Peirce's Theory of Signs

T. L. Short

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521843201

This page intentionally left blank

Peirce's Theory of Signs

In this book, T. L. Short corrects widespread misconceptions of Peirce's theory of signs and demonstrates its relevance to contemporary analytic philosophy of language, mind, and science. Peirce's theory of mind, naturalistic and nonreductive, bears on debates of Fodor and Millikan, among others. His theory of inquiry avoids foundationalism and subjectivism, while his account of reference anticipated views of Kripke and Putnam. Peirce's realism falls between "internal" and "metaphysical" realism and is more satisfactory than either. His pragmatism is not verificationism; rather, it identifies meaning with potential growth of knowledge. Short distinguishes Peirce's mature theory of signs from his better-known but paradoxical early theory. He develops the mature theory systematically on the basis of Peirce's phenomenological categories and concept of final causation. The latter is distinguished from recent and similar views, such as Brandon's, and is shown to be grounded in forms of explanation adopted in modern science.

T. L. Short is Chairman of the Board of Advisors to the Peirce Edition Project (Indiana University–Purdue University, Indianapolis). He has published broadly in the philosophy of science, conceptual change, teleology, and aspects of the philosophy of C. S. Peirce in journals such as *The Monist, American Philosophical Quarterly, Grazer Philosophische Studien, the Transactions of the Charles S. Peirce Society, Biology and Philosophy*, and *Synthese*.

for

Mike, Polly, Ben, Becky, and Dave

Peirce's Theory of Signs

T. L. SHORT

CAMBRIDGE
UNIVERSITY PRESS

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

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

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

www.cambridge.org
Information on this title: www.cambridge.org/9780521843201

First published in print format 2007

ISBN-13 978-0-511-27364-3 eBook (EBL)
ISBN-10 0-511-27364-9 eBook (EBL)

ISBN-13 978-0-521-84320-1 hardback
ISBN-10 0-521-84320-0 hardback

Contents

Preface *page* ix

Acknowledgments xvii

1 Antecedents and Alternatives 1
- 1. *Peirce* 1
- 2. *Sources of Peirce's Semeiotic in Locke and Kant* 2
- 3. *Brentano on Intentionality* 6
- 4. *Chisholm, Quine, et al. on Intentionality* 11
- 5. *Saussure's Semiology* 16
- 6. *Aristotle, the Stoics, St. Augustine* 21

2 The Development of Peirce's Semeiotic 27
- 1. *1865–1866: Thoughts as Representations* 28
- 2. *1867: The 'New List'* 31
- 3. *1868–1869: Thought-signs* 32
- 4. *1859–1877: Nominalism versus Realism* 36
- 5. *Three Flaws in the 1868–1869 Doctrine of Thought-signs* 42
- 6. *Derrida* et cie 45
- 7. *1877–1885: The First Flaw Corrected* 46
- 8. *After 1885: Consequences of the Foregoing* 51
- 9. *1903: The Second Flaw Corrected* 53
- 10. *1907: The Last Flaw Corrected* 56

3 Phaneroscopy 60
- 1. *The 1902 Architectonic* 61
- 2. *The Phaneron and Phaneroscopic Method* 66
- 3. *The Language of Phaneroscopy* 71
- 4. *1stness and 2ndness* 75
- 5. *Two Forms of Generality* 78

6. *The Experience of Continuity* 80
7. *The Experience of Causing* 82
8. *3rdness* 84
9. *The Categories Interpreted Metaphysically* 86
10. *The System of Categories* 89

4 A Preface to Final Causation 91
1. *Strange Objects of Desire* 92
2. *What Is Mechanical?* 94
3. *Teleology's Locus Classicus* 98
4. *A Budget of Errors* 103
5. *Hume's Ghost* 105
6. *Ordinary Purposes* 108
7. *The Mysterious Case of the Surplus Body* 112

5 Final Causation 117
1. *Explanation in Statistical Mechanics* 117
2. *Reflections on the Preceding* 124
3. *Natural Selection* 128
4. *Evolution and Entropy* 133
5. *Peirce's Concept of Final Causation* 136
6. *Comparison to Recent Views* 139
7. *Purpose's Realm* 144

6 Significance 151
1. *Teleology as Conjectural and Empirical* 152
2. *Valuation as Teleological* 153
3. *'Interpret' Defined* 156
4. *'Sign' Defined* 159
5. *'Significance' Defined* 162
6. *The Breadth of These Definitions* 162
7. *Peirce's Definitions of 'Sign'* 164
8. *Peirce's 1907 View* 168
9. *Significance and Purpose* 172
10. *Intentionality Explained* 174

7 Objects and Interpretants 178
1. *Much Groping, No Conclusion* 180
2. *Immediate, Dynamic, and Final Interpretants* 187
3. *Immediate and Dynamic Objects* 191
4. *Peirce's Realism* 196
5. *Emotional, Energetic, and Logical Interpretants* 200

8 A Taxonomy of Signs 207
1. *Qualisign, Sinsign, Legisign* 208
2. *Icon, Index, Symbol* 214
3. *Iconic, Indexical, and Symbolic Legisigns* 222

4. *A Common Error Corrected* 225
5. *Rheme, Dicisign, Argument* 231

9 More Taxa 235
1. *Principles of Semeiotic Taxonomy* 235
2. *Dicisigns and Assertion* 242
3. *Six Trichotomies* 248
4. *Ten Trichotomies* 256
5. *Where We Are Now* 260

10 How Symbols Grow 263
1. *Hypostatic Abstraction* 264
2. *The Hiddenness of Abstraction* 270
3. *A Very Virtuous Variety of Vagueness* 274
4. *Abstraction and Rigid Designation* 276
5. *Incommensurability and Meaning's 'Location'* 279
6. *Pragmatism and the Growth of Symbols* 285

11 Semeiosis and the Mental 289
1. *Contemporary Philosophy of Mind* 291
2. *Functionalism's Problem with Content* 295
3. *On Being Simple-minded* 301
4. *Beyond Biology* 303
5. *Consciousness and Subjectivity* 311

12 The Structure of Objectivity 317
1. *Antifoundationalism* 318
2. *Objectivity* 323
3. *Peirce's Concept of Science* 326
4. *A Fixation on Truth* 330
5. *How Theories Are Tested* 333
6. *Why Observe?* 337
7. *Realism, Not Relativism* 341
8. *How Aims Are Tested* 344
9. *Objectivity and Freedom* 346

Bibliography 349

Name Index 361

Subject Index 365

Preface

Peirce's theory of signs, or semeiotic, misunderstood by so many, has gotten in amongst the wrong crowd. It has been taken up by an interdisciplinary army of 'semioticians'[1] whose views and aims are antithetical to Peirce's own, and meanwhile it has been shunned by those philosophers who are working in Peirce's own spirit on the very problems to which his semeiotic was addressed. Those problems are two: to construct a naturalistic but nonreductive account of the human mind, and to explain and defend the claim that the sciences are objective in their mode of inquiry and in fact yield knowledge of an independently existing reality. In the following pages, I attempt to show how contemporary discussions in the philosophies of mind and science might benefit from a deeper study of Peirce's ideas. The purpose of this book is to say what Peirce's theory of signs is and to suggest what its philosophical significance may be.

As to the philosophy of mind: Peirce's mature theory of signs (as opposed to his early theory) is germane to the issues framed by Putnam, Searle, Dretske, Dennett, Fodor, and others. Obviously, a detailed taxonomy of signs, such as Peirce provided, might be of some help to anyone attempting to account for thought as a form of representation. Much more importantly, however, the mature semeiotic was developed in an attempt to explain, on a naturalistic basis, what we (not Peirce) call the 'intentionality' of mind. I argue that that attempt succeeds

[1] I use 'semeiotic', in Peirce's occasional spelling, for his theory or theories of signs, and the more usual 'semiotic' for that movement which originated in Europe (chapter 1, section 5) independently of Peirce and that later appropriated him, with confusion all around.

where similar, more recent attempts falter, because it was in one respect bolder.

