indistinguishable from Oscar. Davidson can thus consistently maintain that (i) mental events with propositional contents are physical events; (ii) such mental events have their propositional contents essentially; and (iii) mental events with different etiologies can have different propositional contents even if the events are otherwise indistinguishable. See Burge 1993, esp. pp. 105–6, for a different view and a helpful discussion.
16. See also Hornsby 1985 for a discussion of underlying assumptions about the (nonbasic) events that we ordinarily talk about and their relation to events that might literally satisfy predicates from the physical sciences.

6 Knowledge of Self, Others, and World

ERNEST SOSA

Davidson's epistemology, like Kant's, features a transcendental argument as its centerpiece. Both philosophers reject any priority, whether epistemological or conceptual, of the subjective over the objective, attempting thus to solve the problem of the external world. For Davidson, three varieties of knowledge are coordinate – knowledge of the self, of other minds, and of the external world. None has priority. Despite the epistemologically coordinate status of the mind and the world, however, the content of the mind can be shown to entail how things are out in the world. More exactly, Davidson argues, we could not possibly have the beliefs we have, with their contents, unless the world around us were pretty much the way we take it to be, at least in its general outline. We are thus offered a way to argue, to all appearances a priori, from how it is in our minds to how things are in the world. The argument is a priori at least in being free of premises or assumptions about contingent particularities concerning the world around us or our relation to it. From premises about the contents of our propositional attitudes, the argument wends its way to a conclusion about the general lines of how the world around us is structured and populated.

Before presenting his own account, Davidson rejects received views of meaning and knowledge. What follows will combine themes from his critique of alternatives with his more positive account and its way of dealing with the skeptic.

1. DAVIDSON'S EPISTEMIC ARGUMENT AGAINST EMPIRICIST THEORIES OF MEANING

Empiricist accounts of meaning, including Quine's and Dummett's, lead to skepticism, warns Davidson, who sees one advantage of his own account as its better response to the skeptic. Why do the earlier accounts lead to skepticism? Why does his own account do better?

According to Davidson (1986a, p. 313),

> Quine and Dummett agree on a basic principle, which is that whatever there is to meaning must be traced back somehow to experience, the given, or patterns of sensory stimulation, something intermediate between belief and the usual objects our beliefs are about. Once we take this step, we open the door to skepticism for we must then allow that a very great many – perhaps most – of the sentences we hold to be true may in fact be false. . . . Take Quine's proposal that whatever there is to the meaning (information value) of an observation sentence is determined by the patterns of sensory stimulation that would cause a speaker to assent to or dissent from the sentence. . . . Quine's proposal, like other forms of verificationism, makes for skepticism. For clearly a person's sensory stimulations could be just as they are and yet the world outside very different. (Remember the brain in the vat.)

Just how is this supposed to substantiate the charge that rival theories of meaning lead to a radical skepticism about objective external reality? According to Davidson, those theories lead us astray by opening a logical gap between our subjectivity and objective externalia: that is to say, between intrinsic descriptions of the contents of our minds and contingent facts about the world around us. For we are then necessarily unable to close this logical gap (not without vicious circularity).[1] These central themes are sounded repeatedly in Davidson's writings on epistemology. Consider, for example, this passage:

> There is at least a presumption that we are right about the contents of our own minds; so in the cases where we are right, we have knowledge. But such knowledge is logically independent of our beliefs about a world outside, and so cannot supply a foundation for science and common sense beliefs. This is how skeptics, like Hume, reason, and I think they are right: knowledge of the contents of our own minds cannot be the basis for the rest of our knowledge. If this is correct, then our beliefs about the world must, if they are to count as knowledge, stand alone. Yet it has seemed obvious to many philosophers that if each of our beliefs about the world, taken alone, may be false, there is no reason **all** such beliefs might not be false. (Davidson 2001a [1991], p. 194)

Here we have two main points: first, that if external reality is logically independent of the contents of our minds, then the contents of our minds can give no foundation for our beliefs about that external reality. And there is also a second point. Even if no particular contingent empirical belief is

guaranteed to be right, we may still be able to show how we cannot be generally wrong about the world around us. Despite the universal fallibility of individual empirical beliefs, we might still be able to secure a guarantee that lots of our beliefs must be right, that we are inevitably, massively right about the world around us.

The gap that yawns between our subjectivity and the external world, according to earlier views, is repeatedly blamed by Davidson as a source of radical skepticism. In his view, this sort of skepticism bedevils not only supernaturalist, classical foundationalists, such as Descartes, but also contemporary externalists, coherentists, and naturalists: Quine himself, for example. What exactly is the argument for so surprising a pairing and so unexpected a charge? Given the interest and importance of this issue, we are fortunate to have it explained by Davidson himself, with his customary pith, in an account worth quoting and considering in full:

> According to Quine's "naturalized epistemology" we should ask no more from the philosophy of knowledge than an account of how, given the evidence we have to go on, we are able to form a satisfactory theory of the world. The account draws on the best theory we have: our present science. The evidence on which the meanings of our sentences, and all our knowledge, ultimately depend is provided by stimulations of our sense organs. It is these stimulations that provide a person with his only cues to "what goes on around him." Quine is not, of course, a reductionist: "we cannot strip away the conceptual trappings sentence by sentence." Nevertheless, there is according to Quine a definite distinction to be made between the invariant content and the variant conceptual trappings, between report and invention, substance and style, cues and conceptualization.
>
> What matters, then, is not whether we can describe the data in a neutral, theory-free idiom; what matters is that there should be an ultimate source of evidence whose character can be wholly specified without reference to what it is evidence for. Thus patterns of stimulation, like sense-data, can be identified and described without reference to "what goes on around us." If our knowledge of the world derives entirely from evidence of this kind, then not only may our senses sometimes deceive us; it is possible that we are systematically and generally deceived.
>
> It is easy to remember what prompts this view: it is thought necessary to insulate the ultimate sources of evidence from the outside world in order to guarantee the authority of the evidence for the subject. Since we cannot be certain what the world outside the mind is like, the subjective can keep its virtue – its chastity, its certainty for us – only by being protected from contamination by the world. The familiar trouble is, of course, that

> the disconnection creates a gap no reasoning or construction can plausibly bridge. Once the Cartesian starting point has been chosen, there is no saying what the evidence is evidence for, or so it seems. Idealism, reductionist forms of empiricism, and skepticism loom. . . .
>
> Instead of saying it is the scheme-content dichotomy that has dominated and defined the problems of modern philosophy, then, one could as well say it is how the dualism of the objective and the subjective has been conceived. For these dualisms have a common origin: a concept of the mind with its private states and objects. . . . (Davidson 2001b [1988], p. 42–3)

According to Davidson's reasoning, certain philosophical positions – Quine's, for example – lead to skepticism by allowing a certain crucial possibility. In his study of Davidson's epistemological views, Thomas Nagel endorses this reasoning, and joins in taking the crucial "skeptical possibility" to be this: that the external world could differ radically despite presenting subjectively indistinguishable appearances – that a logical chasm divides our subjective appearances from the world beyond. By leaving that possibility open, one smooths the way for the skeptic. Nagel reminds us of the many ways in which the skeptic has been opposed without success. One might try to refute him by reducing external reality to subjective experience, for example, which is the way of phenomenalists, verificationists, pragmatists, transcendental idealists, and internal realists. Reductionists deny that it is really possible for one's experience to remain indistinguishable even while external reality diverges as broadly as the skeptic imagines. And such reductionism is one traditional way in which philosophers have tried to oppose the skeptic.

An alternative way does not rely on any ontological reduction of the world to the mind. But it attempts to argue from the internal to the external nevertheless, either in the way of Descartes or in some other way.

According to Nagel, Davidson has a third way of dealing with the skeptic, one that attempts to relate the external to the subjective neither by deduction nor by reduction. Some might be misled into seeing this as an attempted reduction not of the external to the subjective, but in the opposite direction. However, Davidson is in fact no reductionist at all, not even in the way of the behaviorist. Despite renouncing any such reductionism, Davidson's way still does manage to yield an a priori argument that we cannot be as radically mistaken about the external as the skeptic would have us believe. If the attempt succeeds, therefore, it does, amazingly enough, refute the skeptic, and does so from the armchair.

Such an a priori argument is said to be crucially required, since we cannot fall back on retail empirical beliefs when arguing against a radical

skeptic. Any such appeal to empirical beliefs would just beg the question, since the skeptic puts in doubt all such beliefs in one fell swoop.

> The *a priori* argument is needed because the empirical reasons for particular beliefs are not by themselves sufficient. It makes sense to think about each of a great many of my beliefs, taken one at a time, that it might be false, in spite of the evidence. Some reason must be given to show that these individual possibilities can't be combined into the possibility that most of them are false. That reason can't be just the sum of the particular reasons for each of them, since these are just further beliefs in the set, and the whole question is whether most of them might be false. If they were, their apparent support of one another would be systematically misleading. So we cannot demonstrate empirically that this is not the case, as is proposed by naturalized epistemology; it must be *proved* to be impossible if skepticism is to be ruled out. We need an *a priori* argument, and Davidson has given us one. It is an argument which does *not* rely on the reduction of truth to coherence. (Nagel 1999, p. 203)

Davidson's response is to agree – indeed, effusively so – and even to reject the one gesture by which Nagel tries to distance him from any such purported a priori refutation of the skeptic.[2] Nagel had ventured that Davidson would resist viewing his reasoning as designed to run *from* thought *to* objective reality. In response, Davidson counters by avowing his intention to argue in precisely that way. Davidson concludes his response with a caveat of his own, but one that would not give Nagel pause, nor much gladden the skeptic.

> Nagel is understandably astonished that *a priori* reasoning should show that our general picture of the world around us "covering vast tracts of history, natural science, and ordinary lore," is largely true. Of course, as he notes, there is an empirical premise, the *cogito*. There is not an *a priori* proof that there is a world more or less as I think of it. Nor is the empirical premise a small one. The conclusion that I know that the world, both in general and in many particular ways, is as I think it is, depends on the fact that I have just the beliefs I do. (Davidson 1999i, p. 209)

Whether this removes the "a priori" character of the reasoning is a matter of definitional opinion. In what follows, let us take reasoning to be "a priori" so long as it relies neither on any substantive commitment concerning the external world, nor on external observation as a mode of justified belief acquisition. Reasoning is thus not reduced from a priori status simply because it rests on contingent commitments concerning the contents of the

reasoner's own mind. It is in this sense of 'a priori' that Nagel had tried to distance Davidson from a priori reasoning designed to counter the skeptic. And it is taken in *this* sense that Davidson rejects that gesture.

Regardless of how we choose to speak, the substance of Davidson's claim remains amazing. Who would have thought that, just on the basis of otherwise a priori reflection – reasoning from our knowledge of what we believe, of how it is within our own minds – we should be able to arrive at substantive conclusions about the objective and independent external world around us? This is indeed the sort of reasoning that Cartesians once sought.

Nagel is not fully persuaded to join Davidson in arguing thus against the skeptic, but he does think that Davidson has come up with a deep paradox. The only ways out that he can see are, first, a Platonism that is anathema to contemporary naturalism, and second, a radical "form of skepticism about whether one was really capable of significant thought" (p. 205).

2. WHAT IS DAVIDSON'S ARGUMENT?

How does Davidson propose to refute the skeptic and solve the problem of the external world? His reasoning has two main sources. One is his account of radical interpretation, the other his externalism. Though intimately related, the two are separable on close inspection. Let us first take up radical interpretation.

What is involved in attributing propositional attitudes to someone else? How can we manage it if we do not presuppose a common language giving us easy access to the mind behind the words? In undertaking such radical interpretation, there is no substitute for considering the other's variable assents and dissents upon correspondingly varying occasions of speech. We must then assess meaning in the light of external promptings by the saliently variable features of those occasions. Attributable meaning thus comes in a package together with causation by externalia.

This connects with two other important themes in Davidson's account of meaning and knowledge: charity and triangulation. Since we can attribute such observational beliefs and knowledge only if we interpret through what we see to prompt the believer's assent, the very nature of radical interpretation commits us to interpret the other as largely right in his beliefs. Moreover, the meanings that an interpreter attributes to a speaker's utterances are then bound to reside in the commonly shared nodes of the causal trees that prompt the respective attitudes of speaker and interpreter.

Meaning – attributed meaning – is thus bound to derive from such triangulation. From this, Davidson draws an important lesson:

> It should now be clear what ensures that our view of the world is, in its plainest features, largely correct. The reason is that the stimuli that cause our most basic verbal responses also determine what those verbal responses mean, and the content of the beliefs that accompany them. The nature of correct interpretation guarantees both that a large number of our simplest beliefs are true, and that the nature of those beliefs is known to others. Of course many beliefs are given content by their relations to further beliefs, or are caused by misleading sensations; any particular belief or set of beliefs about the world around us may be false. What cannot be the case is that our general picture of the world and our place in it is mistaken, for it is this picture which informs the rest of our beliefs and makes them intelligible, whether they be true or false. (Davidson 2001b [1991], p. 213–4)

"[B]eliefs are by nature generally true." So Davidson sums up his point (Davidson 2001 [1983], p. 153). On this matter, his statements and arguments leave open two possible interpretations, however – or so Barry Stroud has argued plausibly (Stroud 1999). There is a stronger reading and a weaker reading, as follows:

Stronger reading:	No believer's set of beliefs could be massively false.
Weaker reading:	No interpreter could correctly interpret a speaker in such a way that the speaker's beliefs come out massively false in the interpreter's opinion.

Stroud then documents his claim that the stronger reading is endorsed by Davidson, as when the latter argues that it "cannot be the rule" that a speaker be interpreted on the basis of "shared but erroneous beliefs," and concludes that "massive error about the world is simply unintelligible."

That is of course reminiscent of the celebrated and much-discussed Omniscient Interpreter argument; but I will pass over that with little further comment, since Davidson has now effectively disavowed it in print.[3] Recently he has written, for example, as follows: "I also agree ... that the argument that summons up an Omniscient Interpreter does not advance my case. As with Swampman, I regret these sorties into science fiction. ... If the case can be made with an omniscient interpreter, it can be made without, and better" (Davidson 1999c, p. 192).

Stroud argues that the stronger reading does not follow from the weaker reading, since "the conditions of interpretation as Davidson describes them do not alone guarantee that what interpreters take the causes of utterances

to be is what they in fact are . . ., [and] if it is not a necessary condition of interpretation that the actual causes of utterances are what interpreters and speakers take their causes to be, as I think it is not, the guaranteed truth of attributed belief is not supported by the conditions of radical interpretation as described so far" (Stroud 1999, p. 154).

Stroud's argument highlights (a) that the stronger reading does not follow from the weaker, (b) that Davidson should renounce the stronger, and (c) that the weaker will suffice as a good response to the skeptic. Stroud does not think the stronger reading "is needed to block the threat of philosophical skepticism in the highly promising way he [Davidson] has in mind; the weaker reading alone, and the conditions of interpretation as described so far, suffice" (p. 154).

According to Stroud, we must grant the abstract possibility that we are mostly or even wholly wrong in our contingent beliefs about external reality. To reject that possibility is to threaten the objectivity of our beliefs about such reality. What Davidson's account of radical interpretation yields, then, according to Stroud, is not the impossibility that our beliefs are largely false, but rather just that "belief-attribution is in its nature largely truth-ascribing" (p. 155).

Stroud's discussion seems to me illuminatingly right in distinguishing the two readings and in arguing for the independence of the stronger from the weaker. Evidently, one must go beyond the conditions of radical interpretation if one is to derive anything like the stronger reading. Nevertheless, I am not persuaded on two main points. I do not believe that the weaker reading, and the fact that belief-attribution must be truth-ascribing, constitutes much of a response to the skeptic. And, on the other side, I am not persuaded that the stronger reading is either false or unacceptable, even if it does not follow from the weaker. The following section will take up Davidson's externalist argument for the stronger reading. What remains of this section will discuss Stroud's argument for the claim that the weaker reading on its own is enough to block the skeptic.

Here is that argument:

> An inquirer's relation to the apparently innocent possibility from which a skeptical threat is thought eventually to arise is therefore parallel to a speaker's relation to the possibility expressed in the paradoxical sentence 'I believe that it is raining, and it is not raining'. That is not something one could consistently believe or assert. . . . If the apparently innocent possibility from which the epistemological reasoning would begin is not a possibility anyone could consistently believe to be actual, it can be eliminated from

> serious consideration right at the beginning. There would be no need to insist on the stronger view. . . . There can be no general threat [of the sort pressed by the skeptic] because our considering the specific attributed beliefs we [ask about] . . . guarantees that we find those beliefs to be for the most part true. Our having them and their being all or mostly false is not a possibility we could consistently believe to be actual, so it is not a possibility we could be pressed to explain how we know is not actual.
>
> That is not to say that we therefore *know* that all or most of the things we believe are true. That would be the negation of skepticism, and it does not follow from this anti-skeptical strategy. The goal is only to block a familiar route to skepticism, not to show that skepticism is false. A certain possibility is to be removed from consideration as the source of a potentially unanswerable threat. (pp. 136–7)

Several points remain doubtful or unclear in this reasoning. From the fact that <Not-*p* but I believe *p*> is Moore-paradoxical,[4] and cannot coherently be believed or asserted, it does not follow that the corresponding proposition of the following form must also share that fate: <It is possible that both: not-*p* and I believe *p*; that is to say, it might have been that the following two things were so at the same time: that not-*p* while I believed that *p* anyhow>.

Even if this propositional form were in fact incoherent, moreover, so that each of the propositions of that form about each of my beliefs would be incoherent, it still would not follow that <It is possible that my beliefs are massively false> is also incoherent. Here's an analogy. In a certain time span, *T*, I may list five positive integers: 2, 5, 7, 3, 8. Each of the positive integers I list is such that it could not possibly have been greater than 10. However, that still leaves open the following possibility: that in time span *T* I might have listed some positive integer larger than 10. Similarly, even if I cannot coherently think or say that although I believe that *p*, still it is *not* so that *p*, and even if, *only for the sake of argument*, it is granted that I cannot coherently think or say the following: <although I believe that *p*, it is, compatibly with that, possible that not-*p*>, nevertheless it does not follow from any of that, nor is it true, that I cannot coherently think or say *this*: <I might be massively wrong in my beliefs>.

Accordingly, I do not see that simply from the Davidsonian account of radical interpretation – and specifically, from the weaker reading of Davidson's key thesis – anything has been shown to follow that would block the skeptic's route to supposing, as he tends to do, that we might be massively wrong in our contingent beliefs about the external world. Nor has it

even been shown to follow that we are somehow incoherent in supposing, with regard to any actual belief we host about external reality, that we *might* have hosted that belief despite its falsity (unawares, of course).

Accordingly, I conclude that we must go beyond the weaker reading to the stronger if we hope to block the skeptic's progress. However, I agree with Stroud that the weaker reading is distinct from the stronger and does not entail it. That being the case, we need to go beyond considerations involving radical interpretation if we are to reach the conclusions that we need against the skeptic. What are the chances that we can reach any such conclusions a priori? Do we *need* to reach a priori any such conclusion that can possibly be of use against the determined skeptic?

3. SKEPTICISM ABOUT DAVIDSON'S ACCOUNT OF SKEPTICISM

Suppose *arguendo* that we concede the need for an a priori argument if we are to oppose the skeptic with any hope of success – as Nagel supposes, and about which in response Davidson seems to concur. And suppose further that Davidson has indeed given us the sort of argument we need. I wish to argue that we are nevertheless not much better off against the skeptic, and indeed that such a concession would admit the skeptic's Trojan horse.