Peirce was bold in many ways, but the particular boldness that matters here is in the ontological depth of his theory of final causation. But for that depth, his theory would be little different from the views of teleological explanation recently propounded by biologically minded philosophers such as William Wimsatt, Larry Wright, Robert Brandon, and Ruth Garrett Millikan, none of whom denies that the real world is mechanistic *au fond*. Please do not misunderstand: despite his occasional adoption of the language of the romantic idealists (Schelling particularly), Peirce's teleology is not a rejection of the physicalism that prevails in philosophy today. Instead, it challenges contemporary philosophy's unexamined conception of the physical. Peirce argued that physical explanations are not always mechanistic and that what is explained teleologically cannot be explained mechanistically; we shall conclude that what is explained teleologically or otherwise nonmechanistically are irreducibly nonmechanical aspects of physical processes.

Necessarily, we will also touch on issues in the philosophy of language; for they are implicated in contemporary debates in every area of philosophy. Besides, a theory of signs as broad as Peirce's must entail a philosophy of language. In particular, we cannot avoid reconstructing Peirce's defense of a version of realism that, contrary to the usual view taken of his philosophy, falls between 'internal realism' and 'metaphysical realism', as these are defined by Hilary Putnam. What I shall name 'Peirce's realism' rejects that dichotomy. Peirce's realism is essential to his theory of knowledge and philosophy of science, but his argument for it belongs to the philosophy of language; hence, it is to be found within our systematic statement of the mature semeiotic.

Some of Peirce's anticipations of later philosophers – Reichenbach's frequency concept of probability, Popper's idea of theories as conjectures and his propensity concept of probability – are well known, but others, equally important, are not. The 'holistic' account of meaning presupposed in the worries about scientific objectivity raised by Feyerabend and Kuhn was anticipated by Peirce, as was the view sometimes deployed in opposition to holism, namely, the causal account, associated with Kripke and Putnam, among others, of some kinds of reference. Peirce's pragmatism combined those seemingly disparate views, with a third element added, of a potentiality for future growth as essential to present meaning. That is clearer in his semeiotic writings than in those canonically 'pragmatic', and it removes the standard objections that have been made to

his pragmatism. It also shows how scientific inquiry is objective despite observation's being 'theory-laden'.

A thorough discussion of any contemporary issue, let alone so many, is impossible within the limits of this one book, which must cover so much else besides. I do no more than indicate the ways in which Peirce's theory bears on some questions of current interest. That occurs here and there but primarily in the last three chapters. Although I have made those remarks as exact, complete, and persuasive as I could, I do not pretend that they are anything more than sketchy suggestions.

So, why bother? Apart from their possibly being of some use, another virtue I would claim for these suggestions is that they hang together. In one respect, the tenor of Peirce's work runs counter to contemporary philosophical fashion, which is to atomize issues. Every new puzzle disclosed becomes a site for a new flood of specialist debate, pursued largely out of relation, except for the borrowing of techniques, to work on every other puzzle.[2] (Specialization is essential to modern science, but is it appropriate to philosophy?) Not that Peirce had a grand system. He was always dissatisfied. His emphasis was on inquiry, on endless growth of knowledge, in philosophy no less than in the special sciences. But system building is not the only alternative to fragmentation. Here, too, the study of Peirce's thought may prove salutary.

How to Read this Book

This is the plan: the first two chapters are introductory, the next three lay the foundations for the mature semeiotic, which is developed systematically in the succeeding four chapters, and the last three chapters seek to apply the foregoing to contemporary issues. It works out almost that way, but there is a good deal of leakage between compartments.

Some chapters or sections of chapters contain fairly dense textual analyses that readers willing to take my word for what Peirce said may want to skip. These are: all of chapter 2, sections 7–9 of chapter 6, section 1 of chapter 7, and sections 1, 3, and 4 of chapter 9.

Those doubting the value of time spent grappling with Peirce may want to look first at chapters 10–12, and only then, if curiosity has been aroused, read chapters 3–7. But everything depends on Peirce's phaneroscopy

[2] There are of course important exceptions, but as to the general tenor, at least in the philosophy of language, see Scott Soames' Epilogue to vol. 2 of his masterly summation of analytic philosophy in the twentieth century (Soames 2003).

(chapter 3) and the development and defense of his idea of final causation (chapters 4 and 5).

I have not assumed that the reader has any specialist knowledge, whether of Peirce's philosophy or of formal logic or of contemporary philosophy. Thus the book should be accessible to anyone philosophically interested. Yet I cannot claim that it is easy reading. For many difficult issues are discussed in it, all of them concisely.

One last remark in this vein: it may be objected that a great deal of my own thought obtrudes in my account of Peirce's views. I blush and am embarrassed, but I cannot help it. For one cannot make sense of Peirce's semeiotic without filling in the gaps, selecting the variants that make the most sense, and showing how the parts fit together, even if that means making a few corrections. After all, he was never satisfied with his own statements of the doctrine; he never finished any statement of it. And besides, Peirce wrote philosophy 'like a scientist',[3] setting out ideas not intended as final but to be applied and developed, perhaps by others. The argument for those ideas is not wholly on the page but consists in what can be done with them – just as pragmatism prescribes. Everything I say here that is in some sense 'mine', I first thought in an effort to comprehend Peirce's thought.

Other Views of Peirce's Semeiotic

In the interest of setting out my interpretation of Peirce's theory succinctly, I have avoided to a large extent examining contrary views; areas of controversy are indicated by citations of the literature or, often, by citation of my own articles in which that literature is cited and addressed. It may be well, then, to enumerate here the major alternatives to the view I shall present. Despite the vast amount that has been written on or that exploits Peirce's sign theory, its direct expositions are few and brief.

The major alternatives, I would say, are Karl-Otto Apel's 'semiotical transformation of transcendental logic' (1980, 1981, 1995) and David Savan's 'ordinal' interpretation (Savan 1987; cf. Short 1986a and Savan's response, Savan 1986). More or less in the Savan mode are James Jakób Liszka's 1996 book, a comprehensive, systematic exposition, and Gérard

[3] The words are those of the geologist Victor Baker in conversation, explaining why he found reading Peirce more rewarding than reading other philosophers. It got me thinking. I think it explains why philosophers find Peirce's writings frustrating, and I think it indicates how Peirce ought to be read.

Deledalle's 1987 and 2000 books, written with swift *élan* by the leading, recently deceased, French expositor of American philosophy. In Short 1996b, I have disputed earlier expressions of Liszka's view; my objections apply as well to his book, which appeared in the same year. Another alternative, emphasizing semeiotic's application to the analysis of communication, and perhaps overemphasizing the role of that analysis in Peirce's semeiotic, is due to the anthropologist Richard Parmentier (1985, 1994), and is illuminatingly discussed by Mats Bergman (2000); see also Jürgen Habermas's 1995 article and Klaus Oehler's 1995 response thereto. Douglas Greenlee's 1973 monograph continues to be cited despite its having been shown, repeatedly and irrefutably, to be entirely wrong (Oehler 1974, Brock 1977, Ransdell 1977, and some long footnotes in Short 1981a, 1982). Charles Morris (1938, 1946, 1964) is often taken as a guide to Peirce, but wrongly. Morris never claimed to be presenting Peirce's views, and, in fact, his theory, unlike Peirce's, was behavioristic, especially in its earlier formulations.

There have also been many publications less thoroughly opposed to the view I shall develop here. I mention only those that address Peirce's theory as a whole. First in importance are articles of 1978 and 1983 by Max Fisch, the late dean of Peirce scholars (Fisch 1986, chs. 17–18). Although not a systematic exposition of Peirce's semeiotic, John J. Fitzgerald's 1966 book should also be mentioned for its early success in placing that theory in its philosophical context. Joseph Ransdell, in articles but alas no book, forcefully states a view that in some ways is close to mine but that differs from it in interesting and important ways (1976, 1977, 1979, 1981). A 1993 book by the Danish literary theorist Jørgen Dines Johansen contains an extensive and sensitive exposition of Peirce's theory citing many manuscript sources. These authors have not distinguished Peirce's mature from his early theory as decisively as I do – something they may feel is to their credit.

Continental writers, approaching Peirce from a background of Saussurean semiology, have systematically misinterpreted his semeiotic. For the two doctrines are fundamentally incompatible (chapter 1, section 5). The unholy union of Saussure's supposed conventionalism with the breadth of Peirce's mature semeiotic gave bastard birth to an extreme relativism and irrealism – a modern version of sophistry that Saussure and Peirce would both have rejected. I therefore treat those writings not as an alternative reading of Peirce's semeiotic but as an alternative to it. For the most part, it is an alternative I ignore, but see chapter 2, section 6,

for brief comment on Jacques Derrida's and Umberto Eco's reading of Peirce.

One last strain of interpretation of Peirce's semeiotic must be mentioned. With the encouragement of the late Thomas Sebeok, the linguist and American impresario of semiotics, a number of authors, some of them from the natural sciences, have extended the naturalistic view I favor beyond what I take to be intelligible limits. To be sure, the concepts of information theory may be extended to genetics, but that does not mean that Peirce's semeiotic may be so extended; unlike information theory, it accounts for intentionality, but it does not bring intentionality down to the level of DNA and RNA. Nevertheless, Claus Emmeche (1991, 1998), Jesper Hoffmeyer (1996), Emmeche and Hoffmeyer (1991), and Lucia Santaella Braga (1999a, b) are of interest. Helmut Pape's long and ambitious study (1989) properly places Peirce's semeiotic in phenomenological and teleological context but overextends the theory, less biologically than cosmologically.

A Note on Terminology

I avoid technical language where possible and explain such terms as I do use. My slight use of formal logic and occasional references to its apparatus are not sufficient to block the understanding of anyone not familiar with that subject. Peirce's famously rebarbative neologisms are explained where they cannot be avoided. Concepts evolved in the long history of philosophy are another matter. They might be taken to be well established and understood, except for the awkward fact that in every philosophy they are understood differently. Peirce's glosses on such terms as 'real' and 'individual' are of the greatest interest. Perhaps least in need of definition are the nouns 'universal' and 'particular', as their use in philosophy has been fairly uniform. And yet they are so fundamental to every phase of this book's argument that I define them here and then review some of the finer points, so as to forestall misunderstandings.

'Universal' is the standard translation of Aristotle's *katholon* and is universally understood as Aristotle understood the latter (not in all texts equally but in *De Int.* 7 primarily), as that which is said of many. We may gloss this as: that which, as a matter of grammar, not as a matter of real possibility, may be true of many. Being a unicorn is therefore a universal, as it would not be ungrammatical to speak of many unicorns. Opposed to the universal is the particular, which cannot grammatically be said of

many. No two people are Socrates, even if two are named 'Socrates'. (Only figuratively may one say, 'Would there were another Socrates'.) Socrates is a particular, while being old, being red, being a unicorn, being named 'Socrates', or being a particular are universals.

Notice that what is being defined here are these terms as nouns, hence, as naming kinds of thing. This usage is philosophical. The same terms as adjectives are parts of ordinary speech and are used with related meanings, though the adjective 'particular' is used more broadly, while the adjective 'universal' is used more narrowly than are the corresponding nouns. Thus, when two philosophers are debating about universals, it would not be incorrect for one to say to the other, 'Which particular universal have you in mind?' And while *a* universal is that which *may be* true of *many*, something *is* universal only if it *is* true of *all* (all, that is, of some understood class), as in, 'It is a truth universally acknowledged.' The adjective 'general' corresponds more closely, though imperfectly, to the noun 'universal' than does the adjective 'universal'.

It should not be assumed that every philosopher who uses the word 'universal' as a noun is committed to the proposition that universals exist or are real. For one can ask, 'Are universals real?' and 'Do they exist otherwise than in name?' Realists (in one sense of that overworked word) are those who assert that universals are real, that is, that there are universals independently of their being named or thought of, while nominalists are those who assert that universals exist in name (*nomen*) only.

What of something that, by conception, can be true of one particular at most and yet might have been true of some other particular than the one it is true of? There can be only one twenty-sixth president of the United States and yet it might have been someone other than Theodore Roosevelt, for example, had McKinley not been shot. I think we shall have to say that that is a universal, too, since there is more than one of which it could have been true, though it could not have been true of more than one. But notice that a phrase such as 'the twenty-sixth president,' used as the grammatical subject of a sentence, will normally denote a particular – the individual who was the twenty-sixth president in fact – not a universal.

The noun 'universal' tends to be used to refer only to that which may be true of subjects taken one at a time, for example, being human or being red. But relations may be true of things taken two at a time or three at a time, and so on. The sentence 'John is taller than' is ungrammatical; it wants to be completed by Bill. Being taller than is true of some pairs of particulars. We will count relations as universals.

Peirce often used 'general' as a noun, in place of 'universal'. That is awkward, given the military meaning that that noun has in ordinary English. Thus, he spoke of properties, relations, and laws as 'generals'. The motive is not given. Perhaps it was for the sake of agreement with the adjective (see above). Perhaps it was because a law, whether customary, enacted, or natural, is not a universal. It is general in the sense that it applies to many instances, actual or possible; but the law cannot grammatically be said of those instances. What can be said of them is that they conform to the law. The issue between nominalism and realism may nevertheless be extended to laws, hence, to all 'generals', and Peirce did so extend it.

Bibliographical Note

Peirce's writings are cited in the text parenthetically, in the ways that have become standard among Peirce scholars, as follows: citations of the form (*n.m*) refer to paragraph *m* of volume *n* of the *Collected Papers*; (W*n*:*m*) to page *m* of volume *n* of the new, chronological edition of Peirce's *Writings* (regrettably, not yet complete); (EP*n*:*m*) to page *m* of volume *n* of the *Essential Peirce*; (NEM*n*:*m*) to page *m* of volume *n* of the *New Elements of Mathematics*; (LW:*n*) to page *n* of Peirce's letters to Lady Welby in the volume *Semiotic and Significs*; (RLT:*n*) to page *n* of Peirce's 1898 lectures in Cambridge, Massachusetts, in *Reasoning and the Logic of Things*; and, finally, citations of the form (MS*n*) or (L*n*) are to manuscript *n* or letter *n*, as numbered in Robin 1967. See the Bibliography.

Acknowledgments

This book owes much to the instruction, encouragement, support, and friendship that I have received over the years from a number of philosophers and students of Peirce, many now deceased: Josiah Lee Auspitz, Joseph Brent, Jacquelin Brunning, Arthur Burks, Vincent Colapietro, Gérard Delledale, Max Fisch, George Geiger, George Gentry, Susan Haack, Peter Hare, Charles Hartshorne, Risto Hilpinen, Nathan Houser, Christian Kloesel, I. C. Lieb, Cheryl Misak, Robert Palter, Vincent Potter, Joseph Ransdell, Lucia Santaella, David Savan, Thomas Sebeok, and Michael Shapiro, among others. I must single out Richard Robin and his wife, Joanne, as my most persistent and generous cheering section, the warmth of whose support is deeply appreciated. André De Tienne took time to answer detailed queries about Peirce's intellectual history and the dating of his manuscripts. Andrew Reynolds and Nicholas Pappas each made very useful criticisms of an early version of what is now chapters 4 and 5. Mark Migotti read several chapters with scrupulous care, compelling many improvements of substance and style; my debt to Migotti is great. But my largest debt is to Nathan Houser, who read more chapters than anyone and generally kept cracking the whip.

1

Antecedents and Alternatives

The purpose of this chapter is to put Peirce's semeiotic into context. What are its antecedents and alternatives? What is the type of question it is meant to answer? It might seem that we should begin with the theory itself, so that we can know what it is that we are talking about. But that theory is complex, and a preliminary statement of it would only raise objections before they can be answered. Sketching in the background leaves an empty space for the foreground objects. My hope is that this will allow you to see the shape of things to come and that it will provide some motive for enduring those rigors that lie grimly in wait.

1. Peirce

Charles Sanders Peirce (1839–1914), whose surname is pronounced 'purse', was a son of Benjamin Peirce, a Harvard professor of mathematics and astronomy and, at the time, America's foremost mathematician. Benjamin Peirce was also a major figure in, or, more accurately, one of the creators of, the American scientific establishment. With others, he founded the National Academy of Sciences and had a hand in much else of that kind. He recognized Charles' genius and raised him accordingly, with the consequence that the latter developed an extraordinary degree of intellectual discipline and almost no moral discipline. Although trained in chemistry, Charles Peirce made a number of profoundly original contributions of the first importance to mathematical logic, meanwhile earning his living making exacting empirical measurements in astronomy and geodesy (he made several important contributions to the theory and practice of measurement, as well). Formal logic

and experimental work had each its impact on his philosophy. But what is perhaps most remarkable in Peirce's vast output is the number of fertile fields of investigation that he opened up and the startling originality of his ideas.