Consider an unfortunate victim of futuristic technology whose brain is wirelessly controlled by demonic controllers. Suppose this to have come about only hours ago, perhaps overnight, while the victim slept soundly. This is the sort of possibility that Davidson's reasoning does not manage to preclude. So now we face the following outcome. True, we cannot reason that we might be generally wrong just because we might be wrong in any specific instance. But, by the same taken, we cannot reason that we are safe from being wrong in any specific instance just because we cannot be generally wrong. So even if Davidson's reasoning enables us to close the gap between what is accessible to us a priori and our general correctness about the character of external reality, this still leaves in place gaps aplenty between what is accessible to us a priori and the various specific substantive beliefs that we hold about the world and our place in it at any given time. None of these gaps would seem bridgeable just with an argument like Davidson's.

It might be argued that appearances are deceptive here, as they so often are in this dark swamp. After all, Davidson's argument does yield the conclusion that our substantive beliefs about externalia are and must be massively correct. May we not therefore conclude that any particular such belief must have presumptive justification? Such justification might be defeated,

of course, by particular untoward circumstances. Absent such defeat, however, it is certainly not nothing. And so we seem to have made progress against the skeptic after all.

This argument has a certain "blanketing" property that should give us pause. It would render all substantive beliefs presumptively justified, the astrologer's along with the astronomer's. And now the action would shift to what accounts for the difference, what accounts for the defeat of the astrologer's justification but not of the astronomer's justification. Anyone who believes something out of the blue, and has no very good argument against his reliability on the subject matter involved, any trusting soul innocent of much relevant theory, would seem ipso facto to inherit undefeated epistemic justification.

What is more, there is a further problem that Davidson himself has come explicitly to recognize: namely, that if we are thinking of justification as provided by his proposed complex reasoning, then only those enlightened few who grasp and adopt that reasoning would have their knowledge protected against the skeptic. The mass of nonphilosophers – indeed, the mass of non-Davidsonians – no matter how brilliant or otherwise well informed, would remain benighted. In spite of this, Davidson retains hope that his account will still do some epistemic good. So we will need to consider how any such reasoning might accomplish its good work.

4. DAVIDSON'S ARGUMENT AND THE SKEPTIC'S TROJAN HORSE

If we frame our debate with the skeptic as do Davidson and Nagel, therefore, the skeptic wins, regardless of whatever success Davidson's transcendental argument may enjoy. The success of the transcendental argument turns on complex and still-disputed issues in the philosophy of language and mind. Even if that should all turn out favorably, however, once we think of skepticism in the way suggested by the writings of Davidson and Nagel, there is no way ultimately to overcome skepticism. We are unwise if, in effect, we thus allow the skeptic to set the agenda and frame the issue. We should be wary of setting up the dialectic with the skeptic along the following lines, in the fashion often attributed to Descartes:

> A. If we are to know realm *W* it must be via realm *M*.
>
> B. The way to know a realm *X* via a realm *Y* is by knowing *Y* and reasoning validly from one's knowledge of *Y* to conclusions about *X*.

C. Only deductive reasoning is valid.

D. There is a logical gap between *M* and *W* that no deductive reasoning could possibly bridge.

(Here *M* is the realm of one's mind at a given time, and *W* is the realm of one's surrounding objective world.)

However, one does not defeat the skeptic simply by rejecting C, while also adding inductive reasoning to our set of valid forms of reasoning. For it is no more clear how to cross the gap between *M* and *W* through valid *inductive* reasoning. Once we grant a division such as that of *M* and *W*, and concede that any knowledge of *W* would have to be via knowledge of *M*, it will be hard to withstand the attack of the skeptic.

On one straightforward reading, Davidson's way of framing skepticism puts the realm of one's beliefs, *B*, in place of the realm of one's mind, *M*, in the argument just presented. Accordingly, Davidson does inherit the problem of crossing a divide from the subjective to the objective. How then does he propose that we reason our way from *B* to *W*?

The argument is presented in various forms in several places. For example, it appears as follows in Davidson 2001b [1988], pp. 43–5.

> The action has centered on the concept of subjectivity, what is "in the mind." Let us start with what it is we know or grasp when we know the meaning of a word or sentence. It is a commonplace of the empirical tradition that we learn our first words (which at the start serve the function of sentences – words like 'apple', 'man', 'dog', 'water' – through a conditioning of sounds or verbal behavior to appropriate bits of matter in the public domain). The conditioning works best with objects that interest the learner and are hard to miss by either teacher or pupil. This is not just a story about how we learn to use words: it must also be an essential part of an adequate account of what words refer to and what they mean.
>
> Needless to say, the whole story cannot be this simple. On the other hand, it is hard to believe that this sort of direct interaction between language users and public events and objects is not a basic part of the whole story, the part that, directly or indirectly, largely determines how words are related to things. . . . The grasp of meanings is determined only by the terminal elements in the conditioning process and is tested only by the end product: use of words geared to appropriate objects and situations. This is perhaps best seen by noticing that two speakers who "mean the same thing" by an expression need have no more in common than their dispositions to appropriate verbal behavior; the neural networks may be very different. The matter may be put the other way around: two speakers may be alike in all

> relevant physical respects, and yet they may mean quite different things by the same words because of differences in the external situations in which the words were learned . . .; in the simplest and most basic cases words and sentences derive their meaning from the objects and circumstances in which they were learned. A sentence which one has been conditioned by the learning process to be caused to hold true by the presence of fires will be true when there is a fire present; a word one has been conditioned to be caused to hold applicable by the presence of snakes will refer to snakes. Of course very many words and sentences are not learned this way; but it is those that are that anchor language to the world. . . .
>
> The fallout from these considerations for the theory of knowledge is (or ought to be) nothing less than revolutionary. If words and thoughts are, in the most basic cases, necessarily about the sorts of objects and events that cause them, there is no room for Cartesian doubts about the independent existence of such objects and events. Doubts there can be, of course. But there need be nothing we are indubitably right about for it to be certain that we are mostly right about the nature of the world. Sometimes skepticism seems to rest on a simple fallacy, the fallacy of reasoning from the fact that there is nothing we might not be wrong about to the conclusion that we might be wrong about everything. The second possibility is ruled out if we accept that our simplest sentences are given their meanings by the situations that generally cause us to hold them true or false, since to hold a sentence we understand to be true or false is to have a belief. Continuing along this line, we see that general skepticism about the deliverances of the senses cannot even be formulated, since the senses and their deliverances play no central *theoretical* role in the account of belief, meaning, and knowledge if the contents of the mind depend on the causal relations, whatever they may be, between the attitudes and the world. This is not to deny the importance of the actual causal role of the senses in knowledge and the acquisition of language, of course.

In the simplest cases, we are told, words and sentences derive their meanings from the objects and circumstances that are associated with their learning. Just how does this happen? What sort of "derivation" is here in play? In leading up to and defending his "derivation" thesis, Davidson makes some revealing claims (which are quoted as follows, with my emphases in italic):

> [Two] speakers who "mean the same thing" by an expression need have no more in common than their *dispositions to appropriate verbal behavior*; the neural networks may be very different. (p. 44)

> A sentence which one has been *conditioned by the learning process to be caused to hold true by the presence of fires* will be true when there is a fire present; a word one has been conditioned to be caused to hold applicable by the presence of snakes will refer to snakes. (pp. 44–5)

> If words and thoughts are, in the most basic cases, necessarily about *the sorts of objects and events that cause them*, there is no room for Cartesian doubts about the independent existence of such objects and events. (p. 45)

In considering this, and the many other passages in which Davidson makes the same basic points, we need to distinguish between (a) dispositions to appropriate verbal behavior and (b) the process or processes that may cause such dispositions in a certain speaker/believer. The disposition that one hosts in being such that "one is caused to hold a certain sentence true by the presence of fires" may have been put in place by repeated experience of fires. If so, that is just a contingent matter of fact, which might possibly have been otherwise. The actual disposition involves the fact that one *would* be caused to say or think "Fire!" in the presence of a fire (an evident enough fire). This most likely was indeed put in place through some experience on the part of the speaker/thinker with actual fires. Even if that is so, it would seem only a contingent matter of fact, and the disposition might even have been there innately. But now we have a problem for Davidson's transcendental argument. For there is now no evident impossibility in one's understanding the concepts of our commonsense objective reality, where this understanding resides, at least in part, in one's complex dispositions to verbal and other relevant behavior, although one has *not* acquired such dispositions through causal intercourse with exemplars of the concepts possessed.

So it seems at best unestablished that *both* (a) one could possibly have the beliefs that one has only through having in one's possession the concepts constitutive of the contents of those beliefs, and (b) the only way one could have such concepts is through causal intercourse with their exemplars. Further defense would seem required in favor of assumption (b), for there is a plausible argument against it, one indeed that seems suggested already by Davidson's actual words. It can be argued plausibly, after all, that the relevant requirement for possession of a concept – say, a recognitional concept – is only one's hosting a disposition that makes for differential sensitivity to the presence or absence of exemplars. And such a disposition *need not* have been acquired through causal interaction with actual exemplars.

There are indications that Davidson has come to regard his argument as less plausibly a priori than one might have thought, perhaps less so even than he himself originally had thought. Consider this recent passage:

> Is my argument for the "massive" (essential) truth of our perceptual beliefs transcendental? If you accept the steps that lead to my version of externalism, . . . then you cannot, I think, be a skeptic about the existence of an external world much like the one we all believe we share, nor about the existence of other people with minds like ours. But the considerations in favor of semantic realism seem to depend in part not on purely a priori considerations but rather on a view of the way people are. (Davidson 1999c, p. 194)

Indeed, in recent passages Davidson candidly reveals his vacillation about epistemology and skepticism, especially in his more recent thought. The following are particularly revealing:

> I have vacillated over the years on how to describe my attitude toward scepticism. Do I think that if I am right about the nature of thought scepticism is false, or that scepticism simply cannot get off the ground? Passages Stroud quotes suggest the former. At the same time, I have been happy to go along with Rorty in telling the sceptic to get lost. The two poses can be reconciled by pointing out that though I think scepticism as Stroud formulates it is false, I did not set out to show this. Reflecting on the nature of thought and interpretation led me to a position which, if correct, entails that we have a basically sound view of the world around us. If so, there is no point attempting, *in addition*, to show the sceptic wrong. (Davidson 1999d, p. 163)

> Nagel quotes "A Coherence Theory of Truth and Knowledge" as saying, "The agent has only to reflect on what a belief is to appreciate that most of his basic beliefs are true." I was concerned to show that each of us not only has a basis for his picture of the world in his perceptual beliefs, but that he also, on reflection, would see that there was a reason (my arguments) for thinking this. I was trying to fend off the criticism (which perhaps surfaces in Stroud's contribution to this volume) that I might have shown that we do have a large supply of true beliefs, but not have shown that these constitute knowledge. I now think this attempt at fending off criticism was a mistake, if for no other reason than that it would seem to credit only those whose philosophical thinking is correct with knowledge. The right thing to say is rather this: we are *justified* in taking our perceptual beliefs to be true, even when they are not and so when they are true, they constitute knowledge (this is what I meant by saying our perceptual beliefs are "veridical"). But since

> our only reasons for holding them true are the support they get from further perceptual beliefs and general coherence with how we think things are, the underlying source of justification is not itself a reason. We do not *infer* our perceptual beliefs from something else more foundational. (Davidson 1999i, p. 208)[5]

In these passages, we are given to understand that there is a source of justification other than the adducing of a reason for one's belief, or the basing of one's belief on a reason. Reason-based justification is not the sort of justification that Davidson calls to our attention in his many reflections about the bearing of his content externalism on issues of skepticism and epistemology. He has now seen this clearly, and acknowledges it openly. The source of justification that he focuses on is not a *reason*, inasmuch as it is a source of justification that epistemically favors even those who have no *belief* in any Davidsonian theory about how our beliefs and sayings acquire content.[6] Nor need one have any reason at all for beliefs that are nonetheless somehow justified. Again, nothing in the causality of perceptual beliefs provides a reason for them, but at the same time "many perceptual beliefs are true, and the explanation of this fact shows why we are justified in believing them." Those innocent of a Davidsonian account of the nature of mental content and meaning would lack any rationale, deriving from any such Davidsonian theory of content, in favor of their empirical beliefs about their environing world. But, if Davidson's account is true, they would still have a source of justification involving the nature of such content. And it is the existence and nature of such justification that Davidson now sees himself as having clarified through his writings on externalism, knowledge, justification, and skepticism.

The Davidsonian justification at issue is *not*, therefore, of either of the sorts that Nagel distinguishes. It is a justification that derives neither from a reduction of the external to the subjective (nor, for that matter, the other way around), nor from a deduction of the external from the subjective. Nor is it a justification that comes with possession of an argument – an a priori argument – through which one gains support for its conclusion. Davidson has concluded that it is hopeless to suppose that this is how people generally avoid the clutches of the skeptic. For people generally adduce no such Davidsonian argument in support of their retail beliefs. So even if a few philosophers, persuaded by Davidson, do adduce such a complex argument concerning the nature of mental and linguistic content, and even if they do *thereby* gain some measure of justification for their empirical beliefs, that will not explain the justification that ordinary folk have for *their* empirical

beliefs, and so it will not explain how it is that *these folk* are safe from the objections of the skeptic.

What then *is* the source of the distinctive Davidsonian empirical justification that a subject's perceptual beliefs get from something other than the support of other empirical beliefs hosted by that subject? Apparently, it is simply the *truth* of the Davidsonian account of how our sayings and attitudes must derive their content, and how this guarantees that one's picture of the environing world *must* be massively correct, especially in its perceptual components.

Two fascinating questions ensue. First of all, isn't Davidson now drawing on reliabilist intuitions? It would seem to be the high level of reliability of our empirical beliefs, given his account of meaning and content, that now serves as the core of the special source of justification invoked to explain the high epistemic status of our empirical, and especially our perceptual, beliefs.[7]

The second interesting question concerns the status of Davidson's theory and his "answer" to the skeptic. If the source of justification should now be viewed as distinct from any reasoning, from any invoking of a justifying argument, then it is no longer clear why it must be a priori (not that it was all that clear in any case). It becomes positively opaque why the a priority of Davidson's epistemologically effective reasoning should be an issue. Now it would seem to matter only that the reasoning establish the theory, for it is just the *truth* of the theory that has turned out to be epistemologically effective. What seems to matter is essentially that as things in fact stand in our contingent circumstances, content is set externalistically through causal linkages with our external environment. For it is through this fact that the reliability of our beliefs is assured. And it is from their assured reliability that their presumptive justification derives. Of course, if in no possible world could content derive in any other way, then the reliability of our beliefs would be assured with alethic necessity. But it is far from clear that Davidson's account, or any such content-externalist account, is true with alethic necessity.

Note, finally, that through this new approach we have a way to understand an epistemology naturalized that avoids the objection, voiced for example by Nagel, that such naturalization of epistemology would involve a vicious circularity. This is also reminiscent of the long-standing controversy over whether Descartes' supernaturalization is similarly vicious in its circularity. Consider how interesting in this connection is Davidson's new reason-avoiding approach, on which justification derives somehow from a source other than the subject's actual reasoning. As we have seen, the new

approach strongly suggests a reliabilism for which justification can derive from the reliability of the sorts of beliefs at issue (perceptual beliefs most importantly, though Davidson also comes to generalize beyond these). And Descartes' epistemological reasoning can be viewed similarly, as proposing a way of understanding our forming of beliefs (in the lap of an epistemically benevolent God) as necessarily reliable. (Of course, for special reasons, Descartes *did* aim for alethic necessity, and for a priori reasoning, but the present comparison is independent of that.) For if Descartes' epistemological theorizing is meant to identify a way in which our beliefs come to be justified that is precisely *not* through any reasonings from which we derive them as conclusions, then Descartes too can avoid vicious circularity by responding to the skeptic that our beliefs have a source of justification that need not involve the use of reasoning. In Descartes' case, the effective fact would involve assent attendant upon sufficiently clear and distinct perception and favored by God's epistemic benevolence. In Davidson's case, the effective fact would involve rather that we would not form beliefs having the contents of our empirical beliefs if we did not interact appropriately with surroundings characterized generally as are our surroundings in this world.

The main remaining question concerns the epistemic status of our empirical beliefs once we have reasoned along with Davidson, while leaning presumably on adequate empirical support. Do our empirical beliefs gain any epistemic status through such reasoning? It might be thought that they obviously do not. How could they do so without vicious circularity? How could such theoretical beliefs as to the nature of content add to the status of one's empirical beliefs generally, if it is granted that the theoretical beliefs must in turn gain their own status through reliance on empirical, perceptual beliefs?

Here again the comparison to Descartes is instructive. Descartes did obviously think by the end of the *Meditations* that he had improved himself epistemically. But it is hard to see how he could possibly have avoided the vicious circularity of which he has so often been accused. Because Descartes, early in the *Meditations*, puts so much in doubt, including even his simplest arithmetical and geometrical beliefs, it is hard to see how he could possibly dig himself out of so deep a skeptical hole without at some point falling into vicious circularity. Descartes does have a way out, however, one in fact open both to the advocates of common sense, such as Moore, and to those who advocate an epistemology naturalized – either Quine's, or, now, Davidson's.[8] The response is indeed a kind of "coherence theory of knowledge," after all. For it is by adding interestingly to the coherence of

one's picture of the world and one's place in it that one is able to gain a *further* measure of distinctive, epistemically valuable justification for one's own empirical beliefs, a measure of justification that goes beyond the mere reliability of those beliefs derivative from how we must acquire contents and form beliefs. The additional measure of justification goes beyond any delivered by sheer reliability, and does so by bringing to consciousness a well-founded account of how our nature and emplacement do yield such reliability. Whether this is done in the way of Descartes, or in the way of Moore, or of Quine, or, now, of Davidson, the result would be, structurally, the same: a more satisfyingly coherent account of our nature and our place in the scheme of things.[9]

Notes

1. Compare Nagel's discussion of Davidson's epistemology (Nagel 1999, p. 203); we shall return to this passage later.
2. See his "Reply to Thomas Nagel" (Davidson 1999i, pp. 207–9).
3. There is an instructive critical literature on this, including Foley and Fumerton 1985.
4. See Schilpp 1942, p. 543.
5. Compare Davidson's "Reply to John McDowell":

   > My view is that particular empirical beliefs are supported by other beliefs, some of them perceptual and some not. Perceptual beliefs are caused by features of the environment, but nothing in their causality (except in special, derivative cases) provides a reason for such belief. Nevertheless, many basic perceptual beliefs are true, and the explanation of this fact shows why we are justified in believing them. We know many things where our only reasons for believing them are further beliefs. (Davidson 1999h, pp. 105–6)

   "A Coherence Theory of Truth and Knowledge" had taken a rather different view of the matter:

   > What is needed to answer the skeptic is to show that someone with a (more or less) coherent set of beliefs has a reason to suppose his beliefs are not mistaken in the main. What we have shown is that it is absurd to look for a justifying ground for the totality of beliefs, something outside this totality which we can use to test or compare with our beliefs. The answer to our problem must then be to find a *reason* for supposing most of our beliefs are true that is not a form of *evidence*. (Davidson 2001 [1983], p. 146)
   >
   > I have not been concerned with the canons of evidential support (if such there be), but to show that all that counts as evidence or justification for a belief must come from the same totality of belief to which it belongs (p. 153).
6. Compare this: "Perceptual beliefs are caused by features of the environment, but nothing in their causality (except in special, derivative cases) provides a reason

for such beliefs. Nevertheless, many basic perceptual beliefs are true, and the explanation of this fact shows why we are justified in believing them" (Davidson 1999h, pp. 105–6).

7. That at a deep level Davidson is a reliabilist is suggested in my Sosa 1986, esp. pp. 395–7.
8. More details are offered in two recent papers of mine: Sosa 1997a; 1997b.
9. For further discussion of Davidson's epistemological views, see Klein 1986; Ludwig 1992; McGinn 1986; see also Burge 1999; Genova 1999; McDowell 1999.