Pragmatism, for which Peirce is now best known, is the only major philosophical movement, barring the religious philosophies of the East, to have originated outside Europe. His work on the logic of relations, following Augustus De Morgan's and developed in turn by Ernst Schröder, contributed to Russell and Whitehead's epoch-making *Principia Mathematica* (1910–13). Slightly prior to Edmund Husserl, Peirce invented a phenomenology, or 'phaneroscopy', as he came to call it, that is comparable to Husserl's yet fundamentally different; we shall rely on it extensively in this book. He anticipated later developments, by such philosophers as Hans Reichenbach, Karl Popper, and Stephan Toulmin, in probability theory and in the theory of the natural sciences and their methods. I argue, in chapter 12, that he is still in advance of contemporary philosophy of science with respect to the issues raised in the 1960s by Paul Feyerabend and Thomas Kuhn. Moreover, he was the first of modern philosophers to recognize chance as being a basic feature of existence; at the same time, and deliberately against the modern temper, he revived Duns Scotus' realism as opposed to William of Ockham's nominalism. And so on.

All of this was accomplished even while his professional career and personal life fell into disarray. In the end, Peirce was impoverished and isolated, endlessly revising essays that he never finished. He never succeeded in bringing his ideas into systematic unity; he never published a philosophical book. The incompleteness, digressiveness, and profusion of technical detail of his writings accounts for the educated public's ignorance of his life and work and for the relative neglect of his philosophy even by professional philosophers.

2. Sources of Peirce's Semeiotic in Locke and Kant

Peirce's theory of signs had its origin in Kant's theory of knowledge. However, the term 'semeiotic' is almost certainly a transliteration of the Greek word that Locke used, at the end of his 1690 *Essay*, to name a new 'doctrine of signs'. That doctrine, Locke said, will be 'another sort of logic . . . than what we have been hitherto acquainted with' (Bk. IV, Ch. XII). This is a problematic legacy.

Among signs, Locke included both words and ideas, words being 'signs of ideas' by which we convey ideas to one another, and ideas being 'signs the mind makes use of for the understanding of things'. Locke's reason for treating ideas as signs was that

> since the things the mind contemplates are none of them, besides itself, present to the understanding, it is necessary that something else, as a sign or representation of the thing it considers, should be present to it; and these are *ideas*.

Does that make sense? It seems correct to distinguish ideas from things. My idea of an elephant is not the elephant itself. My idea may embody some error and is in any case incomplete; nor does it weigh as much. But is not my idea, for all of its defects, precisely how the elephant *is* 'present to [my] understanding'? Locke wrote as if I contemplated my idea, and not the elephant, and then inferred the elephant from it, much as I might infer an elephant from its footprint. But that is not how we employ ideas.

And is it not startling to be told that ideas are signs? A sign, in ordinary parlance, is something that makes us think of something else. Thus the footprint is a sign of the elephant: I see it in my garden and think, 'Elephant!' But an idea is not one thing that makes us think of another. It *is* the thought of that other. How, then, can it be a sign?

Locke explained, in the earlier and better-known parts of his *Essay*, that ideas are derived from particular experiences of sense or of reflection and that they are related to their objects or 'archetypes' as effect to cause and, in some cases, by resembling their causes. Presumably, it is through these two relationships, of causality and resemblance, that ideas are signs. For, in ordinary usage, we call the footprint a sign of the beast that produced it, and we infer which beast that was from the resemblance of print to foot. As causal relations and resemblances ground signification in that sort of case, one might suppose that they do also in the case of ideas.

But that lands us in the same difficulty over again. For the ground of signification is one thing, and signification itself is another. Causal relations and resemblances make something, X, a sign of something else, Y, only because they cause us to think of Y once we apprehend X. Whatever the ground of its power to cause us to think of Y, X signifies Y only because it has that power. Now, if X itself is a thought of Y, then Y *is* being thought of. Another step is not required to make Y an object of thought. Hence, X does not have to produce a thought – a further thought – of Y. But, then, it is not a sign of Y. It is of no relevance in that case, that X is either caused by its object, Y, or that it resembles Y. Those relations,

even if they obtain and even if they are somehow involved in X's being a thought of Y, do not make X a sign.

These reflections are sufficient to justify serious doubt about any attempt to analyze thought as being a species of sign. Of course, thoughts sometimes are signs. If you notice that thoughts of food keep recurring to you, you might take that as a sign that you are hungry. If I notice that you are thinking a great deal about death, I might take that as a sign that you are depressed. But what these thoughts signify is something other than what they are thoughts of.

Locke's semeiotic theory of mind faced another difficulty as well. Since he held that all ideas derive from particular experiences, as if they were lingering images thereof, he had difficulty accounting for general ideas, for example, of triangularity in general rather than of this or that particular triangle. Below (chapter 3, section 5), we distinguish a sense of generality in which an image, its cause being ignored, is general; but that is still not the generality of a common noun or of a concept, which comprises a continuum of possible variations. The concepts in which we think are always general – gray in general, elephant in general – even when they are applied, through perception, to particular objects. To think, 'This elephant is gray', is to conceive of it as one of the varied class of elephants and as having some particular shade of the color, gray; but that thought does not distinguish this elephant from all the others, nor its shade of gray from other shades. So far as we think of it as 'an elephant' and as 'gray', we are not thinking of it *in* its particularity. To be sure, we think of it *as being* particular; but nothing is more general than being particular. Being particular is a property that every particular shares.

Kant, unlike Locke, supposed ideas to be general, but he did not say what ideas are. As a term for any mental content, Kant followed Christian Wolff in using *Vorstellung*, or presentation; but he did not say what *Vorstellungen* in themselves are. Peirce developed the Kantian doctrine in a contemporary (though also Platonic) way, by identifying thought as internalized discourse: we think in the words of the language we have learned. Peirce did not limit thinking to the verbal – it can be diagrammatic and otherwise in images – but we think mostly in words, and thus our capacity to think is dependent on our having learned a language. With a different language to think in, we would think somewhat differently. (Peirce did not entertain the very speculative hypothesis, now in vogue, that there is a language common to all minds – 'mentalese' – distinct from the languages people speak.) But common nouns, verbs, adjectives, and adverbs are general, and we cannot say anything without using some

of those general terms. There is no meaningful sentence (unless elliptical) wholly couched in proper names and demonstrative pronouns. That holds for any language (but not for diagrams or images, except so far as, by verbal commentary, they are made to stand for variations they do not themselves comprise). Hence, thought is inherently general.

Now, words no more than ideas are normally called 'signs'. Nevertheless, words, when spoken or written, do conform to the general idea of signs as being that which leads one to think of something else. One hears or reads the word 'elephant' and thinks, not of that sound or inscription, but of a large, gray mammal. And there is a philosophical tradition, going back to Aristotle, of talking about words as being signs. But if thought is essentially verbal and if words are signs, then thoughts are signs. Thus we may reconstruct Peirce's return to Locke's implausible doctrine, albeit on a new basis and modified.

But how do words signify? Equivalently, how do they acquire their meanings? Locke and many others have supposed that the answer is that words express ideas, by convention. Recall our earlier distinction, between significance and its ground. The conventional relation of word to thought is a third ground of significance, alternative to causality and resemblance. As grounding the significance of a word, the idea must exist already (not necessarily in the mind of the speaker, and never only in his mind, but within the stock of ideas possessed by that community to which speaker and hearer belong). It is expressed, not produced, by the word spoken. Thus, linguistic meaning is accounted for by assuming an independent realm of thought. First there is thought, and then there is language, the primary function of which is to express thought. The first statement of this view was by Aristotle:

> Now spoken sounds are symbols of affections in the soul, and written marks symbols of spoken sounds. And just as written marks are not the same for all men, neither are spoken sounds. But what these are in the first place signs of – affections in the soul – are the same for all; and what these affections are likenesses of – actual things – are also the same. (*De Int.* 1, Ackrill trans.)

For 'affections in the soul' we may read thoughts and sensations. As they are likenesses of their objects, the relationship is natural and universal; words, being conventional, vary from nation to nation. Words obtain objects only by standing, by convention, for affections in the soul.

Peirce could not adopt that view. For him, we learn to think in learning to speak. Thought therefore depends on words having a meaning. Meaning therefore cannot depend on thought. But if the meaning of

a word is not the thought that it expresses, then how do words mean? On what is their significance based? Possibly, neither thoughts nor words exist without the other; possibly, they are significant together or not at all. But even if so, whence is their significance?

We have, now, two questions and two problems. First, what is significance? Is it for one thing to produce a thought of something else? If so, then if a thought is a sign, we have an infinite *progressus*: each thought must produce another, *ad infinitum*. Second, what is the ground of significance? If a word signifies by expressing a thought and thoughts are words, then we have an infinite *regressus*: each thought-word must express a preceding thought-word, *ad infinitum*.

3. Brentano on Intentionality

What makes one thing to be of or about another? Being of or about is a peculiar property, hard to explicate. It is also a property that thoughts and signs share. We need a name for it. Peirce, since he held that thoughts are signs, could rely on the word 'significance' to cover all cases of being 'of' or 'about'. If we wish to hold the question open, whether thoughts are signs, we shall need another term. 'Intentionality' is the best candidate. It has come to be commonly used in lieu of the term 'intentional inexistence', which was introduced in this connection by a contemporary of Peirce's, the Austrian philosopher Franz Brentano (1838–1917), in his 1874 book, *Psychologie vom empirischen Standpunkt*. Peirce appears never to have read that book; he never referred to Brentano and never used 'intentional' or its cognates in Brentano's sense, though he did occasionally refer to the Scholastic doctrine of first and second intentions, from which Brentano derived the term.

Intentionality in this sense, at least so far as Brentano saw, does not imply a purpose, as does the English word 'intend'. They have the same root, however: *intendo*, the Latin for stretching or straining toward something. Thus Brentano:

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call . . . reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. . . .
>
> This intentional inexistence is characteristic exclusively of mental phenomena. No physical phenomenon exhibits anything like it. We can, therefore, define mental phenomena by saying they are those phenomena which contain an object intentionally within themselves. (Brentano 1973[1874], pp. 88–9)

There are three things to notice in this passage: Brentano took intentionality to be coëxtensive with the mind; he understood intentionality as involving an 'inexistent' object; and he supposed that that object is somehow contained in the thought or other mental act in which it is, as we shall say, intended. We will return to the first and third propositions anon. For the present, let us try to understand what is meant by 'intentional inexistence'.

To think is to think of something, which is the object of that thought. 'Object' is here being used in a broad sense, for anything about which one can think, and not for physical objects only. The point is, one cannot say what a thought is without mentioning its object. Two thoughts are distinguished from one another by having different objects. And so also for desires, which are for this or that; fears, which are of this or that; and so on. Each has an object (normally; though there are also nameless dreads and restless yearnings for one-knows-not-what). And yet these objects need not exist; one fears the unreal and desires the impossible. Consider unicorns and griffins. Neither is real, neither exists at all, yet the concept of the one is distinguished from the concept of the other by the fact, and only by the fact, that a unicorn is one thing and a griffin is something else. One has a horn, the other has wings. These objects are 'inexistent' in this sense: existence is immaterial to their being objects of thoughts, fears, desires.

But how can something be an object without existing? That is the sort of question, it would seem, that led Brentano to declare that the inexistent object is 'in' a thought or other intention, as its 'content'. It exists after all, only not where it was thought to exist. But that is clearly wrong. A unicorn is not the sort of thing that could be in a thought; what would it find there to eat?

A way to avoid attributing existence to the inexistent is to shift from the 'material mode' of talk about things to the 'formal mode' of talk about talk. Very roughly: something has an inexistent object if (a) it cannot be fully described without mentioning that object and (b) that object need not exist in order for the description to be true. More on this in the next section. On the assumption that some such explication can succeed, we proceed to speak freely of the intentionally inexistent.

In addition to his conception of intentionality, we can distinguish two theses that Brentano propounded about intentionality. The first is his claim that intentionality is distinctive of the mental. Physical things and events do not have objects of or about which they are. According to Brentano, everything mental possesses intentionality, and nothing that is

not mental does. The second thesis is the one he was primarily concerned to establish in his book, that the mental is so fundamentally different from the physical that it eludes any attempt at naturalistic explanation: it must be made the subject of a science, descriptive psychology, that is not explanatory and that differs in method from the natural sciences. Let us examine both theses, beginning with the second.

In the natural world, relational properties, such as fatherhood or sitting, require real objects. One is not really a father if there is not a real child that he fathered, and one is not really sitting if there is not something real on which he sits. But one can really be thinking of unicorns. Brentano concluded that the mind is inexplicable by the natural sciences. At what point in a physical explanation can something that does not exist be introduced? If at no point, then physics cannot account for intentionality. Nor, thought Brentano, can intentional states be observed in accordance with the canons of observation in the natural sciences. We can see, locate, and measure only what does exist. Therefore, we could never detect mental states by such means, as they have inexistent objects. The dichotomy Brentano discerned between the psychical and the physical is a methodological variant of Descartes' ontological dualism of mind and body.

Some of Brentano's students, but most importantly Husserl, replaced his idea of psychology with that of a new science, of phenomenology, in which the identification of intentionality with the mental is a fundamental principle, and in which the exclusion of naturalistic explanation is developed and strengthened. In all its variants, even Peirce's, phenomenology is merely descriptive, not explanatory. But in Continental phenomenology it is usually maintained that intentionality can only be grasped in a self-reflective consciousness (the phenomenologists reject Brentano's reference to mental *phenomena*), and never explained. These additional assertions are omitted from Peirce's phenomenology.

I take the philosophy of Brentano, Husserl, and Continental phenomenology generally to be a major alternative to Peirce's mature semeiotic. The former denies the very possibility of a naturalistic explanation of intentionality, whereas a central thrust of Peirce's mature semeiotic is that intentionality may be explained naturalistically. Peirce rejected all dualisms, on the principle that, by positing inexplicables, they block the road of inquiry. As a corollary of that principle, he in later years proposed a doctrine of 'synechism', of the continuity of all things. The principles, of synechism and of not blocking the road to inquiry, are grand pronouncements. Of more moment would be the concrete development of

a theory that does exhibit mind's continuity with nature. And that is what I shall argue Peirce's mature semeiotic does.

As between Peirce and the phenomenologists, the *crux criticorum* is Brentano's first thesis, of which one part is that intentionality is not to be found outside of the mind. For if there were extra-mental examples of intentionality, they would provide that element of continuity, satisfying Peirce's synechism, by which the natural and the mental, the observable and the introspectable can be bridged.

The identification of intentionality with the mental is often treated as if it were a tautology. That was not Brentano's view. He began with a rough enumeration of mental phenomena *not* overtly identified by their intentionality. And his account of intentional inexistence makes no allusion to the mind. The definition of the mental, as 'those phenomena which contain an object intentionally within themselves', is a conclusion for which Brentano argued. Hence, it presupposes that we understand what intentionality is without reference to the mind. Thus it is conceivable that one might find mental phenomena lacking intentionality and/or nonmental phenomena possessing intentionality. And, indeed, apparent examples of both sorts – apparent counterexamples to Brentano's thesis – have been discussed in the literature, and to some extent by Brentano himself.

There are types of phenomena Brentano counted as mental that seem to lack intentionally inexistent objects and some that lack any object at all. For example, seeing is mental, but if something does not exist we cannot see it (though we may think we see it). And pains, though they have locations, have no objects at all; we simply suffer them. Brentano dealt with seeing easily: it involves having an image of, or thinking of, an object, and the objects of images and thinking are inexistent. The mental, then, can be said to be either that which has an inexistent object or that which has a part that has an inexistent object. That way, intentional inexistence is still part of anything mental. Pains require a further stretch. To accommodate them, Brentano denied that there is a sharp boundary line between feeling and striving (1973 [1874], pp. 235ff.). Pain is hardly separable from the desire to be rid of it, and desire has intentionally inexistent objects.

We have no need to form an opinion about those topics. Of more importance to us are the examples of nonpsychical phenomena that seem to possess intentionality. They fall into two groups. First, spoken or written words, as well as natural signs such as smoke or a falling barometer, are physical, and yet they have objects that, in typical cases, need not exist.