7 Language and Literature

SAMUEL C. WHEELER III

Although Donald Davidson has not written extensively about literature as such,[1] his thought has consequences for literary interpretation and theorizing, and it supplies elegant accounts of various topics, such as figuration and rhetoric, that have interested literary critics. His austere account of meaning appears to be an unpromising tool with which to explicate the idiosyncrasies of literary writing. This appearance is an illusion. His individualist account of language emphasizes the role of innovation and linguistic creativity, while accommodating the phenomena that make social-language analyses persuasive. While he agrees with the post-structuralists in rejecting meanings as entities, he also rejects important assumptions about the social nature of language common to most post-structuralists and other contemporary theorists. Most importantly, Davidson provides literary thought with a flexible and commonsense conception of what language is and how to conceptualize literary phenomena.

The first section of this chapter explains some of Davidson's basic ideas about how to cast a theory of meaning as a theory of truth, and explains how that theory connects language to the world. This section discusses the significance of Davidson's making meaning reside in language, rather than in something behind language, and explains why Davidson thinks that having a language requires a shared world.

The second section discusses interpretation. The fundamental idea of Davidsonian interpretation is that interpreting utterances is interpreting speech acts. Thus language interpretation is a special case of action interpretation. According to Davidson, a speech act is the production of sequences of sentences with (in the basic case) truth conditions, which production has reasons and is done with an intention. Writing acts are likewise productions of sequences of sentences, with intentions of various kinds. Interpretation is coming to understand those intentions and the truth conditions that make the sentences an appropriate instrument for those intentions.

The third section explains how Davidson's account of interpretation is applied in accounting for metaphor and other figures, and how it could be

extended to give accounts of irony, sarcasm, and other characteristics of speech acts. Much of this discussion deals with his famous article "What Metaphors Mean" (Davidson 1984 [1978]). This section discusses the notion of "first meaning," discusses what the primacy of intention means for Davidson, and discusses what Davidson takes malapropism to reveal about language and its understanding.

The fourth section discusses the sense in which, for Davidson, language is social, and the ways in which it is not. We will see how Davidson treats texts as products of actions, whose interpretation is as determinate as that of other actions. For a Davidsonian theory, texts, construed as repetitions of actions – that is, as reproduced and cited – can be indeterminate in ways beyond the indeterminacy of interpretation of utterances.

The fifth section summarizes the "anti-theoretical" side of Davidson's work in relation to literature. The fundamental thrust of Davidson's writing on literature and interpretation is that language use cannot be bound by convention, and is therefore not completely rule-governed, and so is not describable completely by a theory that spells out rules of use. I will show briefly how Davidson would deal with phenomena such as allusion and the existence of genre, which are explained by theorists who take public languages to be primary.

At various points in the last three sections I will extrapolate from Davidson's published work to conjecture what he would say about issues in literary theory and criticism that he has not expressly addressed.

1. MEANING AND TRUTH CONDITIONS

Davidson's theory of meaning shares with Cleanth Brooks's the idea that meaning is given in words.[2] Meaning is not something else that words carry, which could be put more perspicuously as sets of sense impressions, neural activation patterns, Platonic Forms, or anything else. Rather, the concept of meaning, together with those of the interrelated phenomena of belief, desire, intention, and other mental states form an autonomous family of concepts. That is, while the phenomena are nothing over and above the underlying brain and world states, physically described, there is no reduction of the concepts of the mental to the concepts of physics or physiology. The notions of belief, desire, intention, action, truth, and meaning are interrelated. The concepts of truth and meaning are central in giving an account of belief, desire, intention, and action.[3]

Giving meaning by giving truth conditions is the formal reflection of the irreducibility of meaning.[4] The meaning of an utterance of another person is given by a sentence of one's own with the same truth conditions. (In order to keep things simple, I'll ignore complications introduced by the context sensitivity of natural language expressions such as tense and indexicals – see Chapter 1 for discussion.) One does not give truth conditions by giving verification conditions, or anything other than one's words. In particular, to give truth conditions does not require that it be possible to determine truth values. So, for Davidson, the specification of the meaning of 'The infinite illuminates the finite' would be given by the biconditional ' "The infinite illuminates the finite" is true if and only if the infinite illuminates the finite'. (This would seem more illuminating in the context of a complete theory for the language and if the object language were distinct from the metalanguage.) Given a correct interpretation of another speaker on which 'finite' is true of an entity if and only if it is finite, and so on, this biconditional, given that it is a consequence of a theory that generates canonically a similar biconditional for every object language sentence,[5] and given that it meets other appropriate constraints, would be all there is to saying what the person's sentence means. Thus, to a first approximation, meanings of sentences are given by giving other sentences that have the same meaning. Thus the meaning of another person's utterances can be given only by using language rich enough to say what he has said. If I state another's meaning, I do so in a language that I speak, using language that matches his.

Davidson recognizes that 'meaning' is a word with wide and varied application. His theory is a theory of one sense of 'meaning' – roughly, 'conceptual content'. Conceptual content, he thinks, can be given a theory. Other "senses" of 'meaning' rest on interpretation. Interpretation, because it is essentially "coming up with an explanation in response to data," cannot be the working out of an algorithm.[6] Davidson would not deny, for instance, that there are *meaningful* glances, but would argue that there is no finite theory that takes the positioning of eyebrows and lids and yields *the meaning of a glance*. On the other hand, meaning construed as the conceptual content of a sentence or utterance can be, on Davidson's view, explicated in the form of a finitely specifiable truth theory for the language, and plays a role in the description of other kinds of meaning.

How can we justify a claim of "correctness" of an interpretation, if we just have the other person's language and my language? Davidson connects people's languages to one another by their necessary link to a common world. Briefly, Davidson argues that in order to interpret responses to the world as having truth conditions and truth values, we have to suppose that

the speaker recognizes the possibility of mistakes.[7] Davidson's version of the "private language argument" holds that, in the case of a single responder, what a person is responding to is indeterminate. If we take response to be just a causal relation, the production of an effect on the organism by a cause in the environment, then an organism responding to the environment can be regarded as responding to everything along the chain of stimulation. In order to generate a determinate interpretation, therefore, there must be some objective way to select one of the links as the correct one. Only another speaker, Davidson argues, noting the environment, his own responses, and the learner's responses, can provide the objective criterion for correctness of interpretation: the object of thought or speech is the intersection of the chains leading to the speaker's and interpreter's responses to a common environment.[8] Thus the speaker can acquire linguistic ability only by means of another person's "triangulating" among his own responses, the speaker's responses, and the world being responded to.

Triangulation grounds the possibility of making a mistake by providing an objective criterion for what someone is talking or thinking about. This provides for a distinction between getting it right and getting it wrong. Only another speaker, responding to what you are responding to, can provide that distinction. That is, because whatever you respond to in the same way is the same "for you," an outside criterion is needed to get "objectively the same." Only an outside agent, responding both to what you're responding to and to your responses, can provide such a criterion. Thus a presupposition of language is a common world shared with other speakers. A "common world" is essentially a common pattern of categorization, that is, a common pattern of judgments about sameness. The context of communication that provides this criterion then also provides scope for the application of the concept of a mistake, and so for the concept of truth, which Davidson has argued is necessary for possession of a language. While Davidson, unlike other advocates of "private language arguments," holds that no specifically linguistic norms or rules are essential to communication, a wide body of shared conceptualization is essential. Any two speakers must agree, by and large, on what is the same as what, even though they may disagree on what is called what.

Davidson's private language argument has further features: the basic idea that Davidson shares with Derrida, that meaning is fundamentally linguistic, and not something behind or expressed by language, is also implicit in the argument. That is, beliefs, thoughts, and any other items that are true or false also need to be provided a criterion of truth and falsity. Since language provides the only truth bearers that are overt, and that can be

checked by other speakers, the idea that language is prior to thought, belief, and intention follows, since any assignment of propositional content to such nonlinguistic items must be via something that has a criterion of truth and falsity. So, truth definition, giving a meaning by producing a sentence, is the means of giving a theory of meaning.

The thesis that language is prior to or essential for thought does not imply that meaning, seeing that, desiring, and intending are entirely verbal, or that thought is nothing but sequences of words. The thesis is, rather, that whatever else there is to meaning, thinking, and having something in mind – such as picturing, seeing as, and the like – is secondary to language possession in the sense that possession of the concepts of truth and falsity, which are essential to thoughts' being of an objective world, derives from language possession.

Whereas most versions of the private language argument suppose that it establishes that any speaker must follow social linguistic rules or accept the linguistic norms of a community, Davidson argues that there is no reason to suppose that each of the two speakers must make the same response as the other. It is only required that each speaker make what the other recognizes as the same responses to what the other recognizes as the same in order that each speaker have the notion of truth and falsehood. Davidson argues, in effect, that a common world is required for thought, but not a common language.

Thus Davidson makes idiolects primary. Speech acts and writing acts are utterances and inscriptions in an individual idiolect. Such idiolects are the basic objects of truth definition and the primary loci of meaning. Since the evidence-basis for truth definition – that is, determination of meaning – is what a speaker or writer says or writes in which situations, and since any two people have different beliefs, experiences, and desires, any two idiolects will differ. Furthermore, as our experience, desires, and opinions change, our individual idiolects change. How a person understands another, then, will be represented by what Davidson calls a "passing" theory.[9]

An objective world, which can allow true and false beliefs relative to it, is only identifiable as the common world, which is the same as another's. Only relative to common beliefs can there be beliefs at all, since it is only relative to something other than one's own beliefs that the distinction between true and false beliefs can be drawn. Belief is basic to all thought, so all thought requires an objective and shared world to be about. That is, the "objective" world's contrast, the subjective, can be generated only when there are beliefs or other truth bearers. If all truth bearers presuppose the common world, then the "subjective world" depends on the objective. Every conception

of the world presupposes substantial dependence on the world that any thinker or language user must share with all others. Davidson thus takes the supposition that there are "different worlds" for different cultures, or that there is "incommensurability" between one literary tradition and another, to be either literal nonsense or a hyperbolic way of saying that people and traditions can be very different. A general way of putting his point is that, in order to treat another as a language user at all, a substantial amount of agreement on "what is what" must be presupposed. Otherwise – since there are no meanings as logoi or magic language tokens,[10] and since therefore "meaning" is assigned by applying the "mental" family of predicates, and since that family of predicates is applied by detecting a point of contact with a common world – there would be no basis for supposing that the alleged "other" was speaking a language at all.

Davidson does not deny that there can be subject matters in another culture's discourse that are incomprehensible to someone from Ohio, for instance. But the knowledge that this subject matter was a subject matter and was incomprehensible would be founded on a body of agreement about other subject matters. We can only know that higher mathematicians are talking over our heads if we have some basis for thinking that the mathematicians are talking. And that basis must be their connection with us on the mundane topics of whether it is time for lunch and what and where their car keys are.

The same is true for historical understanding, or for understanding what a text meant to its authors. While it is difficult to say exactly how each member of the sequence of redactors and transmitters and retellers understood the story of Jacob's wrestling match (Genesis 32), and impossible to say exactly what the text itself means, we know that Jacob is taken to be a man, that he is in a confrontation with his brother, that brothers are male siblings, and so on. The most mysterious narrative, to be recognized as a narrative at all, must be largely comprehensible.

2. INTERPRETATION

According to Davidson, correct interpretation of a speech act reflects at least part of the intention with which the act was performed.[11] The intention might be, for instance, the intention to inform the hearer that Davidson holds a theory of speech act interpretation in which intention is prominent. An interpretation determines what the words mean and what the person is doing in using the words.

"Intention," for Davidson, does not represent a foundational, self-interpreting mental "language of thought," but rather picks out one of the states of the "mental" family of concepts that interpretation employs in ascribing speech acts and other actions to a being who is treated as a rational agent. Intentions are ascribed along with beliefs, desires, and meanings of utterances, in a holistic way.[12] Intentions are not foundations of meaningfulness in virtue of being magic language inscriptions that are essentially meaningful and require no interpretation, but are rather on a par with beliefs, desires, utterances, and meanings. None of these notions is explicable without reference to the others; no one notion is the bedrock from which the other notions are constructed or understood. Thus, although some of Davidson's account sounds rather like that of Hirsch (1967), Davidson's theory does not suppose that intentions are made up of some version of "logoi" that lie behind words.

How does an intention attach to an utterance so as to make a meaningful utterance? A meaningful utterance is an attempt to communicate. An attempt is qualified as an attempt only if the author believes that the utterance can succeed in being understood. In the same way, my tapping on an egg is not an attempt to fry the egg unless I believe that my tapping will succeed. Not every coincidence of action and desire is an attempt. In the case of communication, the relevant beliefs connecting the utterance and the result must include beliefs about the audience and its knowledge and capacities.

Davidson is thus able to accommodate the sensible thought that, for instance, Shakespeare's text means (by and large) what Shakespeare meant by it, without supposing that 'what Shakespeare meant by it' is a self-interpreting intention, a mental state whose interpretation is given by its very nature.[13] Intention itself may be indeterminate, but relative to a text disconnected from a producer, a text with a producer has a more determinate meaning.

Texts with authors are likewise interpreted as products of actions. The intention that is the reason for the action gives the correct interpretation of the text act. Intentions themselves are textlike, that is, language-like, in contrast to alleged logoi, but a text as a text act of an author is relatively determinate. That is, since we can sometimes know what a person intended in writing, even though the words in isolation would be subject to numerous interpretations, the author's context and intentions narrow the indeterminacy of the text. While it is true that the text is "legible after the death of the author," it is less determinately legible, in the general case.

It is still true for Davidson that the meaning of a speech act or writing act is not entirely a matter of what is in the speaker's or writer's head. Since the point of speech and writing is communication, and an intentional action is an attempt only if there is some reasonable expectation of success, there have to be sufficient clues in the text that the intended audience can reasonably be expected to get the message right. In effect, this means that what the speaker's words mean depends, at least in part, on the text and the context. Davidson denies Humpty Dumpty's intentional theory, which says that a speaker can mean whatever he chooses by his words.[14] Language use and comprehension is part of getting around in the world. One's knowledge of one's fellows is an essential component of communication. One's reasonable beliefs about one's fellows are essential to attempted communication, that is, to any use of language at all.

We can now state more clearly the project of interpretation. The data of interpretation are speech acts or writing acts. The project of interpretation is to find out what the speech acts are. For instance, the utterance 'The rose bush is right behind you' might be correctly interpreted as a warning that Fred is about to run over the rose bush, if that was the intent of the speaker. Interpretation is thus essentially simultaneously rhetorical and "logical." We are given actions in our common world. We interpret by determining truth conditions and the reasons that utterances with those truth conditions are being presented.

Given this notion of truth conditions and meaning, Davidson has available the following picture of the various rhetorical forces with which things are said and written: an utterance is the production of (in the standard case) a sentence, with truth conditions,[15] with an intention. In fact, there is usually a nested array of intentions providing the reason for producing a sentence. I say 'Your pants are unzipped' to let you know that your pants are unzipped, to embarrass you, to get back at you for what you did yesterday, and so forth.

Davidson treats the various "speech act types" as utterances having truth values and truth conditions, but as offered for other reasons than informing the hearer that the truth value is truth. So, warnings, declaratives (such as 'I hereby declare you man and wife'), and other speech acts that are treated as neither true nor false by Austin and Searle are treated as having truth values. Theories according to which such acts have no truth values, according to Davidson, confuse truth valuedness with assertion. For Davidson, the truth conditions of 'You are about to run over the rose bush' are what make the warning a warning that you are about to run over the rose bush. One utters such a sentence for a variety of purposes, but perhaps a typical purpose

would be warning. Even utterances that are never uttered to inform still have truth conditions or satisfaction conditions. Truth conditions or satisfaction conditions give the "conceptual content" of utterances with every rhetorical force.[16]

Davidson denies that conventions determine what speech act is being performed (Davidson 1984 [1983]). His argument against the efficacy of conventions is that, if there were a conventional linguistic device indicating a particular force, that device immediately would be available to be used in speech acts with other forces.[17] The device could be used on stage, for instance. Rhetorical force cannot be in the text qua text. The general point is that linguistic devices are equipment to be used for purposes. The equipment itself does not fix the purpose; rather, given the properties of the equipment, various purposes can be served. The properties of sentences that make them usable for the various purposes to which they are put are their truth conditions. The truth conditions of 'This room will be vacuumed by this evening' allow that sentence to be used to command that the hearer vacuum the room, as well as to predict that the room will be vacuumed.

3. METAPHOR AND OTHER FIGURES

Davidson's best-known contribution on a topic that is clearly part of literary theory is his account of metaphor.[18] Given that the meaning of an utterance is given by the truth conditions assigned in an appropriate truth definition, there is nothing for words to mean except what can be given in words themselves. If there were such a thing as special metaphorical meaning, that metaphorical meaning would be given with a truth definition using words. If there were metaphorical meaning, then metaphors could be paraphrased without loss. However, virtually every serious thinker on metaphor (including Davidson) agrees that metaphors cannot be exhaustively paraphrased. If there were meanings that metaphors expressed, those meanings could be expressed literally, by coining a word; but this seems obviously to miss the mark.

Davidson thus treats metaphor as a rhetorical phenomenon: a metaphorical utterance is an utterance of a sentence that is usually literally false, but is uttered for the purpose of bringing something to the attention of the hearer.[19] The description of the utterance or text as metaphorical describes the intention of the utterance or text. It is a further question whether or not the metaphor is an effective instrument for that intention, just as it is a further question about a foul shot whether or not it went through the

hoop. So a person can utter the sentence 'The war reformatted the hard drive of their civilization' with the intent of getting his audience to see the war's destruction of a culture in a certain way. This metaphor would be inept and ineffective, for reasons that might be plumbed, but Davidson would argue that there can be no general theory of what makes metaphors wonderful and what makes them ludicrous. His theory is just designed to provide a framework for "metaphorical meaning" that avoids supposing that there is some kind of semantical meaning that metaphors have that makes them metaphors or that carries their significance. Metaphorical meaning is a rhetorical, not a semantic phenomenon.

Davidson compares the effect of a metaphor – getting a person to see something in a new way – to that of a bumping or other nonsemantic event. Specifically, metaphor is a speech act intended to get another to see something as something else – to see what the author of the metaphor sees. Davidson points out that seeing-as does not cash out into propositions: seeing men as pigs does not cash out into some finite set of propositions about men, derived from the properties of pigs. Seeing-as is not propositional at all, in fact. In Davidson's words: "a picture is not worth a thousand words, or any other number. Words are the wrong currency to exchange for a picture" (Davidson 1984 [1978], p. 263).

We can note that the success of the intention of a metaphor will depend on the correctness of the author's assessment of his audience.[20] The speaker's theory of the audience leads him to reasonably expect that saying 'Men are pigs' will lead them to see what he sees about men. The intention of a metaphor must be something that the author can reasonably expect to accomplish, just as an intended action of any kind must be. A metaphor, like a joke, generally requires a context. The reception of 'Margaret Thatcher needs a heart transplant' will depend on whether the circumstances and the audience are medical or political. Knowledge of the audience, and the prediction of the effects of the words, is not part of semantics, according to Davidson, in part because it is part of the speaker's interpretation of the theories of the audience. Roughly, the maker of a metaphor must know how his audience feels about pigs in order to be successful. But this body of knowledge is not a set of propositions.

Davidson's account holds that a metaphor, qua metaphor, has no special cognitive content. That is, 'Men are pigs' has no more propositional content than is given by the formula, ' "Men are pigs" is true if and only if men are pigs'. The new way of seeing men that the author may induce in an audience is not a cognitive content. Many would argue that paraphrase and explication give (admittedly partial) accounts of the cognitive content

contained in the metaphor. Such paraphrases draw parallels, point out resemblances, and show why the metaphor is appropriate for changing one's view of the subject. On such accounts, metaphor depends on, and implicitly has as content, the features of men and pigs that might make saying 'Men are pigs' effective. Davidson's account of such helpful paraphrase is that it is just another way to get an audience to have the "seeing as" experience that the metaphor is designed to produce.