Words can lie, not all smoke is caused by fire, and so on. Such examples clearly conform to Brentano's definition of intentionality and yet they are not thoughts or other mental acts or states. They may nevertheless be accommodated to Brentano's thesis easily, if it can be shown that their intentionality is derivative from that of the thoughts that words and other signs either express or elicit. That is the line that Roderick Chisholm (1952) took, and it is implicit in Husserl.[1] It leads straight back to the view shared by Aristotle and Locke, that significance derives from thought. By its means, Brentano's thesis is amended, but its underlying idea is preserved intact: there is intentionality outside of the mind, it can be said, but that sort of intentionality is utterly dependent on the mind.

The other candidate for extra-mental intentionality is animal behavior, which seems always to be goal-directed. Dogs look for bones they have buried and salmon swim upstream toward the beds from which they were spawned. But a bone might be gone and a spawning bed destroyed. Can the behavior of the dog and the salmon be adequately described without referring to the objects sought, which might not exist? If not, then those actions would seem to be intentional, since they have inexistent objects. Suppose that is so. Brentano's thesis survives nonetheless if we can plausibly attribute something like thought to the lower animals.[2] The idea here may be that goal-directed behavior is always directed by some thought or image, and so on, of the goal. However, that stratagem becomes increasingly implausible as we work down the animal kingdom, from dogs to salmon to lice to paramecia. For paramecia, too, exhibit goal-directedness, for example, swimming up chemical gradients toward a food source. But is it plausible to attribute ideas of their goals even to salmon, much less to paramecia?

An alternative strategy, proceeding from the same assumption, that goal-directedness always requires direction by something mental, is to deny that animal behavior, at least beneath a certain level, roughly mammalian, is genuinely goal-directed. On that view, what appears to be goal-directedness is purely mechanical, the operation of mechanisms indifferent to consequences. The description of those operations will be a complete account of the animal's behavior, and in that description there will be no reference to anything that does not exist or that need not exist.

[1] E.g., 'A thing is only properly an indication if and where it in fact serves to indicate something to some *thinking* being'. Husserl 1970 [1900–1], p. 270, emphasis added.

[2] Brentano does that forthrightly (1973 [1874], pp. 40–1); for a more recent statement of the same, see Searle 1983, p. 5.

If that can be done, then talk of salmon, and the like, as 'seeking' anything may be dismissed as anthropomorphic. Such talk, it will be said, attributes something like human purposefulness to what is only a scaly machine. (The next step often taken is to argue that human purposefulness, also, is simply a machine function.)

However, can a mechanistic reduction of apparent purposefulness be carried out? Of course, mechanisms are involved. But can animal behavior be *fully* described and understood without mentioning intentionally inexistent objects – those things that are sought, fled, or attempted? The plausibility of Peirce's mature semeiotic depends on a negative answer to that question. The question itself, however, needs clarification and refinement, as follows.

4. Chisholm, Quine, et al. on Intentionality

Clarification of Brentano's idea of intentionality, so as to avoid attributing existence to the inexistent, was begun by Brentano himself, in essays of 1911 (1973 [1874], pp. 271–2 and 89n11). His student and editor, Oskar Kraus, in his introduction and notes to a 1924 edition of the *Psychologie* (ibid., pp. 370–1), shifted the question from material to formal mode, and formal mode analyses were pursued vigorously by a number of analytic philosophers, most notably by Roderick Chisholm in various publications from 1952 to 1967.[3]

There are many forms of expression that violate supposed canons of scientific language. For example, in the *Tractatus*, Wittgenstein supposed that in an ideal language the value, true or false, of any proposition is a function of the values of logically simple (elementary) propositions; the value of the whole is determined by the values of the parts. But 'Jones believes that *p*' has a value that is not a function of *p*'s value. Jones is notorious for believing false things as well as true ones. Another supposed canon of scientific language is the principle of substitutivity of identicals, *salva veritate*: if $x=y$, S is a sentence in which x occurs, and S' is obtained from S by substituting y for x, then S is true if and only if S' is true. And yet it may be true that you believe that your father is a good man and false that you believe that the head of Murder, Inc. is a good man, though in fact your father is the head of Murder, Inc. (He is good enough to hide his business from you.) Facts of these kinds about the logic of 'believes'

[3] Chisholm 1957, ch. 11, and Castañeda 1967, pp. 11–35. See also the papers by Chisholm and others collected in Marras 1972 and Marras's bibliographical note, pp. 506–8.

gave hope of a linguistic criterion by which to distinguish the intentional from the physical.

Before descending to details about what such a criterion might be, let us enumerate the components of the dialectical situation defined by Chisholm's project.

1. There are assumptions made, and that might be questioned, about what constitutes an ideal or canonical scientific or physicalistic language. One might also question the equation of scientific with physicalistic and/or the very idea of there being a canonical form of scientific language.
2. There is an intuitive idea of intentionality, expressed by Brentano, for which a formal-mode counterpart is wanted. It is assumed that this will be a classification of idioms – let us call them 'intentional idioms' – that do not conform to the aforementioned canons.
3. There are many expressions that do violate those canons, not all of which fit our intuitive idea of intentionality. For example, it is logically necessary that your father is a male but it is not logically necessary that the head of Murder, Inc. is a male; like 'believes', 'being logically necessary' is noncanonical, yet it seems not to be an intentional idiom. Thus, part of the task facing Chisholm was to distinguish the subset of noncanonical idioms that are intentional idioms.
4. Chisholm wished to defend Brentano's thesis that intentionality is the mark of the mental, and therefore he had not only to isolate the class of intentional idioms but also to show that all of these are psychological idioms and that no psychological idiom fails to be in the proposed class of intentional idioms. His analyses and those of others were therefore subject to test by counterexample – assuming that we can identify psychological and nonpsychological idioms independently of relying on criteria of intentionality. But what if one should find a psychological idiom that is not, by a proposed criterion, intentional? Does that refute the proposed criterion, or does it refute, instead, Brentano's thesis? Or does it call into question our intuitive idea of the psychological? The same alternatives are faced if one should find an idiom that is intentional by a proposed criterion but is evidently not psychological.
5. Chisholm wished to defend Brentano's thesis that the intentional is irreducible to the physical; in the formal mode, the task is to show that, while some uses of intentional idioms may in fact be

> replaceable without loss of meaning by physicalist description, not all can be so replaced. It follows that the argument can never be conclusively ended, as it depends on the resourcefulness of two sets of disputants, those who devise physicalistic reductions and those who show that these reductions fail or who find new examples that they think will resist reduction. Every seeming conclusion is open to challenge by someone bringing new resources to bear on the question. That is not an objection to the enterprise, merely a feature of it.

Some other philosophers in the analytic tradition, most notably, W. V. O. Quine, have agreed with Chisholm that intentional idioms are irreducible to nonintentional idioms, but have drawn a different moral: namely, that intentional idioms do not denote anything real, or that the irreducible part of their meaning does not. Quine admitted that they are practically indispensable, but maintained that they do not belong in science:

> 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. (1960, p. 221)

Direct quotation reports the words actually spoken, as if they were merely physical events. Indirect quotation provides the propositional content of such 'propositional attitudes' as believing that . . . , knowing that . . . , fearing that . . . . Quine also banished from scientific discourse all expressions that are referentially opaque (1960, §32), that is, that violate the principle of substitutivity of identicals, *salva veritate*. To take an example that is not of a propositional attitude, you may be looking for your father and yet, in one sense of 'looking for', not looking for the head of Murder, Inc.

In effect, Quine denied that believing, seeking, and so on are real conditions or activities, though he did not deny that there is something real that we misleadingly describe in such language; the realities so described are perfectly physical and lack all intentionality. Two assumptions, (a) that reality is physical and (b) that physicalist language is wholly free of intentional idioms, are shared by most contemporary philosophers of mind, whether or not they deny the reality of mind: see chapter 11. This stands as a second major alternative to Peirce's view. For Peirce held that mind is real and yet his semeiotic implies that it cannot wholly be described without recourse to intentional idioms.

The major alternatives, then, are (i) Brentano/Chisholm's, that mind is real, irreducibly intentional, and inexplicable naturalistically, (ii) that of Quine et al., that whatever is real is nonintentional and explicable naturalistically; and (iii) Peirce's, that mind is real, irreducibly intentional, and explicable naturalistically.

The dialectical situation defined by Peirce's view is less demanding than is that of Chisholm's project. Components (1) and (4) drop out. No assumption need be made about what scientific language ought to be. As the thesis that intentionality marks the mental is not to be defended but, rather, challenged, a thicket of possible counterexamples is evaded. In fact, some of the examples that occasioned Chisholm and his colleagues the most difficulty are grist for a Peircean mill. And (2) plus (3) is softened, as a complete theory of intentional idioms is not needed. It suffices to identify a broad class of idioms that (a) violate Quine's canons and (b) are plausibly a formal mode counterpart of Brentano's material mode conception of intentional inexistence. We can then state a sufficient (not a necessary) condition, in the formal mode, of material-mode intentionality. It is this: something possesses intentionality if it cannot fully be described without implying the truth of a proposition that cannot be stated without employing one or another idiom of the defined class of intentional idioms.

Dialectical component (5) still applies in full force: determining intentionality by this criterion depends on the resourcefulness of two sets of disputants, those who try to find instances of seemingly irreducible intentionality and those who try to reduce those instances. The question faced by Quine et al. is whether one can plausibly maintain that no information about the world is lost if we refrain from using intentional language. Brentano's thesis will be refuted if a class of examples is found that are not mental in his sense but that are by this criterion irreducibly intentional.

For our purposes, a sketch will suffice. Consider the transitive verbs 'throw' and 'look for'. What is thrown must exist, what is looked for need not exist. There are grammatical contexts in which neither verb takes an object that must exist, for example, 'Joe does not throw a ball', 'Joe should throw a ball', 'Grandpa would like to see Joe throw a ball'. And there are ways of specifying the object ('the ball', 'his ball') that imply its existence in any case. But in the grammatically simplest singular affirmations where the object is indefinite ('a ball'), the difference between these two verbs is clear. From 'Joe throws a ball', we may infer that a ball exists; from 'Joe looks for a ball', that same inference may not be made. We shall call any

verb of the latter kind 'intentional', regardless of the context in which it occurs.

A parallel distincton can be made between verbs ('believes that', 'sees that') that take propositional objects, the question being whether, in a simple singular affirmation, the proposition may be inferred. (Some verbs, e.g., 'disproves', license inference to the denial of the propositional object; for our purposes, we may, for the sake of simplicity, mandate restatement by which, e.g., '*S* disproves *p*' is replaced by '*S* proves not-*p*', making *p*'s negation the propositional object.) Wrapping up the two cases together, let us say that a verb is intentional if its use in simple singular affirmations does not license an inference that its object, if indefinite, exists or, if propositional, obtains.

This definition is stipulative and not subject to disproof by counterexample (of course, it might be open to other kinds of objection). It makes no difference that many verbs used to ascribe intentionality are not by this definition intentional. It makes no difference that instances can be found of verbs that are intentional by this definition, for example, 'wants', that are not invariably used to ascribe intentionality ('This frying pan wants a handle').

The claim is that something possesses intentionality if it cannot fully be described without implying the truth of a simple affirmation *about it* that employs an intentional verb (or gerund, etc., derived from that verb). (No one such verb is privileged; the same thing can be said in many ways. Also, the implied affirmation may be specified or it may be an unspecified member of a specified class.)

The fact expressed in 'This frying pan wants a handle' can be stated without making any simple affirmation that is about the pan and employs an intentional verb. (It may be that such a verb must be applied to something else, e.g., that someone *wishes* that the pan had a handle.) There are a great many examples not so easily decided. Can the behavior of a paramecium be described fully without implying one or another proposition on the order of 'It seeks *X*'?

The criterion suggested bears comparison to many that have been entertained by Chisholm and others;[4] G. E. M. Anscombe (1965) stated something like it as one of three conditions she thought jointly sufficient for a verb's being intentional. The reason it or equivalent or

[4] William Lycan gives a fairly comprehensive account of what has been done in this area (Marras 1972, ch. 6). My focus on the dialectical situation is adapted from his.

near-equivalent conditions were not held to be sufficient alone is that they do not exclude nonmental examples. Our point exactly.

5. Saussure's Semiology

Another alternative to Peirce's semeiotic was provided by another of his contemporaries, the Swiss linguist Ferdinand de Saussure (1857–1913). Whereas the one implies a semeiotic philosophy of mind, the other is a theory of signs that takes mental functions largely for granted. Peirce was unlikely to have heard of Saussure, whose fame and influence are due to his posthumously published lectures, *Cours de linguistique générale* (1916). That volume was a primary source of the broad intellectual movement, now *passé*, known as structuralism. It was also the source of European semiotics. The latter subsequently discovered Peirce, whom it claimed, contrary to the usual conventions of paternity, as a second father. That, unsurprisingly, has generated much confusion. The main task of this section is to establish that Saussure's view is fundamentally different from and incompatible with Peirce's.

Saussure distinguished *langage* and *parole*, or speech, from a particular tongue, *langue*, which he defined as a system of linguistic rules employed in speaking. And he argued that, in a scientific approach, the study of the former must be based on a prior and independent study of the latter. For it is the rules of the language used that give acts of speech their meaning and thereby explains their occurrence. He also distinguished the diachronic study of the development of languages over time from the synchronic study of a given *langue*. Again, it is the understanding of a *langue* that is prior, scientifically, to the study of its evolution. Or so, at least, Saussure argued. That was the inspiration of structuralism: the idea that the multitudinous, concrete, historical facts of human existence can be explained on the basis of an abstract representation – precise, complete, certain – of an underlying structure that is largely unknown to those whose actions it determines. Marxism and Freudianism lent themselves to that interpretation – the underlying structures being material or psychic, respectively – thus swelling the structuralist movement, for awhile.

Saussure suggested that his approach to linguistics might be generalized into a study of all the sign systems embedded in social behavior: customs, gestures, modes of artistic representation, and so on. Here, too, it is a system of rules, yet to be articulated, that makes particular behavior possible by giving it its meaning. Saussure named this broader study 'semiology', after the Greek word for sign, *sēmeîon*.

There are two assumptions in semiology that Saussure carried over from his linguistics. One is that a word or other sign is a two-part entity, consisting of a material signifier (*signifiant*) coupled with a signification (*signifié*). Since the object studied is *langue* not *parole*, the material signifier is not a particular sound but is a sound pattern; ergo, it is not itself material but is, rather, an abstraction. For the same reason, Saussure took the signification to be a concept rather than a particular thing that satisfies the concept. Thus, 'tree' signifies tree in general and smiling signifies amiability. The other assumption is that the relation between these two parts is essentially arbitrary. Saussure sometimes wrote of arbitrariness as a matter of convention but at other times implied the opposite, since the growth of language is an unconscious evolution and since linguistic rules cannot be changed by fiat. In any case, he argued that onomatopoeia and the like play inessential and indeed marginal parts in language.

It might seem, at first, that for Saussure, linguistic meaning does not derive from thought. He wrote, 'No ideas are established in advance, and nothing is distinct, before the introduction of linguistic structure'. For, 'setting aside its expression in words, our thought is simply a vague, shapeless mass'. An idea, he said, 'is fixed in sound' and only thereby does the sound becomes a sign of the idea (1983, pp. 110–11; pp. 155–7 in the pagination, now standard, of the 2nd French ed.). But we shall see below that Saussure does assume thought's intentionality, independent of language.

Saussure also held that human vocalization does not fall into distinct, repeatable units before being matched with thoughts (pp. 111–20; 157–69). It is their subordination to forming meaningful words that gives sound units (let us follow current usage in calling them 'phonemes') definition: a range of sounds count as variants of the same phoneme when the substitution of one for another forms the same parts of the same words, where a word is an expression of fixed meaning. The pairing of sound and sense establishes a system of mutual exclusions on the sound level and another system of mutual exclusions on the conceptual level. A range of sounds is one phoneme by virtue of being distinguished from those ranges that constitute other phonemes, and a range of meaning is one concept by virtue of its not being other concepts. (Thus, in French, *vague* includes the vacant, which the English concept of vagueness does not include, as vacancy is reserved, in English, for another word.) It is this paradoxical suggestion, that meaning is created out of two systems of pure differences arbitrarily paired, which accounts for so much of the excitement that Saussure has generated.