How can metaphors not have special metaphorical cognitive content, given that real-world connections and the beliefs and desires of the audience seem to be called into play by understanding a metaphor? A relevant analogy would be to jokes. Jokes, like metaphors, can be successful or unsuccessful. When I say to my students, "Your final grade will be determined by my mood while I'm grading your exam," many of my students may take me seriously and have their worst suspicions about professors confirmed. On the other hand, I could be speaking truthfully. The propositional content of a lame joke differs not at all between someone who gets the joke and someone who doesn't. Even very well-known and highly regarded jokes go beyond the text. Youngman's "Take my wife. Please," qua text, could be completely serious.[21]

So to call an utterance or a text metaphorical is to describe an intention of the author. The author may have further intentions, consisting of his reasons for wanting his audience to see the world in a particular way. Various other meanings, given by the layered intentions with which that utterance was produced, yield other notions of meaning.[22]

The account of figures other than metaphor would in many cases be analogous. A metonymy, for example, would likewise be an utterance uttered without the intent of asserting that is it literally true. 'A hungry stomach has no ears' is, of course, true, but it is not written with the intent of pointing out that stomachs don't ever have ears, so that, a fortiori, hungry ones don't. Jokes, sarcasm, irony, and so forth do not change cognitive content; they characterize the intention. Many other rhetorical devices counted as figures by theorists of literature[23] are intentional violations of conventions of syntax, punctuation, spelling, and so on. Few would claim that misspellings convey distinctive cognitive content, for instance. Davidson's theory – which, as we will see, denies any constitutive or essential function to conventions – can be seen as assimilating metaphor and metonymy to such figures.

Broader-scale literary phenomena such as parody and irony could be treated in much the same way. Just as some metaphors could hardly be anything else, given that the author is a human being with normal beliefs about

his audience, so some works of literature can be known to be parodic without special knowledge of the author. Some interpretations need use only what Davidson calls our "prior theory," the general theory we have of a speaker before knowing anything about him except that he speaks English. The text of *Tristram Shandy*, for instance, is rich enough in data for interpretation that there is little question that it is supposed to be humorous.[24]

In other cases we may not be so sure, and only information about the author would tell us. It is possible that the entire corpus of the works of Kim Il Sung, for example, could turn out to be a parody of Stalin's writings, if it turned out that Kim Il Sung was actually a sophisticated but totally cynical and evil person. Likewise, some of the effusive dedications to royalty one finds in seventeenth-century texts would be parodic or ironic if they were not formulaic, and if the power relations between author and sovereign were not taken into account.[25]

4. PUBLIC LANGUAGES, INTENTIONS, AND THE AUTHOR

Davidson holds that the central feature of language use is communication. Communication takes place whenever there is successful interpretation – that is, whenever the audience gets the speech act or writing act right. He conceives of "communication," sensibly enough, as a speech or writing act that results in successful interpretation by an audience. The audience gets what the author had in mind. Given Davidson's notions of the mental, intention, and interpretation, this "communication" is not a matter of conveying self-interpreting logoi to another person. The correct interpretation of a speech act does indeed convey the intended meaning, but the intended meaning is determinate only to the extent that radical interpretation could arrive at limited options for "best interpretation." That is, since meaning is nothing over and above what people would do and say in which circumstances – even though it is not reducible to such patterns of circumstances – "correct interpretation" and "communication" are internal to the intentional family of concepts.

The central issue between Davidson and the majority of modern literary theorists concerns the role of the conventions, norms, and social practices connected with language use. Most contemporary literary theory, as well as most contemporary philosophy of language, regards language as essentially a social phenomenon. The fundamental notion of "language" for such theories thus refers to entities such as "English" and "Czech." Such languages are standardly regarded as systems of collective practices

governed by semantical rules. Quine, Davidson, and Derrida argue that the "semantical rule" idea is either a mystification or an appeal to logoi – that is, language-transcendent meanings that various languages express in various ways. Derrida and other "public language" theorists therefore regard languages as more loosely systematized collective practices.

For theories that deny such logoi but retain the notion that language is primarily social, the meaning of a linguistic expression, to the extent that it is determinate, is a function of its role in the culture at a given time. Thus an individual author or speaker is a speaker of a language, which has meanings that are beyond that speaker's control. An author's intentions to mean something by her words can only be intentions to mean what the culture means, since the meaning of her words is determined by public and cultural phenomena. What she means and what her words mean are identical. What she means is given by what her mental text says according to the public language.[26] That is, the "public" meaning, rather than her intended meaning, determines the meaning of what the author wrote. So one can say that it is the language that is speaking or writing, according to the possibilities that that set of cultural practices determines.[27]

As Davidson[28] and Chomsky[29] have observed, a language construed as a set of social practices is an ill-defined and vague entity. Davidson does not deny that there are speakers of English and speakers of Czech. His view is that such languages are to be identified (vaguely) with sets of overlapping idiolects. Cultures and populations are never uniform in vocabulary, nor in the beliefs and desires that form the basis for the conceptual connections among items of vocabulary. Any two individuals will thus mean different things by their words, if we take "meaning" to be the inferential and valuational connections that speakers make. Without the existence of logoi, that is, there is no difference in kind between inferences that are due to the beliefs and desires that people have and inferences that follow from the meanings of their words. Without the analytic-synthetic distinction and the "rules" that mysteriously establish logoi, the social nature of language, given the diversity of beliefs and desires within a society, entails indeterminacy of meaning and sense within that language. So, relative to a conception of language as social and meaning as practices, it would be no surprise that interpretation is indeterminate.[30] If a text is legible in the absence of the author, and texts have meaning on the basis of social relationships and practices, and social relationships and practices differ within a society and over time, then a text's meaning likewise shifts.

There are two responses to this paradoxical result of the combination of the denial of logoi and the thesis that language and meaning are a function

of cultural practices: one could accept it as a surprising discovery. The tradition has thought either that meaning was fixed by individual thoughts or that languages were somehow systems of rules administered by some group spirit. Both of these suppositions have been theorized, since Plato, as various forms of direct relationship to logoi. If language and meaning are social, and only logoi-supported rules could determine meaning in an objective way, then the meaning of a person's utterances and writings is not a function of individual intentions but a matter of social phenomena. Since the social is constantly shifting and incompletely determinate at any given time, meaning itself shifts politically and in other ways over time and across audiences at a given time. Thus the meaning of any writing or utterance is not determinate absolutely, but determinate only relative to an interpretation or a reception. This is the reaction of popular[31] "postmodernism."

The other response is Davidson's: since we sometimes do understand perfectly clearly what a text or a person means, the individual's meaning must be primary, since the network of social practices – that is, the collection of the community's idiolects – would never allow a determinate meaning. Davidson would agree that texts without authors are less determinate than texts with authors. Committee documents, the Torah, and quoted fragments are more indeterminate than an individual utterance or a novel by Agatha Christie.

There are two independent theses at issue. First, there is the issue of whether the truth conditions of utterances are determined socially or by individuals. Davidson, unlike the vast majority of philosophers of language, holds that languages can be private, in the sense that the truth conditions of what is said are determined by individual intentions rather than by the practices of a linguistic community. Second, there is the issue of whether there are logoi, or some other reduction of meaning to something else, that allow fixing the truth conditions of utterances. Those who think that there are such fixers of truth conditions include many naturalists, most Platonists, and some of those who view language as a system of norms. Those who deny such fixing of truth conditions by reductions or by mystification of "language" or by logoi hold that, while meaning is nothing over and above what people say in which circumstances, there is no reduction of meaning to such circumstances, social or otherwise. Such theorists, who include Derrida, Quine, Davidson, and perhaps Wittgenstein, hold that the assignment of meaning is sometimes indeterminate.

Davidson denies that the truth conditions of an individual's utterances are socially determined, but he also denies that there are logoi or any other such reductions of linguistic meaning to something else. His argument that

thought and language depend on the existence of other speakers denies that logoi are present to the mind. Derrida, on the other hand, holds that a language is a social entity. He takes the denial of logoi to mean that indeterminacy is pervasive to the extent that the public language, in which thoughts are expressed, is pervasive. Derrida does not deny that context can narrow indeterminacy, but he argues that the remaining indeterminacy is, while "governed" by dissemination and differance, present in every utterance and thought. That is, since the language of thought is the public language, and the public language is shifting and indeterminate at any given time (albeit not randomly), Derrida takes all linguistic events to be as indeterminate in meaning as Davidson takes texts detached from intention to be.

The real issue between Davidson and Derrida is the primacy of the idiolect over the public language. If the public language were primary – and if the public language is as vague, variable, and indeterminate in itself as both Davidson and Derrida acknowledge – then literary texts would indeed be radically indeterminate. The author would be writing in a language whose meanings depended on much more than his interior thoughts and intentions. That is, since the language is public, and since nothing unambiguously determines what that public object means, the text itself would be subject to exactly the kind of dissemination and drift that Derrida describes.

Davidson's main objection to the "social" nature of language, as it is usually conceived, is that uniformity of vocabulary or syntax is not necessary for communication. If understanding communication is primary for understanding language, then agreement on vocabulary and syntax – that is, speaking according to some average of the society (or according to the "norms" of the society) – is not essential to language.

Davidson argues that interpretation often succeeds when there is intentional or unintentional violation of norms and regularities. His arguments can be regarded as extensions of his rhetorical view of figuration. Just as we can be understood when we utter falsehoods in speaking metaphorically, so we can be understood when we use idiosyncratic vocabulary.

Davidson's article "A Nice Derangement of Epitaphs" argues that, because there is no "violation" of conventions or norms that interpretation cannot succeed in spite of, such conventions and norms are not essential to language use. Admittedly, regularities make interpretation easier and more routine. With familiarity, we can understand a deviant speaker or a group of deviant speakers with ease. The "normal" situation, in which someone is "speaking perfect English," is just the limiting case of such familiarity. Even in this limiting case, though, there are differences in vocabulary, nuance, and general knowledge, so that the "same language," in the

philosophical logician's sense, never obtains between two persons' idiolects in real life.

Davidson does argue that language is social, as explicated earlier: without other people to establish an objective world, and thus without a contrast between belief and true belief, there is no thought or language. The "norms" of language are for Davidson like rules of thumb – rough guidelines that are used to facilitate communication, but ones that can be successfully violated. That is, the rules of language are not like the rules of chess, except to the extent that "develop knights before bishops" is a prudential rule of chess.

Thus, contra Derrida, Davidson in effect privileges speech, in the sense that speech and speechlike texts are determinate in a way that texts detached from authors are not. However, unlike others whom Derrida attacks, Davidson does not hold that an author or a speaker resolves all indeterminacy by supplying the logoi behind the words. So, while Davidson would deny that texts are legible (accurately) in the absence of the author – that is, considered apart from the author, her intentions, and her context – he would agree with Derrida that iterability is essential to language. "Iterability," given that the language is an idiolect, just means that other possible tokenings in the idiolect-at-a-time, those that a passing theory would understand correctly, would have the same meaning.[32]

Davidson's view is rather that the actions of a speaker are relatively determinate – that is, determinate relative to texts apart from authors. In other words, on Davidson's view, intentions do not fix interpretations by being in a magic language, since they depend on language, but they nevertheless fix interpretations much of the time.

So, Davidson conceives of a literary work as a written utterance in the idiolect of the author. A text is interpreted, that is, in a way analogous to the way we interpret a speaker.

5. LITERARY THEORY AND CREATIVITY[33]

Davidson would acknowledge that much illuminating literary criticism has operated under the guidance of theories that take language and meaning to be based on rules or loosely formulated features of a public language. Accounts using notions of intertextuality, allusion, and so forth give accurate interpretations, seemingly on the basis of taking the public language to be primary. An account of how an important poet stretches the form of the sonnet seems to imply a set of rules for being a sonnet. Allusion and citation seem to require a public object that is shared.

Davidson can generate any explanation that a public language theorist generates using a public language while still taking the idiolect to be primary. Davidson's idea is that the "rules of language," taken by such theorists to be constitutive of the public language, are actually prudential rules and methodological guidelines for interpreters and those who wish to be accurately interpreted. Since the primary function of language is communication,[34] a speaker or writer will, given an intention to communicate, communicate in ways that he reasonably expects will be understood by his audience. Since the audience has theories about how authors will write, and the author shares those theories and knows that his audience holds them, the effect of "rules" is achieved just by mutual expectations, together with the intent to communicate. So, as Davidson remarks, most of the time it suits the intentions of a communicator to speak and write in ways that the audience will comprehend without difficulty.

At the most basic level of such methodological guidelines, syntax and lexicon, it usually behooves an author to use most of his words in the ways the audience does, employing more or less familiar syntax, so that the audience can use the "homophonic" interpretation scheme, interpreting the author's sentences by identical sentences of their own.[35] Authors adjust to the audience. Letters to the editor use different vocabulary than professional journal articles. We may want to emphasize or deemphasize class differences, perhaps using 'It don't matter none to me' (if it can be done effectively) to establish solidarity in a biker bar. The "rules" are like the rules of painting. When we learn to communicate, we learn what will happen in what circumstances given what utterances. Then we select something appropriate for our ends, just as an understanding of color relationships allows a painter to achieve the effects she wants.

The more "difficult" the author, the more interpretation is demanded of the audience. James Joyce's *Finnegans Wake*, for instance, is a work that demands a great deal of interpretive skill on the part of the reader. Davidson's account of Joyce is a demonstration of how his theory is ideally suited to handle idiosyncratic syntax, vocabulary, obscure allusion and citation (Davidson 1991b).

What would genres be, according to Davidson? If one desires to communicate, it is helpful to have ways of letting the audience know what sort of thing is being communicated. Most of the time, we can identify novels at a glance, tell that a work is a work of poetry, or detect an academic article. To write an academic article in philosophy for an analytic audience is to include footnotes, arguments, numbered sections, and so on, and to use words with the intended extensions of those same words as spoken and written by the

audience. Authors need not be regarded as under any obligations imposed by the culture or by the language, broadly construed, to write philosophy, for random example, in this way. For the existence of "norms," it suffices that reasonable authors who are aware of the expectations of their readers will write in a way that aids readers in interpretation. Of course, highly admired authors can be idiosyncratic, since they can reasonably expect an audience to take special pains in understanding. Nietzsche and Wittgenstein spring to mind.

Davidson's account of poetic forms would be similar to his account of genres. An author wants his work to be identifiable as a sonnet, for instance, and to be evaluated as such. A violation of the rules for being a sonnet must have a product that is close enough to being a sonnet that the intention is recognizable, in order to be evaluated as a sonnet. Davidson's account of intertextuality and allusion would have to observe only that part of the knowledge an author acquires of his audience is what they know and what they've read.

Everything that can be explained by appeal to the rules of a public language can likewise be explained by the "prudential" guidelines of successful communication. The difference between Davidson and a "public language" theorist is that, according to Davidson, the adherence to regularities and expectations, although practically useful much of the time, is not essential. Davidson has an easier task explaining the deviations from "norms" that we find in interesting literature, since his authors are not constrained by anything in particular; they can find innovative ways of communicating without stooping to saying what everyone else says. Of course, there will also be writing that tries to accomplish unchallenging comprehension. Prescriptions for medication, permission slips, sabbatical proposals, and other such writing, where there is a high premium on communication and a high cost to miscommunication, will reasonably be expected to proceed in ways that require little innovative interpretation. Literature, however, is often intended to be interesting.

The fundamental thrust of Davidson's later writing on language and literature is that language is not bound by convention, by rules, or by precedents. This does not mean that anything goes. We will close by considering what the conditions and limits are, according to Davidson, on an author's meaning something idiosyncratic by his words.

Meaning something by some words is trying to do something – specifically, trying to communicate. Just as in action, there has to be some believed connection between the behavior and the result. More is required than a simultaneous want and a behavior. So I can try to bring rain by

reciting a spell, but not by blinking my eyes and simultaneously wanting it to rain. The beliefs that make the behavior a 'trying' do not have to be correct, but they have to be ascribable as beliefs, and thus they have to have some connection to true beliefs. Trying to do something is not the mysterious attachment of a desire to a behavior.

The same applies to speech and writing. Meaning something by one's words is not a mysterious mental attachment of a message to an act, but an attempt to get something across. If a speech act or writing act is essentially an attempt to communicate, and if an attempt must be something that the actor can believe has some chance of succeeding, then Humpty Dumpty did not actually mean 'a knockdown argument' by 'glory', at least as far as Carroll's story indicates. However, if Humpty was deluded about his audience, and had some unusual correlation in mind – for instance, that in his favorite novel, 'glory' occurred in the same position as 'knockdown argument' but on the verso of the page, and this seemed to Humpty something everyone would notice – then we could say that Humpty's communication attempt was meaningful, but unsuccessful. It is a familiar situation in literature that something about a text indicates a hidden message, but that the author has provided too few clues for any actual interpreter. These texts are meaningful, but they have no determinate meaning.[36] Davidson's account treats understanding language, texts, and utterances as of a piece with understanding anything else. We get information and make an inference. But there is no algorithm for such inference.

Notes

1. The exceptions are his Davidson 1991b and Davidson 1993a.
2. Brooks (1947, pp. 72–3), writes:
   > The poem communicates so much and communicates it so richly and with such delicate qualifications that the thing communicated is mauled and distorted if we attempt to convey it by any vehicle less subtle than that of the poem itself. . . . if we are to speak exactly, the poem itself is the *only* medium that communicates the particular 'what' that is communicated. The conventional theories of communication offer no easy solution to our problem of meanings: we emerge with nothing more enlightening than this graceless bit of tautology: the poem says what the poem says.
3. The fundamental paper on this topic is Davidson 1980c [1970]. See also Davidson 1990d.
4. The fundamental paper on this topic is Davison 1984 [1967]. See Chapters 1 and 3 for further discussion of the role of a truth theory in a meaning theory.
5. Such a theory would be an account of the contribution of segments of sentences to the truth conditions of whole sentences, i.e., an account of how the truth

conditions of whole sentences are built up from the satisfaction conditions of predicates. What simple predicates mean is given by a word that means the same, or by a compound predicate that means the same. Such a thesis – that words are as good as "representation" gets – means that words are explicated in words, not by relating them to some other items. Davidson appeals neither to forms nor to sets of sense data nor to stimulations in accounting for the meaning of logically simple predicates such as 'is a frog': 'is a frog' is true of an object if and only if that object is a frog. See Chapter 1 for further discussion.

6. The basic point, that there cannot be an algorithm for "inductive logic," is due to Carl Hempel. He argued that, since some adjustments to new data require new theoretical terms, and since new theoretical terms cannot be produced by an algorithm, there can be no routine for induction. An inductive logic, after all, would be a theory that gave a measure of the conditional probability of *A* given *B* for arbitrary *A* and *B*. It is not surprising that there can be no such algorithm. Interpretation, like getting along in the world in general, is more than following a program.
7. Davidson's argument is given in several places. See Davidson 2001 [1992] and Davidson 1994c.
8. For Davidson's most succinct exposition of triangulation and its consequences, see his Davidson 1994c, p. 15.
9. "Prior" and "passing" theories are discussed in Davidson 1986c. A "passing theory" is typically an adjustment in a "prior theory" – that is, the theory we had before actually encountering the person. The "theory" we form of another will of course be incomplete, but it will involve hypotheses about what the person would say in other circumstances. Counterfactuals about another's speech behavior thus enter Davidson's account at the ground level of understanding of language, and so at the ground level of the possibility of thought. To speak of "theory" in this context is to describe how we would characterize a certain ability. Davidson does not imagine that interpreters have available to them anything like an axiomatic system. However, given an ability to understand, a theory would be our representation of that ability.
10. "Logos" (plural "logoi") is Derrida's term for items of thought about which there is no question of interpretation. Such terms, which I have characterized elsewhere as "magic language" terms, mean what they mean in virtue of their very essence. Locke's "ideas" and Frege's "senses" are such terms.
11. Intentions are layered. A liar's intention, for instance, is to make an assertion that *p*, which assertion itself has purposes. At every level, "correct interpretation" gets the intention right.
12. "Holistic" amounts to the following: a family of concepts is applied on the basis of data, but all of the data are "criteria" for all of the concepts. For example, when we observe some speech acts, we interpret the acts by assigning beliefs and desires to the agent and meanings to the words. The same set of data, that is, determine the application of all of the concepts. There are constraints on how those assignments are done – the "maximization constraints" – that make the speaker believe mostly truths, seek the good, be consistent, and so forth.

Indeterminacy, for Quine and Davidson, results when the maximization constraints determine no single best solution. Since beliefs, desires, and other members of the "intentional" family of concepts have their use exhausted by their role in keeping track of behavior, there is no "fact of the matter" regarding what a person believes, desires, or means when there is indeterminacy – or, rather, no fact of the matter regarding which of the various theories that meet all the constraints is correct. All do an equally good job in capturing all the facts there are.

13. The relationship between Davidson's view and Derrida's argument in Derrida 1977 is complicated. Derrida is sometimes taken to argue that, since there is indeterminacy even in the interpretation of an intention, unless there are logoi fixing the meanings of inner states, meaning is therefore loosened to the extent that no text or speech act has a determinate meaning. If this means that we never know what someone's utterance means, it seems uncharitable to ascribe this view to Derrida. Unless determinate meaning requires the absolute anchors of logoi, there is no reason to suppose that denying that there are logoi amounts to denying that meaning is sometimes determinate. Davidson and Derrida, on my interpretation, agree in being pragmatists about the ascription of meaning. Although nothing anchors meaning that is not itself anchored by its relations to other unanchored states of the agent, agents still sometimes are correctly interpretable. Both Davidson and Derrida recognize that there is indeterminacy sometimes, and that "dissemination" and drift of meaning is thus a fact. The differences between them are differences of emphasis. Davidson, starting from Quine's indeterminacy of translation and its denial of mentalistic foundations, is concerned to rescue the commonsense view that we often know what people are doing and saying. Derrida is primarily concerning to argue against mentalistic foundations.

    Because intentions have logical structure (an intention to go to the store and to the gas station has as a component an intention to go to the gas station), they are essentially language-like. Construed as entities or states, intentions can be understood only by assigning them conceptual content, i.e., by giving truth definitions for them.

14. See Davidson's discussion of Donnellan's reply to McKay in Davidson 1986c, p. 439.
15. In order to cover all utterances, there are obvious generalizations of Davidson 1984 [1968] to the presentation of open sentences. For instance, 'Whom did you dance with?' could be treated as the presentation of the open sentence 'You danced with __' with the intention that the hearer will fill in the blank. The mood indicator would be treated as in Davidson 1984b [1979]. Note 12 of Davidson 1984b [1979] sketches a theory of questions. That note indicates that there is much to be done with the semantics of 'How many frogs invaded your living room?' Such utterances as 'Jesus H. Christ!' would pose other problems.
16. As remarked in note 15, Davidson's program remains incomplete in the case of questions and expletives.
17. This point has been made again and again. See Wittgenstein's *Tractatus Logico-Philosophicus*, 4.442 (Wittgenstein 1961).

18. The fundamental article on metaphor is Davidson 1984 [1978].
19. Since "literally false" would often be understood as "false in the public language," and since Davidson's distinction is supposed to apply primarily to an idiolect in which a speaker may use a word idiosyncratically but mean it, Davidson has changed his terminology to "first meaning." The first meaning of someone's utterance is given by the truth definition applied to that utterance.
20. For a farm boy from Ohio, seeing men as pigs may raise his estimation of men, since pigs are intelligent, clever, and resourceful, relative to other domestic beasts. In addition, pigs are rarely malevolent and seem not to be obsessed with domination of other pigs, except in practical matters such as getting food. A metaphor may misfire if the author is mistaken about his audience.
21. A mad killer has cornered a couple. He vows his intent to kill one or the other. The ungentlemanly husband says: "Take my wife. Please."
22. Still further notions of meaning fall within the scope of the English term 'meaning'. For instance, what the utterance reveals about the author's culture, gender attitudes, economic position, and so forth would be "meanings" of his utterance. Davidson has nothing special to say about such notions of meaning, except to observe that there is nothing especially literary about them. A person's stock market transactions or his pattern of daily life would have similar "meanings." What it says about a person that he bicycles to work instead of driving is analogous to what it says about an author that he writes experimental novels.
23. The most entertaining of the modern treatises on this topic is Arthur Quinn's (1982).
24. The question of whether the humor is strictly in the text, though, is not so clear. Borges (1965) argues that who the author is, and his historical context, is part of what determines the rhetorical character even of a text as complex as *Don Quixote*. Borges imagines that a post–World War I author writes the very text of *Don Quixote*. As a text of that time, various passages will reasonably be taken differently.
25. Note that this would not be a question about how the author really felt about his sovereign, but rather a question about what the author intended to communicate.
26. Supporting this view, for example, would be sentences such as 'May I have your hand', which might be a proposal, if uttered in a context with an appropriate woman, whether or not the speaker was serious.
27. An influential advocate of this view of language is Roland Barthes. In Barthes 1967, p. 14, explicating Saussure's concept of "langue," Barthes writes: "It is the social part of language, the individual cannot by himself either create or modify it; it is essentially a collective contract which one must accept in its entirety if one wishes to communicate." The kernel of truth in this conception – that I have been rude when I characterize a woman as a broad, even though I was under the impression that the term meant "having wide interests" – would be accommodated by Davidson's accounts of how we form theories of our interlocutors, and of how successful intentions require some luck in estimating the effects of what we say.

28. Davidson 1986c, pp. 444–5:

> A person's ability to interpret or speak to another person consists [in] . . . the ability . . . to construct a correct, that is convergent, passing theory for speech transactions with that person. . . . This characterization of linguistic ability is so nearly circular that it cannot be wrong: It comes to saying that the ability to communicate by speech consists in the ability to make oneself understood, and to understand. It is only when we look at the structure of this ability that we realize how far we have drifted from standard ideas of language mastery. For we have discovered no learnable common core of consistent behaviour, no shared grammar or rules, no portable interpreting machine set to grind out the meaning of an arbitrary utterance. We may say that linguistic ability is the ability to converge on a passing theory from time to time. . . . But if we do say this, then we should realize that we have abandoned not only the ordinary notion of language, but we have erased the boundary between knowing a language and knowing our way around in the world generally. For there are no rules for arriving at passing theories, no rules in any strict sense, as opposed to rough maxims and methodological generalities. . . . I conclude that there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed.

29. Chomsky (1995, p. 48) writes: ". . . common, public language . . . remains mysterious . . . , useless for any form of theoretical explanation."
30. Whether further, even more radical results follow is questionable. Derrida supposes that the changes in a public language are not random, but operate by differance and dissemination. It is not clear to me whether this means more than that the changes are nonrandom. Derrida certainly holds that they are not predictable. Whether it follows from this that texts always have alternative, equally good interpretations; that we never know what someone means except relative to a choice of interpretative scheme; or that what someone has written depends on its reception, would depend on further arguments with other premises.
31. I exclude Derrida from this category. Derrida's position on meaning, determinacy, and the social is too subtle to be dealt with here. He does hold that texts are legible apart from their authors, that language is essentially social, and that there are no logoi; but his conclusions are nuanced beyond summary.
32. The "passing theory" for Davidson is the description of someone's ability to understand another. Davidson does not suppose that such theories are formulated by the interpreter. Yet they are not like the "theory" that our eyes and brain use to construct objects out of visual inputs. Theories are sought for and arrived at by the standard ways that we have of understanding one another. For Davidson, like Heidegger, understanding is primitive.
33. I should point out that many of the "Davidsonian" theories offered in this section are extrapolations rather than summaries of views that Davidson has actually expressed.
34. See Davidson 1994c, p. 11: ". . . what matters, the point of language or speech or whatever you want to call it, is communication, getting across to someone

else what you have in mind by means of words that they interpret (understand) as you want them to. Speech has endless other purposes, but none underlies this one. . . ."

35. Except for demonstratives and proper names.
36. Debates about the real meaning of Revelation or about the occurrence of Jewish festivals in the Gospel of John are familiar examples. When these are taken to be holy texts, the issue is as serious as the issue of code breaking. Just as we know that the enemy is communicating something, so we know that the allegorical aspects of Revelation mean something.

Bibliography of Davidson's Publications

Davidson's most influential work has appeared in the form of relatively short essays. Most of these essays have been or are being collected in five volumes of essays published by Oxford University Press. The following bibliography is organized into five sections. Section I lists the volumes of the collected essays, including those projected for publication. Section II lists Davidson's major philosophical publications in essay form, giving details of their first publication. The volume in which an article is collected is indicated by a superscript arabic numeral attached to the date; the number represents the number in order of publication of the volume of collected essays in which it appears or will appear. Section III lists books Davidson has authored or coauthored. Section IV lists books Davidson has coedited. Section V lists videotaped conversations with Davidson. Author/date citations to Davidson's papers in the text should be referred to the Bibliographic References. An extensive history of reprintings of Davidson's essays can be found at the end of the volume on Davidson in the Library of Living Philosophers series, *The Philosophy of Donald Davidson* (Chicago: Open Court, 1999).

I. Davidson's Collected Papers

1980. *Essays on Actions and Events*. New York: Clarendon Press. Second edition published in 2001. The second edition adds two appendices, "Adverbs of Actions" and "Reply to Quine on Events."

1984. *Inquiries into Truth and Interpretation*. New York: Clarendon Press. Second edition published in 2001. The second edition adds an appendix, "Replies to Lewis and Quine," to Chapter 10, "Belief and the Basis of Meaning."

2001. *Subjective, Intersubjective, Objective*. New York: Clarendon Press.

Forthcoming. Volume 4, *Problems of Rationality.* New York: Clarendon Press.

Forthcoming. Volume 5, *Truth, Language and History*. New York: Clarendon Press.

II. Articles

1952. Why Study Philosophy? *View Point, 2*, 22–4.

1955. Outline of a Formal Theory of Value. *Philosophy of Science, 22*, 140–60. (With J. J. C. McKinsey and Patrick Suppes.)

1955. The Return of Reason in Ethics. In T. Kimura (Ed.), *Analysis of the American Way of Thinking*. Tokyo: Tokyo University Press. (In Japanese.)

1956. A Finitistic Axiomatization of Subjective Probability and Utility. *Econometrica*, *24*, 264–75. (With Patrick Suppes.)

1959. Experimental Tests of a Stochastic Decision Theory. In C. W. Churchman and P. Ratoosh (Eds.), *Measurement: Definitions and Theories* (pp. 233–69). New York: Wiley and Sons. (With J. Marschak.)

1963.[1] Actions, Reasons, and Causes. *The Journal of Philosophy*, *60*, 685–99.

1963. The Method of Intension and Extension. In A. Schilpp (Ed.), *The Philosophy of Rudolf Carnap* (pp. 311–49). La Salle, Ill.: Open Court.

1966.[1] Emeroses by Other Names. *The Journal of Philosophy*, *63*, 778–9.

1966.[2] Theories of Meaning and Learnable Languages. In Y. Bar-Hillel (Ed.), *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science* (pp. 383–94). Amsterdam: North Holland.

1967.[1] Causal Relations. *The Journal of Philosophy*, *64*, 691–703.

1967.[1] The Logical Form of Action Sentences. In N. Rescher (Ed.), *The Logic of Decision and Action* (pp. 81–95). Pittsburgh: University of Pittsburgh Press.

1967.[2] Truth and Meaning. *Synthese*, *17*, 304–23.

1968.[2] On Saying That. *Synthese*, *19*, 130–46.

1969. Facts and Events. In J. Margolis (Ed.), *Fact and Existence* (pp. 74–84). Oxford: Blackwell.

1969.[1] The Individuation of Events. In N. Rescher (Ed.), *Essays in Honor of Carl G. Hempel* (pp. 216–34). Dordrecht: D. Reidel.

1969.[2] True to the Facts. *The Journal of Philosophy*, *66*, 748–64.

1970. Action and Reaction. *Inquiry*, *13*, 140–8.

1970.[1] Events as Particulars. *Nous*, *4*, 25–32.

1970.[1] How Is Weakness of the Will Possible? In J. Feinberg (Ed.), *Moral Concepts* (pp. 93–113). Oxford: Oxford University Press.

1970.[1] Mental Events. In L. Foster and J. Swanson (Eds.), *Experience and Theory* (pp. 79–102). Amherst: University of Massachusetts Press.

1970.[2] Semantics for Natural Languages. In *Linguaggi nella Societa e nella Tecnica*. Milan: Comunita.

1971.[1] Agency. In R. Binkley, R. Bronaugh, and A. Marras (Eds.), *Agent, Action, and Reason* (pp. 3–37). Toronto: University of Toronto Press.

1971.[1] Eternal vs. Ephemeral Events. *Nous*, *5*, 335–49.

1973.[1] Freedom to Act. In T. Honderich (Ed.), *Essays On Freedom of Action* (pp. 137–56). London: Routledge and Kegan Paul.

1973.[2] In Defence of Convention T. In H. Leblanc (Ed.), *Truth, Syntax and Modality* (pp. 76–86). Dordretch: North-Holland.

1973.[1] The Material Mind. Paper presented at the Fourth International Congress for Logic, Methodology, and Philosophy of Science, Dordrecht, Holland.

1973.[2] Radical Interpretation. *Dialectica*, *27*, 314–28.

1974.[2] Belief and the Basis of Meaning. *Synthese*, *27*, 309–23.

1974.[2] On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association*, *47*, 5–20.

1974.[1] Psychology as Philosophy. In S. C. Brown (Ed.), *Philosophy of Psychology* (pp. 41–52). London: Macmillan.

1974.[2] Replies to David Lewis and W. V. Quine. *Synthese*, *27*, 345–9.

1975.[2] Thought and Talk. In S. Guttenplan (Ed.), *Mind and Language* (pp. 7–23). Oxford: Oxford University Press.

1976.[1] Hempel on Explaining Action. *Erkenntnis*, *10*, 239–53.

1976.[1] Hume's Cognitive Theory of Pride. *The Journal of Philosophy*, *73*, 744–56.

1976.[2] Reply to Foster. In G. Evans and J. McDowell (Eds.), *Truth and Meaning: Essays in Semantics* (pp. 33–41). Oxford: Oxford University Press.

1977.[2] The Method of Truth in Metaphysics. *Midwest Studies in Philosophy*, *2*, 244–54.

1977.[2] Reality without Reference. *Dialectica*, *31*, 247–53.

1978.[1] Intending. In Y. Yovel (Ed.), *Philosophy of History and Action* (pp. 41–60). Dordretch: D. Reidel and The Magnes Press.

1978.[2] What Metaphors Mean. *Critical Inquiry*, *5*, 31–47.

1979.[2] The Inscrutability of Reference. *The Southwest Journal of Philosophy*, *10*, 7–20.

1979.[2] Moods and Performances. In A. Margalit (Ed.), *Meaning and Use* (pp. 9–20). Dordrecht: D. Reidel.

1979.[2] Quotation. *Theory and Decision*, *11*, 27–40.

1980.[1] Comments and Replies. In *Essays on Actions and Events* (pp. 239–44). New York: Clarendon Press.

1980.[5] Toward a Unified Theory of Meaning and Action. *Grazer Philosophische Studien*, *11*, 1–12.

1982.[3] Empirical Content. *Grazer Philosophische Studien*, *17*, 471–89.

1982.[4] Paradoxes of Irrationality. In R. Wollheim et al. (Eds.), *Philosophical Essays on Freud* (pp. 289–305). Cambridge: Cambridge University Press.

1982.[3] Rational Animals. *Dialectica*, *36*, 317–28.

1983.[3] A Coherence Theory of Truth and Knowledge. In D. Henrich (Ed.), *Kant oder Hegel?* (pp. 423–38). Stuttgart: Klett-Cotta.

1983.[2] Communication and Convention. *The Journal of the Indian Council of Philosophical Research*, *1*, 13–25. Published in 1984 in *Synthese*, *59*, 3–18.

1984.[4] *Expressing Evaluations*. Lawrence, Kan.: Lindley.

1984.[3] First Person Authority. *Dialectica*, *38*, 101–12.

1985.[1] Adverbs of Action. In B. Vermazen (Ed.), *Essays on Davidson* (pp. 230–41). Oxford: Clarendon Press.

1985.[4] Deception and Division. In E. Lepore (Ed.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 138–48). Oxford: Blackwell. Also published in 1986 in J. Elster (Ed.), *The Multiple Self* (pp. 79–92). Cambridge: Cambridge University Press.

1985.[4] Incoherence and Irrationality. *Dialectica*, *39*, 345–54.

1985. A New Basis for Decision Theory. *Theory and Decision*, *18*, 87–98.

1985.[1] Reply to Quine on Events. In E. Lepore (Ed.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 172–6). Oxford: Blackwell.

1986.[4] Judging Interpersonal Interests. In J. Elster and A. Hylland (Eds.), *Foundations of Social Choice Theory* (pp. 195–211). Cambridge: Cambridge University Press.

1986.[5] A Nice Derangement of Epitaphs. In E. Lepore (Ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Cambridge: Blackwell. Also published in 1986 in R. E. Grandy (Ed.), *Philosophical Grounds of Rationality* (pp. 157–74). Oxford: Oxford University Press.

1987.[3] Afterthoughts, 1987. In A. Malichowski (Ed.), *Reading Rorty* (pp. 120–38). Cambridge: Blackwell.

1987.[3] Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*, *60*(3), 441–58.

1987.[4] Problems in the Explanation of Action. In P. Pettit (Ed.), *Metaphysics and Morality: Essays in Honour of J. J. C. Smart* (pp. 35–49). New York: Blackwell.

1988.[3] Epistemology and Truth. Proceedings of the Fourth Panamerican Philosophy Conference, Cordoba, Argentina.

1988. Reply to Burge. *The Journal of Philosophy*, *85*, 664–5.

1989. The Conditions of Thought. In J. Brandl and W. L. Gombocz (Eds.), *The Mind of Donald Davidson* (pp. 193–200). Amsterdam: Rodopi.

1989.[3] The Myth of the Subjective. In M. Krausz (Ed.), *Relativism: Interpretation and Confrontation* (pp. 159–72). Notre Dame, Ind.: University of Notre Dame Press.

1989.[3] What Is Present to the Mind? In J. Brandl and W. Gombocz (Eds.), *The Mind of Donald Davidson* (pp. 3–18). Amsterdam: Rodopi.

1990.[5] Meaning, Truth and Evidence. In R. B. Barret and R. F. Gibson (Eds.), *Perspectives on Quine* (pp. 68–79). Cambridge: Blackwell.

1990.[4] Representation and Interpretation. In W. H. Newton-Smith and K. V. Wilkes (Eds.), *Modelling the Mind* (pp. 12–26). Oxford: Oxford University Press.

1990. The Structure and Content of Truth. *The Journal of Philosophy*, *87*(6), 279–328.

1990.[4] Turing's Test. In W. H. Newton-Smith and K. V. Wilkes (Eds.), *Modelling the Mind* (pp. 1–11). Oxford: Oxford University Press.

1991.[3] Epistemology Externalized. *Dialectica*, *45*(2–3), 191–202.

1991.[5] James Joyce and Humpty Dumpty. *Midwest Studies in Philosophy*, *16*, 1–12.

1991.[3] Three Varieties of Knowledge. *Philosophy*, *30*(supp.), 153–66. Also published in 1991 in A. P. Griffiths (Ed.), *A. J. Ayer: Memorial Essays* (pp. 153–66). New York: Cambridge University Press.

1991.[3] What Is Present to the Mind. In E. Villanueva (Ed.), *Consciousness* (Vol. 6) (pp. 97–214). Atascadero, Calif.: Ridgeview.

1992.[3] The Second Person. *Midwest Studies in Philosophy*, *17*, 255–67.

1992.[5] The Socratic Conception of Truth. In K. J. Boudouris (Ed.), *The Philosophy of Socrates: Elenchus, Ethics and Truth* (pp. 51–8). Athens: International Center for Greek Philosophy and Culture.

1993.[5] Locating Literary Language. In R. W. Dasenbrock (Ed.), *Literary Theory after Davidson* (pp. 295–308). University Park: Pennsylvania State University Press.

1993.[5] Method and Metaphysics. *Deucalion*, *11*, 239–48.

1993.[5] Plato's Philosopher. In R. W. Sharples (Ed.), *Modern Thinkers and Ancient Thinkers* (pp. 1–15). Boulder, Colo.: Westview Press. Also published in 1993 in *Apeiron*, *26*(3–4), 179–94.

1993. Reply to Jerry Fodor and Ernest Lepore's 'Is Radical Interpretation Possible?' In R. Stoecker (Ed.), *Reflecting Davidson* (pp. 77–84). New York: de Gruyter.

1993. Thinking Causes. In J. Heil and A. Mele (Eds.), *Mental Causation* (pp. 3–17). New York: Clarendon Press.

1993.[5] The Third Man. *Critical Inquiry*, *19*(4), 607–16.

1994.[5] Dialectic and Dialogue. In G. Preyer (Ed.), *Language, Mind, and Epistemology: On Donald Davidson's Philosophy* (pp. 429–37). Dordrecht: Kluwer.

1994. Donald Davidson. In S. Guttenplan (Ed.), *A Companion to the Philosophy of Mind* (pp. 231–6). Oxford: Blackwell.

1994. Exchange between Donald Davidson and W. V. Quine following Davidson's Lecture. *Theoria*, *60*(3), 226–31.

1994. On Quine's Philosophy. *Theoria*, *60*(3), 184–92.

1994. Radical Interpretation Interpreted. In J. Tomberlin (Ed.), *Philosophical Perspectives: Logic and Language* (Vol. 8) (pp. 121–28). Atascadero, Calif.: Ridgeview.

1994.[5] The Social Aspect of Language. In B. McGuinness (Ed.), *The Philosophy of Michael Dummett* (pp. 1–16). Dordrecht: Kluwer.

1994.[5] What Is Quine's View of Truth? *Inquiry*, *37*(4), 437–40.

1995.[4] Could There Be a Science of Rationality? *International Journal of Philosophical Studies*, *3*(1), 1–16.

1995. Laws and Cause. *Dialectica*, *49*(2–4), 263–79.

1995.[4] The Objectivity of Values. In C. Gutiérrez (Ed.), *El Trabajo Filosófico de Hoy en el Continente* (pp. 59–69). Bogatá: Editorial ABC.

1995.[4] The Problem of Objectivity. *Tijdschrift voor Filosofie*, *57*(2), 203–20.

1995.[5] Pursuit of the Concept of Truth. In P. Leonardi and M. Santambrogio (Eds.), *On Quine: New Essays* (pp. 7–21). New York: Cambridge University Press.

1996.[5] The Folly of Trying to Define Truth. *The Journal of Philosophy*, *93*, 263–78.

1996.[3] Subjective, Intersubjective, Objective. In P. Coates and D. Ituito (Eds.), *Current Issues in Idealism* (pp. 155–76). Bristol: Thoemmes.

1997. The Centrality of Truth. In J. Peregrin (Ed.), *The Nature of Truth: Proceedings of the International Colloquium* (pp. 3–14). Prague: Filosofia.

1997.[5] Gadamer and Plato's *Philebus*. In L. Hahn (Ed.), *The Philosophy of Hans-Georg Gadamer* (pp. 421–32). Chicago: Open Court.

1997.[3] Indeterminism and Antirealism. In C. B. Kulp (Ed.), *Realism/Antirealism and Epistemology* (pp. 109–22). Lanham, Md.: Rowman and Littlefield.

1997.[5] Seeing through Language. In J. Preston (Ed.), *Thought and Language* (pp. 15–28). Cambridge: Cambridge University Press.

1997.[4] Who Is Fooled? In J.-P. Dupuy (Ed.), *Self-Deception and Paradoxes of Rationality* (pp. 1–18). Stanford: CSLI Publications.

1998.[3] The Irreducibility of the Concept of the Self. In M. Stamm (Ed.), *Philosophie in Synthetisher Absicht* (pp. 123–30). Stuttgart: Klett-Cotta.

1998.[5] Replies. *Crítica*, *30*, 97–112. Replies to articles by Barry Stroud, John McDowell, Richard Rorty, and Carlos Pereda in *Crítica*, *28* (1998).

1999.[3] The Emergence of Thought. *Erkenntnis*, *51*(1), 7–17.

1999. Intellectual Autobiography. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 3–70). Chicago: Open Court.

1999. Interpretation: Hard in Theory, Easy in Practice. In M. D. Caro (Ed.), *Interpretations and Causes: New Perspectives on Donald Davidson's Philosophy* (pp. 31–44). Dordrecht: Kluwer.

1999. Spinoza's Causal Theory of the Affects. In Y. Yovel (Ed.), *Desire and Affect: Spinoza as Psychologist* (pp. 95–112). New York: Little Room Press.

1999. Is Truth a Goal of Inquiry? Discussion with Rorty. In U. M. Zeglen (Ed.), *Donald Davidson: Truth, Meaning and Knowledge* (pp. 17–19). London: Routledge.

2000.[4] Objectivity and Practical Reason. In E. Ullmann-Margalit (Ed.), *Reasoning Practically* (pp. 17–26). Oxford: Oxford University Press.

2000. Perils and Pleasures of Interpretation. *Ars Interpretandi*, *5*, 21–37.

2000.[5] Truth Rehabilitated. In R. B. Brandom (Ed.), *Rorty and His Critics* (pp. 65–74). Cambridge: Blackwell.

2001. Externalisms. In P. Kotatko, P. Pagin, and G. Segal (Eds.), *Interpreting Davidson* (pp. 1–16). Stanford: CSLI Publications.

III. Books

1957. *Decision Making: An Experimental Approach*. Stanford, Calif.: Stanford University Press. (With Patrick Suppes.)

1990. *Plato's Philebus*. New York: Garland. (1949 Harvard dissertation.)

IV. Edited Books

1969. *Words and Objections: Essays on the Work of W. V. Quine*. Dordrecht: Reidel. (With Jaakko Hintikka.) Rev. ed. 1975. (With Jaakko Hintikka and W. V. Quine.)

1970. *Essays in Honor of Carl G. Hempel.* Dordrecht: D. Reidel. (With Carl Hempel and Nicholas Rescher.)

1972. *Semantics of Natural Language*. Boston: D. Reidel. (With Gilbert Harman.) A second edition was published in 1977 with additional essays by Saul Kripke and others.

1975. *The Logic of Grammar*. Encino, Calif.: Dickenson. (With Gilbert Harman.)

V. Videotaped Conversations

1997. *In Conversation: Donald Davidson* (19 videocassettes, 1460 min.). Introduced and edited by R. Fara. London: Philosophy International Centre for Philosophy of the Natural and Social Sciences, London School of Economics.

Selected Commentary on Davidson

Listed below are journal issues (section I), collections of essays (section II), and books or collections of essays (section III) on aspects of Donald Davidson's philosophical work. Those collections in section II that contain replies by Davidson are marked with an asterisk. Entries are restricted to English-language publications.

I. Journal Issues

Mind & Language, *11* (1996). Forum on Swampman. Edited by Samuel Guttenplan and Sarah Patterson.

Critica, *88* (1998). Papers by Barry Stroud, John McDowell, Richard Rorty, Carlos Pereda, and Akeel Bilgrami presented at the Sixteenth International Symposium of Philosophy on the Philosophy of Donald Davidson.

II. Collections of Essays

Brandl, J., and Gombocz, W. L. (Eds.) (1989) *The Mind of Donald Davidson*. Amsterdam: Rodopi.

Dasenbrock, R. W. (Ed.) (1993) *Literary Theory after Davidson*. University Park: Pennsylvania State University Press.

De Caro, M. (Ed.) (1999) *Interpretations and Causes: New Perspectives on Donald Davidson's Philosophy*. Dordrecht and Boston: Kluwer.

Evans, G., and McDowell, J. (Eds.) (1976) *Truth and Meaning: Essays in Semantics*. Oxford: Clarendon Press.

*Hahn, L. E. (Ed.) (1999) *The Philosophy of Donald Davidson*. Chicago: Open Court.

*Kotatko, P., Pagin, P., and Segal, G. (Eds.) (2001) *Interpreting Davidson*. Stanford: CSLI Publications.

Lepore, E. (Ed.) (1986) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell.

Lepore, E., and McLaughlin, B. (Eds.) (1985) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell.

Platts, M. (Ed.) (1980) *Reference, Truth, and Reality: Essays on the Philosophy of Language*. London: Routledge and Kegan Paul.

Preyer, G. (Ed.) (1994) *Language, Mind, and Epistemology: On Donald Davidson's Philosophy*. Dordrecht: Kluwer.

*Stoecker, R. (Ed.) (1993) *Reflecting Davidson*. Berlin: W. de Gruyter.

*Vermazen, B., and Hintikka, M. B. (Eds.) (1985) *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press.

*Zeglen, U. M. (Ed.) (1999) *Donald Davidson: Truth, Meaning, and Knowledge*. New York: Routledge.

III. Books

Bayer, S. L. (1997) *Confessions of a Lapsed Neo-Davidsonian: Events and Arguments in Compositional Semantics*. New York: Garland.

Davies, M. (1981) *Meaning, Quantification, Necessity*. London: Routledge and Kegan Paul. An "introductory" synthesis of work on philosophical logic influenced by the Davidsonian program in semantics.

Evnine, S. (1991) *Donald Davidson*. Stanford, Calif.: Stanford University Press.

Larson, R., and Segal, G. (1995) *Knowledge of Meaning*. Cambridge, Mass.: MIT Press. This combines Davidson's truth-theoretic approach to semantics with syntactical theory as developed in linguistics.

Lepore, E., and Ludwig, K. Forthcoming. *Donald Davidson: Truth, Meaning, Language and Reality*. Oxford: Oxford University Press.

Letson, B. H. (1997) *Davidson's Theory of Truth and Its Implications for Rorty's Pragmatism*. New York: Peter Lang.

Malpas, J. E. (1992) *Donald Davidson and the Mirror of Meaning: Holism, Truth, Interpretation*. Cambridge: Cambridge University Press.

Platts, M. (1997) *Ways of Meaning: An Introduction to Philosophy of Language*, 2nd ed. Cambridge, Mass.: MIT Press. This is an introduction to the philosophy of language through Davidson's approach to the semantics of natural languages.

Ramberg, B. T. (1989) *Donald Davidson's Philosophy of Language: An Introduction*. Cambridge: Blackwell.

Wheeler, S. C. (2000) *Deconstruction as Analytic Philosophy*. Stanford, Calif.: Stanford University Press.

Bibliographic References

Anscombe, G. E. M. (1957) *Intention*. Oxford: Blackwell.

Antony, L. (1989) Anomalous Monism and the Problem of Explanatory Force. *The Philosophical Review*, *98*, 153–87.

Antony, L. M. (1994) The Inadequacy of Anomalous Monism as a Realist Theory of Mind. In G. Preyer (Ed.), *Language, Mind, and Epistemology: On Donald Davidson's Philosophy* (pp. 223–54). Dordrecht: Kluwer.

Armstrong, D. M., and Malcolm, N. (1984) *Consciousness and Causality*: *A Debate on the Nature of Mind*. New York: Blackwell.

Audi, R. (1993) *Action*, *Intention*, *and Reason*. Ithaca, N.Y.: Cornell University Press.

Barthes, R. (1967) *Elements of Semiology*. London: Cape.

Barwise, J. (1981) Scenes and Other Situations. *The Journal of Philosophy*, *78*, 369–97.

Barwise, J., and Perry, J. (1981a) Semantic Innocence and Uncompromising Situations. *Midwest Studies in Philosophy*, *6*, 387–403.

Barwise, J., and Perry, J. (1981b) Situations and Attitudes. *The Journal of Philosophy*, *78*, 668–90.

Barwise, J., and Perry, J. (1983) *Situations and Attitudes*. Cambridge, Mass.: MIT Press.

Beaney, M. (Ed.) (1997) *The Frege Reader*. Oxford: Blackwell.

Bergstrom, L. (1994) Interview with Donald Davidson. *Theoria*, *60*, 207–25.

Bishop, J. (1989) *Natural Agency*. Cambridge: Cambridge University Press.

Block, N., and Stalnaker, R. (1999) Conceptual Analysis, Dualism, and the Explanatory Gap. *Philosophical Review*, *108*(1), 1–46.

Bolker, E. (1967) A Simultaneous Axiomatization of Utility and Subjective Probability. *Philosophy of Science*, *34*, 330–40.

Boolos, G. (1984) To Be Is to Be a Value of a Variable (or to Be Some Values of Some Variables). *The Journal of Philosophy*, *81*, 430–49.

Boolos, G. (1985) Nominalist Platonism. *Philosophical Review*, *94*, 327–44.

Borges, J. (1965) Pierre Menard: Author of the Quixote. In *Fictions* (pp. 42–51). London: John Calder.

Brand, M. (1984) *Intending and Acting*. Cambridge, Mass.: MIT Press.

Bratman, M. (1985) Davidson's Theory of Intention. In E. Lepore (Ed.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 14–28). Oxford: Blackwell.

Bratman, M. (1987) *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.

Bratman, M. (1999) *Faces of Intention: Selected Essays on Intention and Agency*. New York: Cambridge University Press.

Brooks, C. (1947) What Does Poetry Communicate? In *The Well Wrought Urn: Studies in the Structure of Poetry*. New York: Reynal and Hitchcock.

Burge, T. (1993) Mind-Body Causation and Explanatory Practice. In J. Heil and A. R. Mele (Eds.), *Mental Causation*. New York: Clarendon.

Burge, T. (1999) Comprehension and Interpretation. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 229–50). Chicago: Open Court.

Cappelen, H., and Lepore, E. (1997) Varieties of Quotation. *Mind*, *106*(423), 429–50.

Cappelen, H., and Lepore, E. (1999a) Semantics for Quotation. In K. Murasugi (Ed.), *Philosophy and Linguistics* (pp. 209–21). Boulder, Colo.: Westview Press.

Cappelen, H., and Lepore, E. (1999b) Replies to the Commentaries. In K. Murasugi (Ed.), *Philosophy and Linguistics* (pp. 279–85). Boulder, Colo.: Westview Press.

Carlson, G. (1984) Thematic Roles and Their Role in Semantic Interpretation. *Linguistics*, *22*, 259–79.

Castañeda, H. (1967) Comments. In N. Rescher (Ed.), *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press.

Chihara, C. S. (1976) Truth, Meaning, and Paradox. Nous, 10, 305–12.

Child, W. (1993) Anomalism, Uncodifiability, and Psychophysical Relations. *Philosophical Review*, *102*, 215–45.

Child, W. (1994) *Causality, Interpretation, and the Mind*. New York: Clarendon Press.

Chomsky, N. (1995) Language and Nature. *Mind*, *104*, 1–61.

Church, A. (1943) Carnap's Introduction to Semantics. *Philosophical Review*, *52*, 298–305.

Church, A. (1951) The Need for Abstract Entities in Semantic Analysis. *Proceedings of the American Academy of Arts and Letters*, *80*, 100–12.

Church, A. (1956) *Introduction to Mathematical Logic*. Princeton, N.J.: Princeton University Press.

Costa, M. (1987) Causal Theories of Action. *Canadian Journal of Philosophy*, *17*, 831–54.

Dasenbrock, R. W. (1993) *Literary Theory after Davidson*. University Park: Pennsylvania State University Press.

Davidson, D. (1963) The Method of Intension and Extension. In A. Schilpp (Ed.), *The Philosophy of Rudolf Carnap* (pp. 311–50). La Salle, Ill.: Open Court.

Davidson, D. (1970) Action and Reaction. *Inquiry*, *13*, 140–8.

Davidson, D. (1980a) *Essays on Actions and Events*. New York: Clarendon Press.

Davidson, D. (1980b) Toward a Unified Theory of Meaning and Action. *Grazer Philosophische Studien*, *11*, 1–12.

Davidson, D. (1980 [1963]) Actions, Reasons, and Causes. In *Essays on Actions and Events* (pp. 3–20). New York: Clarendon Press. Original publication (1963): *The Journal of Philosophy*, *60*, 685–99.

Davidson, D. (1980a [1967]) Causal Relations. In *Essays on Actions and Events* (pp. 149–62). New York: Clarendon Press. Original publication (1967): *The Journal of Philosophy*, *64*, 691–703.

Davidson, D. (1980b [1967]) The Logical Form of Action Sentences. In *Essays on Actions and Events* (pp. 105–21). New York: Clarendon Press. Original publication (1967): In N. Rescher (Ed.), *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press.

Davidson, D. (1980 [1969]) The Individuation of Events. In *Essays on Actions and Events* (pp. 163–80). New York: Clarendon Press. Original publication (1969): In N. Rescher (Ed.), *Essays in Honor of Carl G. Hempel* (pp. 216–34). Dordrecht: D. Reidel.

Davidson, D. (1980a [1970]) Events as Particulars. In *Essays on Actions and Events* (pp. 181–8). New York: Clarendon Press. Original publication (1970): *Nous*, *4*, 25–32.

Davidson, D. (1980b [1970]) How Is Weakness of the Will Possible? In *Essays on Actions and Events* (pp. 21–42). New York: Clarendon Press. Original publication (1970): In J. Feinberg (Ed.), *Moral Concepts*. Oxford: Oxford University Press.

Davidson, D. (1980c [1970]) Mental Events. In *Essays on Actions and Events* (pp. 207–24). New York: Clarendon Press. Original publication (1970): In L. Foster and J. Swanson (Eds.), *Experience and Theory*. Amherst: University of Massachusetts Press.

Davidson, D. (1980a [1971]) Agency. In *Essays on Actions and Events* (pp. 43–62). New York: Clarendon Press. Original publication (1971): In R. Binkley, R. Bronaugh, and A. Marras (Eds.), *Agent, Action, and Reason* (pp. 3–37). Toronto: University of Toronto Press.

Davidson, D. (1980b [1971]) Eternal vs. Ephemeral Events. In *Essays on Actions and Events* (pp. 189–204). New York: Clarendon Press. Original publication (1971): *Nous*, *5*, 335–349.

Davidson, D. (1980a [1973]) Freedom to Act. In *Essays on Actions and Events* (pp. 63–82). New York: Clarendon Press. Original publication (1973): In T. Honderich (Ed.), *Essays On Freedom of Action*. London: Routledge and Kegan Paul.

Davidson, D. (1980b [1973]) The Material Mind. In *Essays on Actions and Events* (pp. 245–60). New York: Clarendon Press. Original publication (1973): In P. Suppes, L. Henkin, and G. C. Moisil (Eds.), *Proceedings of the Fourth International Congress for Logic, Methodology, and Philosophy of Science*. Dordrecht: D. Reidel.

Davidson, D. (1980 [1974]) Psychology as Philosophy. In *Essays on Actions and Events* (pp. 229–38). New York: Clarendon Press. Original publication (1974): In S. C. Brown (Ed.), *Philosophy of Psychology*. London: Macmillan.

Davidson, D. (1980a [1976]) Hempel on Explaining Action. In *Essays on Actions and Events* (pp. 261–76). New York: Clarendon Press. Original publication (1976): *Erkenntnis*, *10*, 239–53.

Davidson, D. (1980b [1976]) Hume's Cognitive Theory of Pride. In *Essays on Actions*

and Events (pp. 277–90). Oxford: Oxford University Press. Original publication (1976): *The Journal of Philosophy*, *73*, 744–56.

Davidson, D. (1980 [1978]) Intending. In *Essays on Actions and Events* (pp. 83–102). New York: Clarendon Press. Original publication (1978): In Y. Yovel (Ed.), *Philosophy of History and Action*. Dordretch: D. Reidel and The Magnes Press.

Davidson, D. (1982) Paradoxes of Irrationality. In R. Wollheim et al. (Eds.), *Philosophical Essays on Freud* (pp. 289–305). Cambridge: Cambridge University Press.

Davidson, D. (1984a) *Expressing Evaluations*. Lawrence, Kan.: Lindley.

Davidson, D. (1984b) *Inquiries into Truth and Interpretation*. New York: Clarendon Press.

Davidson, D. (1984 [1966]) Theories of Meaning and Learnable Languages. In *Inquiries into Truth and Interpretation* (pp. 3–16). New York: Clarendon Press. Original publication (1966): In Y. Bar-Hillel (Ed.), *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science*. Amsterdam: North Holland.

Davidson, D. (1984 [1967]) Truth and Meaning. In *Inquiries into Truth and Interpretation* (pp. 17–36). New York: Clarendon Press. Original publication (1967): *Synthese*, *17*, 304–23.

Davidson, D. (1984 [1968]) On Saying That. In *Inquiries into Truth and Interpretation* (pp. 93–108). New York: Clarendon Press. Original publication (1968): *Synthese*, *19*, 130–46.

Davidson, D. (1984 [1969]) True to the Facts. In *Inquiries into Truth and Interpretation* (pp. 37–54). New York: Clarendon Press. Original publication (1969): *The Journal of Philosophy*, *66*, 748–64.

Davidson, D. (1984 [1970]) Semantics for Natural Languages. In *Inquiries into Truth and Interpretation* (pp. 37–54). New York: Clarendon Press. Original publication (1970): In *Linguaggi nella Societa e nella Tecnica*. Milan: Comunita.

Davidson, D. (1984a [1973]) In Defense of Convention T. In *Inquiries into Truth and Interpretation* (pp. 65–76). New York: Clarendon Press. Original publication (1973): In H. Leblanc (Ed.), *Truth, Syntax and Modality*. Dordretch: North-Holland.

Davidson, D. (1984b [1973]) Radical Interpretation. In *Inquiries into Truth and Interpretation* (pp. 125–40). New York: Clarendon Press. Original publication (1973): *Dialectica*, *27*, 314–28.

Davidson, D. (1984a [1974]) Belief and the Basis of Meaning. In *Inquiries into Truth and Interpretation* (pp. 141–54). New York: Clarendon Press. Original publication (1974): *Synthese*, *27*, 309–23.

Davidson, D. (1984b [1974]) On the Very Idea of a Conceptual Scheme. In *Inquiries into Truth and Interpretation* (pp. 183–98). New York: Clarendon Press. Original publication (1974): *Proceedings and Addresses of the American Philosophical Association*, *47*, 5–20.

Davidson, D. (1984 [1975]) Thought and Talk. In *Inquiries into Truth and Interpretation* (pp. 155–70). New York: Clarendon Press. Original publication (1975): In S. Guttenplan (Ed.), *Mind and Language*. Oxford: Oxford University Press.

Davidson, D. (1984 [1976]) Reply to Foster. In *Inquiries into Truth and Interpretation* (pp. 171–80). New York: Clarendon Press. Original publication (1976): In G. Evans, and J. McDowell (Eds.), *Truth and Meaning: Essays on Semantics*. Oxford: Oxford University Press.

Davidson, D. (1984a [1977]) The Method of Truth in Metaphysics. In *Inquiries into Truth and Interpretation* (pp. 199–214). New York: Clarendon Press. Original publication (1977): *Midwest Studies in Philosophy*, *2*, 244–54.

Davidson, D. (1984b [1977]) Reality without Reference. In *Inquiries into Truth and Interpretation* (pp. 215–26). New York: Clarendon Press. Original publication (1977): *Dialectica*, *31*, 247–53.

Davidson, D. (1984 [1978]) What Metaphors Mean. In *Inquiries into Truth and Interpretation* (pp. 245–64). New York: Clarendon Press. Original publication (1978): *Critical Inquiry*, *5*, 31–47.

Davidson, D. (1984a [1979]) The Inscrutability of Reference. In *Inquiries into Truth and Interpretation* (pp. 227–42). New York: Clarendon Press. Original publication (1979): *The Southwest Journal of Philosophy*, *10*, 7–20.

Davidson, D. (1984b [1979]) Moods and Performances. In *Inquiries into Truth and Interpretation* (pp. 109–22). New York: Clarendon Press. Original publication (1979): In A. Margalit (Ed.), *Meaning and Use*. Dordrecht: D. Reidel.

Davidson, D. (1984c [1979]) Quotation. In *Inquiries into Truth and Interpretation* (pp. 79–92). New York: Clarendon Press. Original publication (1979): *Theory and Decision*, *11*, 27–40.

Davidson, D. (1984 [1983]) Communication and Convention. In *Inquiries into Truth and Interpretation* (pp. 265–80). New York: Clarendon Press. Original publication (1983): *The Journal of the Indian Council of Philosophical Research*, *1*, 13–25.

Davidson, D. (1985a) Adverbs of Action. In B. Vermazen and M. Hintikka (Eds.), *Essays on Davidson: Actions and Events* (pp. 230–41). Oxford: Clarendon Press.

Davidson, D. (1985b) Deception and Division. In E. Lepore (Ed.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 138–48). Oxford: Blackwell.

Davidson, D. (1985c) Incoherence and Irrationality. *Dialectica*, *39*, 345–54.

Davidson, D. (1985d) A New Basis for Decision Theory. *Theory and Decision*, *18*, 87–98.

Davidson, D. (1985 [1982]) Rational Animals. In E. Lepore (Ed.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 473–81). Oxford: Blackwell. Original publication (1982): *Dialectica*, *36*, 317–28.

Davidson, D. (1985e) Replies to Essays I-IX. In B. Vermazen and M. Hintikka (Eds.), *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press.

Davidson, D. (1985f) Reply to Quine on Events. In E. Lepore (Ed.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 172–6). Oxford: Blackwell.

Davidson, D. (1986a) A Coherence Theory of Truth and Knowledge. In E. Lepore (Ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp. 307–19). Cambridge: Blackwell.

Davidson, D. (1986b) Judging Interpersonal Interests. In J. Elster and A. Hylland (Eds.), *Foundations of Social Choice Theory* (pp. 195–211). Cambridge: Cambridge University Press.

Davidson, D. (1986c) A Nice Derangement of Epitaphs. In E. Lepore (Ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp. 433–46). Cambridge: Blackwell.

Davidson, D. (1986d) A Nice Derangement of Epitaphs. In R. E. Grandy (Ed.), *Philosophical Grounds of Rationality* (pp. 157–74). Oxford: Oxford University Press.

Davidson, D. (1987a) Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*, *60*(3), 441–58.

Davidson, D. (1987b) Problems in the Explanation of Action. In P. Pettit (Ed.), *Metaphysics and Morality: Essays in Honour of J. J. C. Smart* (pp. 35–49). New York: Blackwell.

Davidson, D. (1989) The Conditions of Thought. In J. Brandl and W. L. Gombocz (Eds.), *The Mind of Donald Davidson* (pp. 193–200). Amsterdam: Rodopi.

Davidson, D. (1990a) Meaning, Truth and Evidence. In R. B. Barret and R. F. Gibson (Eds.), *Perspectives on Quine* (pp. 68–79). Cambridge: Blackwell.

Davidson, D. (1990b) *Plato's Philebus*. New York: Garland.

Davidson, D. (1990c) Representation and Interpretation. In W. H. Newton-Smith and K. V. Wilkes (Eds.), *Modelling the Mind* (pp. 12–26). Oxford: Oxford University Press.

Davidson, D. (1990d) The Structure and Content of Truth. *The Journal of Philosophy*, *87*(6), 279–328.

Davidson, D. (1990e) Turing's Test. In W. H. Newton-Smith and K. V. Wilkes (Eds.), *Modelling the Mind* (pp. 1–11). Oxford: Oxford University Press.

Davidson, D. (1991a) Epistemology Externalized. *Dialectica*, *45*(2–3), 191–202.

Davidson, D. (1991b) James Joyce and Humpty Dumpty. *Midwest Studies in Philosophy*, *16*, 1–12.

Davidson, D. (1992a) The Second Person. *Midwest Studies in Philosophy*, *17*, 255–67.

Davidson, D. (1992b) The Socratic Conception of Truth. In K. J. Boudouris (Ed.), *The Philosophy of Socrates: Elenchus, Ethics and Truth* (pp. 51–8). Athens: International Center for Greek Philosophy and Culture.

Davidson, D. (1993a) Locating Literary Language. In R. W. Dasenbrock (Ed.), *Literary Theory after Davidson* (pp. 295–308). University Park: Pennsylvania State University Press.

Davidson, D. (1993b) Thinking Causes. In J. Heil and A. Mele (Eds.), *Mental Causation* (pp. 3–17). New York: Clarendon Press.

Davidson, D. (1994a) On Quine's Philosophy. *Theoria*, *60*(3), 184–92.

Davidson, D. (1994b) Radical Interpretation Interpreted. In J. Tomberlin (Ed.), *Philosophical Perspectives: Logic and Language* (Vol. 8) (pp. 122–28). Atascadero, Calif.: Ridgeview.

Davidson, D. (1994c) The Social Aspect of Language. In B. McGuinness (Ed.), *The Philosophy of Michael Dummett* (pp. 1–16). Dordrecht: Kluwer.

Davidson, D. (1994d) What Is Quine's View of Truth? *Inquiry*, *37*(4), 437–40.

Davidson, D. (1995a) Could There Be a Science of Rationality? *International Journal of Philosophical Studies*, *3*(1), 1–16.

Davidson, D. (1995b) Laws and Cause. *Dialectica*, *49*(2–4), 263–79.

Davidson, D. (1995c) The Objectivity of Values. In C. Gutiérrez (Ed.), *El Trabajo Filosófico de Hoy en el Continente* (pp. 59–69). Bogatá: Editorial ABC.

Davidson, D. (1995d) The Problem of Objectivity. *Tijdschrift voor Filosofie*, *57*(2), 203–20.

Davidson, D. (1995e) Pursuit of the Concept of Truth. In P. Leonardi and M. Santambrogio (Eds.), *On Quine: New Essays*. New York: Cambridge University Press.

Davidson, D. (1996a) The Folly of Trying to Define Truth. *The Journal of Philosophy*, *93*, 263–78.

Davidson, D. (1996b) Subjective, Intersubjective, Objective. In P. Coates (Ed.), *Current Issues in Idealism*. Bristol: Thoemmes.

Davidson, D. (1997a) The Centrality of Truth. In J. Peregrin (Ed.), *The Nature of Truth: Proceedings of the International Colloquium* (pp. 3–14). Prague: Filosofia.

Davidson, D. (1997b) Seeing through Language. In J. Preston (Ed.), *Thought and Language* (pp. 15–28). Cambridge: Cambridge University Press.

Davidson, D. (1997c) Who Is Fooled? In J.-P. Dupuy (Ed.), *Self-Deception and Paradoxes of Rationality* (pp. 1–18). Stanford, Calif.: CSLI Publications.

Davidson, D. (1999a) Intellectual Autobiography. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 3–70). Chicago: Open Court.

Davidson, D. (1999b) Interpretation: Hard in Theory, Easy in Practice. In M. D. Caro (Ed.), *Interpretations and Causes: New Perspectives on Donald Davidson's Philosophy* (pp. 31–44). Dordrecht: Kluwer.

Davidson, D. (1999c) Reply to A. C. Genova. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 192–4). Chicago: Open Court.

Davidson, D. (1999d) Reply to Barry Stroud. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 162–6). Chicago: Open Court.

Davidson, D. (1999e) Reply to Edna Ullmann-Margalit. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 497–500). Chicago: Open Court.

Davidson, D. (1999f) Reply to James Higginbotham. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 687–9). Chicago: Open Court.

Davidson, D. (1999g) Reply to Jennifer Hornsby. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 636–40). Chicago: Open Court.

Davidson, D. (1999h) Reply to John McDowell. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 105–8). Chicago: Open Court.

Davidson, D. (1999i) Reply to Thomas Nagel. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 207–9). Chicago: Open Court.

Davidson, D. (1999j) Spinoza's Causal Theory of the Affects. In Y. Yovel (Ed.), *Desire and Affect: Spinoza as Psychologist (Papers Presented at the Third Jerusalem Conference)*. New York: Little Room Press.

Davidson, D. (2000a) Objectivity and Practical Reason. In E. Ullmann-Margalit (Ed.), *Reasoning Practically* (pp. 17–26). Oxford: Oxford University Press.

Davidson, D. (2000b) Perils and Pleasures of Interpretation. *Ars Interpretandi*, *5*, 21–37.

Davidson, D. (2000c) Truth Rehabilitated. In R. B. Brandom (Ed.), *Rorty and His Critics* (pp. 65–74). Cambridge: Blackwell.

Davidson, D. (2001a) Externalisms. In P. Kotatko, P. Pagin, G. Segal (Eds.), *Interpreting Davidson* (pp. 1–16). Stanford: CSLI Publications.

Davidson, D. (2001b) *Subjective*, *Intersubjective*, *Objective*. New York: Clarendon Press.

Davidson, D. (2001a [1982]) Empirical Content. In *Subjective*, *Intersubjective*, *Objective* (pp. 159–76). New York: Clarendon Press. Original publication (1982): *Grazer Philosophische Studien*, *17*, 471–89.

Davidson, D. (2001b [1982]) Rational Animals. In *Subjective*, *Intersubjective*, *Objective* (pp. 95–106). New York: Clarendon Press. Original publication (1982): *Dialectica*, *36*, 317–28.

Davidson, D. (2001 [1983]) A Coherence Theory of Truth and Knowledge. In *Subjective*, *Intersubjective*, *Objective* (pp. 137–53). New York: Clarendon Press. Original publication (1983): In D. Henrich (Ed.), *Kant oder Hegel?* Stuttgart: Klett-Cotta.

Davidson, D. (2001 [1984]) First Person Authority. In *Subjective*, *Intersubjective*, *Objective* (pp. 3–14). New York: Clarendon Press. Original publication (1984): *Dialectica*, *38*, 101–112.

Davidson, D. (2001 [1987]) Knowing One's Own Mind. In *Subjective*, *Intersubjective*, *Objective* (pp. 15–38). New York: Clarendon Press. Original publication (1987): *Proceedings and Addresses of the American Philosophical Association*, *60*(3), 441–58.

Davidson, D. (2001a [1988]) Epistemology and Truth. In *Subjective*, *Intersubjective*, *Objective* (pp. 177–92). New York: Clarendon Press. Original publication (1988): Proceedings of the Fourth Panamerican Philosophy Conference, Cordoba, Argentina.

Davidson, D. (2001b [1988]) The Myth of the Subjective. In *Subjective*, *Intersubjective*, *Objective* (pp. 39–52). New York: Clarendon Press. Original publication (1988): In M. Benedikt and R. Berger (Eds.), *Bewusstsein*, *Sprache und die Kunst*. Vienna: Edition S. Verlag der Österreichischen Staatscruckerei.

Davidson, D. (2001 [1989]) What Is Present to the Mind? In *Subjective*, *Intersubjective*, *Objective* (pp. 53–68). New York: Clarendon Press. Original publication (1989): In J. Brandl and W. Gombocz (Eds.), *The Mind of Donald Davidson*. Amsterdam: Rodopi.

Davidson, D. (2001a [1991]) Epistemology Externalized. In *Subjective*, *Intersubjective*, *Objective* (pp. 193–204). New York: Clarendon Press. Original publication (1991): *Dialectica*, *45*(2–3), 191–202.

Davidson, D. (2001b [1991]) Three Varieties of Knowledge. In *Subjective*, *Intersubjective*, *Objective* (pp. 205–20). New York: Clarendon Press. Original publication (1991): In A. P. Griffiths (Ed.), *A. J. Ayer: Memorial Essays*. New York: Cambridge University Press.

Davidson, D. (2001 [1992]) The Second Person. In *Subjective, Intersubjective, Objective* (pp. 107–22). New York: Clarendon Press. Original publication (1992): *Midwest Studies in Philosophy, 17*, 255–67.

Davidson, D. (2001 [1999]) The Emergence of Thought. In *Subjective, Intersubjective, Objective* (pp. 123–34). New York: Clarendon Press. Original publication (1999): *Erkenntnis, 51*(1), 7–17.

Davidson, D., and Harman, G. (1975) *The Logic of Grammar*. Encino, Calif.: Dickenson.

Davidson, D., and Harman, G. (1972) *Semantics of Natural Language*. Boston: D. Reidel.

Davidson, D., and Harman, G. (Eds.) (1977) *Semantics of Natural Language*, 2nd ed. Dordrecht: D. Reidel.

Davidson, D., and Suppes, P. (1957) *Decision Making: An Experimental Approach*. Stanford, Calif.: Stanford University Press.

Davies, M. (1981) *Meaning, Quantification, Necessity*. London: Routledge and Kegan Paul.

De Finetti, B. (1972) *Probability, Induction and Statistics: The Art of Guessing*. New York: Wiley.

Derrida, J. (1977) Signature, Event, Context. *Glyph, 1*, 172–97.

Descartes, R. (1955 [1643]) Meditations on First Philosophy (E. S. Haldane and G. R. T. Ross, trans.). In E. S. Haldane and G. R. T. Ross (Eds.), *The Philosophical Works of Descartes* (Vol. 1, pp. 131–99). New York: Dover.

Dretske, F. (1988) *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Mass.: MIT Press.

Evans, G. (1985) Does Tense Logic Rest on a Mistake? In *Collected Papers* (pp. 343–63). Oxford: Clarendon Press.

Evans, G., and McDowell, J. (Eds.) (1976) *Truth and Meaning: Essays in Semantics*. Oxford: Clarendon Press.

Feinberg, J. (1965) Action and Responsibility. In M. Black (Ed.), *Philosophy in America* (pp. 134–60). Ithaca, N.Y.: Cornell University Press.

Fodor, J. A. (1970) Three Reasons for Not Deriving 'Kill' from 'Cause to Die'. *Linguistic Inquiry, 1*, 429–38.

Fodor, J. A. (1974) Special Sciences, or the Disunity of Science as a Working Hypothesis. *Synthese, 28*, 97–115.

Foley, R., and Fumerton, R. (1985) Davidson's Theism? *Philosophical Studies, 48*, 83–90.

Foster, J. A. (1976) Meaning and Truth Theory. In G. Evans and J. McDowell (Eds.), *Truth and Meaning: Essays in Semantics* (pp. 1–32). Oxford: Clarendon Press.

Francken, P., and Lombard, L. (1992) How Not to Flip the Switch with the Floodlight. *Pacific Philosophical Quarterly, 73*, 31–43.

Frege, G. (1892) Ueber Sinn und Bedeutung. *Zeitschrift fur Philosophie und Philosophische Kritik, 100*, 25–50.

Frege, G. (1960 [1891]) Function and Concept. In P. Geach and M. Black (Eds.),

Translations from the Philosophical Writings of Gottlob Frege (pp. 21–41). Oxford: Oxford University Press.

Frege, G. (1997 [1891]) Function and Concept (M. Beaney, trans.). In M. Beaney (Ed.), *The Frege Reader* (pp. 130–48). Oxford: Blackwell.

Frege, G. (1997a [1892]) On Concept and Object (M. Beaney, trans.). In M. Beaney (Ed.), *The Frege Reader* (pp. 181–93). Oxford: Blackwell.

Frege, G. (1997b [1892]) On Sinn and Bedeutung. In M. Beaney (Ed.), *The Frege Reader*. Oxford: Blackwell.

Genova, A. C. (1999) The Very Idea of Massive Truth. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 167–91). Chicago: Open Court.

Ginet, C. (1990) *On Action*. Cambridge: Cambridge University Press.

Gödel, K. (1966) Russell's Mathematical Logic. In H. Putnam and P. Benacerraf (Eds.), *Philosophy of Mathematics*. Englewood Cliffs, N.J.: Prentice Hall.

Goldman, A. I. (1970) *A Theory of Human Action*. Englewood Cliffs, N.J.: Prentice-Hall.

Hahn, L. E. (Ed.) (1999) *The Philosophy of Donald Davidson* (Library of Living Philosophers, Vol. 27). Chicago: Open Court.

Higginbotham, J. (1983) The Logic of Perceptual Reports: An Extensional Alternative to Semantics. *The Journal of Philosophy*, *80*, 100–27.

Higginbotham, J. (1985) On Semantics. *Linguistic Inquiry*, *16*, 547–93.

Higginbotham, J. (1986) Linguistic Theory and Davidson's Program in Semantics. In E. Lepore (Ed.), *Essays on Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp. 29–48). New York: Blackwell.

Higginbotham, J. (1995) Tensed Thoughts. *Mind and Language*, *10*(3), 226–49.

Hirsch, E. D. (1967) *Validity in Interpretation*. New Haven, Conn.: Yale University Press.

Hochberg, H. (1975) Explaining Facts. *Metaphilosophy*, *6*, 277–302.

Honderich, T. (1981) Psychophysical Lawlike Connections and Their Problem. *Inquiry*, *24*, 277–303.

Honderich, T. (1982) The Argument for Anomalous Monism. *Analysis*, *42*, 59–64.

Honderich, T. (1988) *A Theory of Determinism: The Mind, Neuroscience, and Life-hopes*. Oxford: Clarendon Press.

Horgan, T. (1989) Mental Quausation. *Philosophical Perspectives*, *3*, 47–76.

Hornsby, J. (1980) *Actions*. London: Routledge and Kegan Paul.

Hornsby, J. (1985) Physicalism, Events and Part–Whole Relations. In E. Lepore and B. McLaughlin (Eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 444–58). Oxford: Blackwell.

Hursthouse, R. (1991) Arational Actions. *The Journal of Philosophy*, *88*, 57–68.

Jeffrey, R. (1983) *The Logic of Decision*, 2nd ed. Chicago: University of Chicago Press.

Joyce, J. M. (1999) *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Kenny, A. J. P. (1963) *Action, Emotion and Will*. London and New York: Routledge and Humanities Press.

Kim, J. (1984a) Concepts of Supervenience. *Philosophy and Phenomenological Research*, *45*, 153–76.

Kim, J. (1984b) Epiphenomenal and Supervenient Causation. *Midwest Studies in Philosophy*, *9*, 257–70.

Kim, J. (1984c) Self-Understanding and Rationalizing Explanations. *Philosophia Naturalis*, *21*, 309–20.

Kim, J. (1985) Psychophysical Laws. In E. Lepore (Ed.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 369–86). Oxford: Blackwell.

Kim, J. (1993a) Can Supervenience and Non-Strict Laws Save Anomalous Monism? In J. Heil and A. R. Mele (Eds.), *Mental Causation* (pp. 19–26). New York: Clarendon Press.

Kim, J. (1993b) *Supervenience and Mind: Selected Philosophical Essays*. New York: Cambridge University Press.

Klein, P. D. (1986) Radical Interpretation and Global Skepticism. In E. Lepore (Ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Cambridge: Blackwell.

Kolmogorov, A. N. (1956) *Foundations of the Theory of Probability*, 2nd English ed. New York: Chelsea.

Koslicki, K. (1999) The Semantics of Mass-Predicates. *Nous*, *33*(1), 46–91.

Kotatko, P., Pagin, P., and Segal, G. (Eds.) (2001) *Interpreting Davidson*. Stanford, Calif.: CSLI Publications.

Landman, R. (1996) Plurality. In S. Lappin (Ed.), *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell.

Larson, F., and Segal, G. (1995) *Knowledge of Meaning*. Cambridge, Mass.: MIT Press.

Latham, N. (1999) Davidson and Kim on Psychophysical Laws. *Synthese*, *118*, 121–43.

Lepore, E. (1986) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell.

Lepore, E., and Loewer, B. (1987) Mind Matters. *The Journal of Philosophy*, *84*, 630–42.

Lepore, E., and Loewer, B. (1989) You Can Say That Again. *Midwest Studies in Philosophy*, *14*, 338–56.

Lepore, E., and Ludwig, K. (2000) The Semantics and Pragmatics of Complex Demonstratives. *Mind*, *109*(434), 199–240.

Lepore, E., and Ludwig, K. (2003) Outline of a Truth Conditional Semantics for Tense. In Q. Smith (Ed.), *Tense, Time and Reference*. Cambridge, Mass.: MIT Press.

Lepore, E., and Ludwig, K. (forthcoming) *Donald Davidson: Truth, Meaning, Language and Reality*. New York: Oxford University Press.

Lepore, E., and Ludwig, K. (n.d.) "The Inutility of Meanings." Unpublished manuscript.

Lepore, E., and McLaughlin, B. (Eds.) (1985) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell.

Loar, B. (1976) Two Theories of Meaning. In G. Evans and J. McDowell (Eds.), *Truth and Meaning: Essays in Semantics* (pp. 138–61). Oxford: Clarendon Press.

Locke, D. (1974) Reasons, Wants, and Causes. *American Philosophical Quarterly*, *11*, 169–79.

Lombard, L. B. (1985) How Not to Flip the Prowler: Transitive Verbs of Action and Actions. In E. Lepore and B. McLaughlin (Eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 268–81). Oxford: Blackwell.

Ludwig, K. (1994) First Person Knowledge and Authority. In G. Preyer, F. Siebelt, A. Ulfig (Eds.), *Language, Mind, and Epistemology: On Donald Davidson's Philosophy* (pp. 367–98). Dordrecht: Kluwer.

Ludwig, K. (1997) The Truth about Moods. *Protosociology, 10: Cognitive Semantics I – Conceptions of Meaning*, 19–66.

Ludwig, K. (2002) What Is the Role of a Truth Theory in a Meaning Theory? In D. Shier, J. K. Campbell, and M. O'Rourke (Eds.), *Meaning and Truth: Investigations in Philosophical Semantics* (pp. 142–63). New York: Seven Bridges Press.

Ludwig, K., and Ray, G. (1998) Semantics for Opaque Contexts. *Philosophical Perspectives*, *12*, 141–66.

Ludwig, K. A. (1992) Skepticism and Interpretation. *Philosophy and Phenomenological Research*, *52*(2), 317–39.

Ludwig, K. A. (1996) Duplicating Thoughts. *Mind and Language*, *11*(1), 92–102.

Lycan, W. (1981) Psychological Laws. *Philosophical Topics*, *12*, 9–38.

Marras, A. (1999) Davidson on Intentional Causation. In D. Fisette (Ed.), *Consciousness and Intentionality: Models and Modalities of Attribution* (pp. 311–24). Dordrecht: Kluwer.

Matthews, R. (1994) The Measure of Mind. *Mind*, *103*(410), 131–46.

McDowell, J. (1999) Scheme-Content Dualism and Empiricism. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 87–104). Chicago: Open Court.

McFetridge, I. (1976) Propositions and Davidson's Account of Indirect Discourse. *Proceedings of the Aristotelian Society*, *76*, 131–45.

McGinn, C. (1986) Radical Interpretation and Epistemology. In E. Lepore (Ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp. 356–68). Oxford: Blackwell.

McLaughlin, B. (1989) Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical. *Philosophical Perspectives*, *3*, 109–36.

McLaughlin, B. (1993) On Davidson's Response to the Charge of Epiphenomenalism. In J. Heil and A. R. Mele (Eds.), *Mental Causation* (pp. 25–40). New York: Clarendon Press.

Melden, A. I. (1961) *Free Action*. London: Routledge and Paul.

Mele, A. R. (1987) *Irrationality: An Essay on Akrasia, Self-deception, and Self-control*. New York: Oxford University Press.

Mele, A. R. (1988) Effective Reasons and Intrinsically Motivated Actions. *Philosophy and Phenomenological Research*, *48*, 723–31.

Mele, A. R. (1992a) Acting for Reasons and Acting Intentionally. *Pacific Philosophical Quarterly*, *73*, 355–74.

Mele, A. R. (1992b) Intentions, Reasons, and Beliefs: Morals of the Toxin Puzzle. *Philosophical Studies*, *68*, 171–94.

Mele, A. R. (1992c) *Springs of Action: Understanding Intentional Behavior*. New York: Oxford University Press.

Mele, A. R. (1995) *Autonomous Agents: From Self-control to Autonomy*. New York: Oxford University Press.

Mele, A. R. (1997a) Agency and Mental Action. *Philosophical Perspectives*, *11*, 231–49.

Mele, A. R. (1997b) Introduction. In A. R. Mele (Ed.), *The Philosophy of Action* (pp. 1–26). Oxford: Oxford University Press.

Mele, A. R. (2000) Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance. *Philosophical Perspectives*, *14*, 279–300.

Mele, A. R. (2003) *Motivation and Agency*. New York: Oxford University Press.

Mele, A. R., and Moser, P. K. (1994) Intentional Actions. *Nous*, *28*, 39–68.

Mill, J. S. (1961) *A System of Logic*, 8th ed., reprint. London: Longmans, Green.

Montague, R. (1974) English as a Formal Language. In R. Thomason (Ed.), *Formal Philosophy* (pp. 188–221). New Haven, Conn.: Yale University Press.

Moore, G. E. (1922) The Conception of Intrinsic Value. In *Philosophical Studies* (pp. 253–75). London: Routledge and Kegan Paul.

Morgan, C. L. (1923) *Emergent Evolution*. London: Williams and Norgate.

Nagel, E. (1961) *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt Brace and World.

Nagel, T. (1999) Davidson's New Cogito. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 195–206). Chicago: Open Court.

Neale, S. (1995) The Philosophical Significance of Gödel's Slingshot. *Mind*, *104*, 761–865.

Neale, S., and Dever, J. (1997) Slingshots and Boomerangs. *Mind*, *106*, 143–68.

O'Connor, T. (1995) Agent Causation. In T. O'Connor (Ed.), *Agents, Causes, and Events* (pp. 173–200). New York: Oxford University Press.

Oppy, G. (1997) The Philosophical Insignificance of Gödel's Slingshot. *Mind*, *106*(421), 121–41.

O'Shaughnessy, B. (1973) Trying (as the Mental 'Pineal Gland'). *The Journal of Philosophy*, *70*, 365–86.

O'Shaughnessy, B. (1980) *The Will: A Dual Aspect Theory*. New York: Cambridge University Press.

Parsons, T. (1990) *Events in the Semantics of English: A Study in Subatomic Semantics*. Cambridge, Mass.: MIT Press.

Peacocke, C. (1985) Intention and Akrasia. In B. Vermazen and M. Hintikka (Eds.), *Essays on Davidson: Actions and Events* (pp. 51–73). Oxford: Clarendon Press.

Pears, D. (1982) How Easy Is Akrasia? *Philosophia*, *11*, 33–50.

Pears, D. F. (1984) *Motivated Irrationality*. New York: Clarendon Press.

Pianesi, F., and Varzi, A. (2000) Events and Event Talk. In J. Higginbotham, F. Pianesi, and A. C. Varzi (Eds.), *Speaking of Events* (pp. 3–48). New York: Oxford University Press.

Pietroski, P. M. (2000) *Causing Actions*. Oxford: Oxford University Press.

Pietroski, P. M. (forthcoming-a) The Character of Natural Language Semantics. In A. Barber (Ed.), *Epistemology of Language*. Oxford: Oxford Univesity Press.

Pietroski, P. M. (forthcoming-b) *Events and Semantic Architecture*. Oxford: Oxford University Press.

Platts, M. (1980) *Reference, Truth, and Reality: Essays on the Philosophy of Language*. London: Routledge and Kegan Paul.

Platts, M. (1997) *Ways of Meaning: An Introduction to Philosophy of Language*, 2nd ed. Cambridge, Mass.: MIT Press.

Preyer, G. (Ed.) (1994) *Language, Mind, and Epistemology: On Donald Davidson's Philosophy*. Dordrecht: Kluwer.

Priest, G. (1981) Review of *Theory and Meaning*. *The Philosophical Quarterly*, *31*, 77–9.

Quine, W. V. (1940) *Mathematical Logic*. New York: Norton.

Quine, W. V. O. (1960) *Word and Object*. Cambridge, Mass.: MIT Press.

Quine, W. V. O. (1969) *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W. V. O. (1985) Events and Reification. In E. Lepore and B. McLaughlin (Eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 162–71). Oxford: Blackwell.

Quine, W. V. (1990a) Comment on Davidson. In R. B. Barrett and R. F. Gibson (Eds.), *Perspectives on Quine* (p. 78). Cambridge: Blackwell.

Quine, W. V. (1990b) Three Indeterminacies. In R. B. Barrett and R. F. Gibson (Eds.), *Perspectives on Quine* (pp. 1–16). Cambridge: Blackwell.

Quinn, A. (1982) *Figures of Speech: 60 Ways to Turn a Phrase*. Salt Lake City: G. M. Smith.

Quinn, W. (1993) *Morality and Action*. Cambridge: Cambridge University Press.

Ramsey, F. P. (1931) Truth and Probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays* (pp. 156–98). London: Routledge and Kegan Paul.

Reichenbach, H. (1947) *Elements of Symbolic Logic*. New York: Macmillan.

Rescher, N. (1962) Plurality Quantification. *Journal of Symbolic Logic*, *27*, 373–4.

Richard, M. (1992) Semantic Competence and Disquotational Knowledge. *Philosophical Studies*, *65*(1–2), 37–52.

Scanlon, T. (1998) *What We Owe to Each Other*. Cambridge, Mass.: Harvard University Press.

Scheffler, I. (1954) An Inscriptional Approach to Indirect Quotation. *Analysis*, *10*, 83–90.

Schein, B. (1993) *Plurals and Events*. Cambridge, Mass.: MIT Press.

Schein, B. (2002). Events and the Semantic Content of Thematic Relations. In G. Preger and G. Peter (Eds.), *Logical Form and Language* (pp. 263–344). Oxford: Clarendon Press.

Schiffer, S. (1987) *Remnants of Meaning*. Cambridge, Mass.: MIT Press.

Schilpp, P. A. (Ed.) (1942) *The Philosophy of G. E. Moore*. New York: Tudor Publishing.

Searle, J. (1983) *Intentionality*. Cambridge: Cambridge University Press.

Smart, J. J. C. (1959) Sensations and Brain Processes. *The Philosophical Review*, *68*, 141–56.

Soames, S. (1989) Direct Reference and Propositional Attitudes. In *Themes from Kaplan* (pp. 393–420). New York: Oxford University Press.

Soames, S. (1992) Truth, Meaning, and Understanding. *Philosophical Studies*, *65* (1–2), 17–36.

Sosa, E. (1984) Mind-Body Interaction and Supervenient Causation. *Midwest Studies in Philosophy*, *9*, 271–82.

Sosa, E. (1986) 'Circular' Coherence and 'Absurd' Foundations. In E. Lepore (Ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp. 387–97). Oxford: Blackwell.

Sosa, E. (1993) Davidson's Thinking Causes. In J. Heil and A. R. Mele (Eds.), *Mental Causation* (pp. 41–50). New York: Clarendon Press.

Sosa, E. (1997a) How to Resolve the Pyrrhonian Problematic: A Lesson from Descartes. *Philosophical Studies*, *85*(2–3), 229–49.

Sosa, E. (1997b) Reflective Knowledge in the Best Circles. *Journal of Philosophy*, *94*(8), 410–30.

Stich, S. (1976) Davidson's Semantic Program. *Canadian Journal of Philosophy*, *6*, 201–27.

Stoecker, R. (Ed.) (1993) *Reflecting Davidson*. Berlin: W. de Gruyter.

Stoutland, F. (1980) Oblique Causation and Reasons for Actions. *Synthese*, *43*, 351–67.

Strawson, P. F. (1965) Truth: A Reconsideration of Austin's Views. *Philosophical Quarterly*, *15*, 289–301.

Strawson, P. F. (1985) Causation and Explanation. In B. Vermazen and M. Hintikka (Eds.), *Essays on Davidson: Actions and Events* (pp. 115–36). Oxford: Oxford University Press.

Stroud, B. (1999) Radical Interpretation and Philosophical Scepticism. In L. E. Hahn (Ed.), *The Philosophy of Donald Davidson* (pp. 139–61). Chicago: Open Court.

Tarski, A. (1944) The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research*, *4*, 341–76.

Tarski, A. (1983 [1932]) The Concept of Truth in Formalized Languages. In *Logic, Semantics, Metamathematics*, 2nd ed. (pp. 152–278). Indianapolis: Hackett.

Taylor, B. (1985) *Modes of Occurrence: Verbs, Adverbs, and Events*. Oxford: Blackwell.

Taylor, C. C. W. (1980) Plato, Hare and Davidson on Akrasia. *Mind*, *89*, 499–518.

Thalberg, I. (1972) *Enigmas of Agency: Studies in the Philosophy of Human Action*. London: Allen and Unwin.

Thalberg, I. (1977) *Perception, Emotion and Action*. Oxford: Blackwell.

Thomson, J. (1971) Individuating Actions. *The Journal of Philosophy*, *68*, 771–81.

Thomson, J. J. (1977) *Acts and Other Events*. Ithaca, N.Y.: Cornell University Press.

Vermazen, B., and Hintikka, M. B. (Eds.) (1985) *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press.

Vlach, F. (1983) On Situation Semantics for Perception. *Synthese*, *54*, 129–52.

Wallace, J. (1970) On the Frame of Reference. *Synthese*, *22*, 117–51.

Wallace, J. (1978) Logical Form, Meaning, Translation. In M. Guenthener-Reutter (Ed.), *Meaning and Translation* (pp. 45–58). London: Duckworth.

Watson, G. (1977) Skepticism about Weakness of Will. *Philosophical Review*, *86*, 316–39.

Weinstein, S. (1974) Truth and Demonstratives. *Nous*, *8*, 179–84.

Whitehead, A. N. (1929) *Process and Reality: An Essay in Cosmology*. Cambridge: Cambridge University Press.

Wiggins, D. (1980) 'Most' and 'All': Some Comments on a Familiar Programme. In M. Platts (Ed.), *Reference, Truth and Reality: Essays on the Philosophy of Language* (pp. 318–46). London: Routledge and Kegan Paul.

Wilson, G. M. (1985) Davidson on Intentional Action. In E. Lepore (Ed.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (pp. 29–43). Oxford: Blackwell.

Wilson, N. (1959) Substances without Substrata. *Review of Metaphysics*, *12*, 521–39.

Wittgenstein, L. (1950) *Philosophical Investigations*. London: Macmillan.

Wittgenstein, L. (1961) *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.

Woods, M. (1972) Reasons for Action and Desire. *Proceedings of the Aristotelian Society*, *46*, 189–201.

Yalowitz, S. (1997) Rationality and the Argument for Anomalous Monism. *Philosophical Studies*, *87*, 235–58.

Yalowitz, S. (1998) Causation in the Argument for Anomalous Monism. *Canadian Journal of Philosophy*, *28*, 183–226.

Name Index

Subject Index