But despite Saussure's insights and subtlety, there is also in his theory a surprising vagueness or, perhaps, vacancy, where Peirce had much to say. Saussure does not explain, or even attempt to explain, how thought and language manage to be about the world. He did not deny that there is a world independent of language and thought about which we think and speak. But he did not account for its being an object of representation. He simply assumed that the undifferentiated mass of preverbal thought is already directed toward the world and that language serves only to carve that mass into discrete units. It is the mental concept of tree that determines, for Saussure, what the word 'tree' refers to. Saussure is therefore in the tradition of Aristotle and Locke in making the intentionality of speech to be dependent on the intentionality of the mind. That is one difference between semiology and Peirce's semeiotic.

In the usual contrast drawn between their theories (e.g., Deledalle 2000, chs. 4 and 9), Saussure is said to have had a dyadic conception of a sign, as consisting of a signifier and what is signified, whereas Peirce had a triadic conception, by the addition of an 'interpretant'. The interpretant is a response to the sign that the sign elicits and in which that sign is taken to be a sign of an object: it is this that accords the sign its significance. Peirce did not say so, but presumably an interpretation comprises equivalent interpretants. Thus, if we see smoke and I think, 'Fire!' and you think, in French, 'Feu!' then we interpret the smoke in the same way, as a sign of fire, though our two interpretants are distinct.[5] The concept of the interpretant is central to Peirce's theory of signs, in all periods of its development.

Nevertheless, it is a basic error to suppose that the difference between the two concepts of sign is simply that one is dyadic and the other triadic. Saussure made the sign a dyad, a two-sided entity. Peirce, on the contrary, made the sign just one relatum of a triadic relation, of which the other two relata are the sign's object and the sign's interpretant. All three items are triadic in the sense that none is what it is – a sign, an object, or an interpretant – except by virtue of its relation to the other two. But that does not mean that any of the three is in itself a triad; if the object, the sign, or the interpretant is in itself triadic, that must be for another reason. Peirce's and Saussure's ideas of sign differ, then, not

[5] As there must always be differences of nuance and as these differences will be for some purposes important, it follows that the equivalence of interpretants that defines 'an interpretation' must be less than absolutely strict. What constitutes equivalence will vary with our interest.

only in number of items combined but in the way they are combined. It is the difference between a composite entity and a relational property. This seems persistently to be misunderstood, and therefore, with some reluctance, I shall belabor the point.

Is it not odd to say that a word is not only a pattern of sound but is also a concept? For it is the sound that is made and that is heard when the word is spoken. That is what we call a 'word' in English or a *mot* in French. To be sure, it is only a word because it has a meaning. But what is had is not always a part of that which has it. A man may have a nose, a wife, and a reputation, but only the first of these is one of his parts. Similarly, 'sign' in English and *signe* in French refer to the signifier alone and not to the signified as well. Saying that its meaning is part of a word or that its significance is part of a sign is like saying that a husband is a union of two persons, rather than one person united with another.

Writers who suppose that Peirce held the sign to be a triad or a triadic relation, rather than one relatum of a triadic relation, must then invent a third part or third relatum, in addition to the sign's object and its interpretant. This they usually call a 'sign vehicle'.[6] They then identify a sign either as vehicle plus object plus interpretant or as the relation that binds the three. But there is no basis for this in any Peircean text. The 'sign vehicle' *is* the sign, as Peirce conceived of signs, and the object and interpretant are other things, distinct from the sign.

Another difference often misconstrued concerns the breadth of semiology and semeiotics. Peirce admitted a variety of grounds of significance, including resemblance and causality. Thus his semeiotic embraces natural signs and images, as well as arbitrary signs. Furthermore, he made particular signs as well as types of signs a primary study. And finally, he admitted that, in addition to thoughts, other responses to signs, such as feelings and actions, can be interpretants. Hence, sign interpreters are not necessarily humans only. This breadth appears to be one of its features that has caused semeiotic to be favored over semiology. However, it does not follow that Peirce's semeiotic embraces Saussure's semiology as a part. If Saussure was right that systems of arbitrary signification can be studied in abstraction from their particular uses and in abstraction from natural signs and other nonarbitrary forms of significance, then Peirce was wrong. And if Peirce was right that language can be understood only in the concrete context of its uses, in coöperation with other

[6] The term appears to have been introduced in Charles Morris (1938), not, however, as an interpretation of Peirce's theory.

kinds of signs, then Saussure was wrong. One system cannot be a part of the other, because one contradicts the other.

And that is something about which Saussure and Peirce, had they known of each other's work, would have agreed. For each wanted to make sign theory a science and each regarded any science as necessarily including a taxonomy that, as the stock phrase has it, carves nature at its joints. Those joints are presumed to exist objectively. Thus Saussure wrote,

> A language as a structured system . . . is both a self-contained whole and a principle of classification. As soon as we give linguistic structure pride of place among the facts of language, we introduce a natural order into an aggregate which *lends itself to no other classification.* (Saussure 1983, p. 10; 25, emphasis added)

Similarly, Peirce:

> If the question were simply what we *do* mean by a sign, it might soon be resolved. But that is not the point. We are in the situation of a zoölogist who wants to know what might be the meaning of 'fish' in order to make fishes one of the great classes of vertebrates. (8.332)

Peirce thereby warns us that he will use 'sign' in a technical sense not necessarily identical with its ordinary usage. Its reference will be determined not by usage but by the needs of science: items explained by the same principles are to be grouped together, those by different principles are to be separated.

As each way of carving the beast is meant to be the only objectively correct way, they cannot both be correct. At least one of the two systems of classification must be factually mistaken. Such facts, however, are not easily established. How are we to choose between these rival claims? Not on the basis that Peirce's semeiotic ropes in more phenomena and thus creates a larger empire on which 'semioticians' may plant their standard. Perhaps he has put things together that should be held further apart. No, the question is: which system is the more illuminating? They are to be decided between (unless both are rejected in favor of some third alternative) in the way that any rival scientific theories are (cf. chapter 12, section 3): in minor degree by their internal coherence and conformity to known truths, in major degree by their respective fecundity in prompting further discoveries and yielding additional insights. Peirce sometimes expressed his pragmatism in the words of Jesus, 'By their fruits ye shall know them'. But let us not confuse the fruits of science with mere numbers of publications and conferences.

The failure to see the fundamental opposition of semeiotic to semiology accounts for the unholy union, effected in many contemporary writings, of Peirce's semeiotic breadth with Saussure's semiological arbitrariness, usually misunderstood as conventionalism. The result is an interpretation of natural signs such as medical symptoms and of forms of pictorial mimesis as being constituted by convention, often referred to as 'semiotic codes'. To be sure, convention, or something like convention, has some role to play – a subordinate role – even in these cases, as we acknowledge in due course (chapter 8, sections 2 and 4). But the conclusion that all significance is conventional, as well as being unjustified, is pernicious. It implies (and often has been meant to imply) that the natural world – human nature most of all – is a human construction, subject to revision at whim. That sort of 'postmodern' relativism is the antithesis of Peirce's philosophy.

6. Aristotle, the Stoics, St. Augustine

Among his other accomplishments, Peirce was in his time unusually – perhaps, outside of the German universities, uniquely – well read in the long history, Hellenic, Hellenistic, and medieval, of logic and sign theory. That tradition ended with the modern period, the period defined by its rejection of tradition. Locke, despite his use of a Greek term, possibly made-up, wrote in historical innocence. If others, such as Descartes, were deeply indebted to such predecessors as Saints Anselm and Augustine, that debt was contrary to their profession to have begun afresh. Peirce, so far as he drew on antique doctrines, is an exception among modern thinkers, deliberately. Let us take a quick peep into a couple of cabinets in logic's dusty museum. They bear on Peirce's thought and deserve to be mentioned lest our hero be made to seem even more original than he was.

A distinction between natural signs and verbal reports, as of illness, goes back at least to Hippocrates (Sebeok 1986, p. 39). That was in the period of Socrates, when, as we know from many sources, nature was distinguished from convention, that is to say, man-made law. The Sophists, like today's 'postmoderns', were claiming for convention many things that had been thought natural. Thus it was natural that they thought verbal meaning conventional – a doctrine that Plato subjected to playful examination in his *Cratylus*.

That examination was playful in style only. Since Plato had described thought as inward discourse (e.g., at *Sophist*, 163E), and therefore as

verbal, had he accepted a conventionalist account of linguistic meaning, he would have had to agree that thought itself is a tissue of conventions and has no purchase on truth. Heraclitus was supposed to have denied that language is conventional, and Cratylus was one of his disciples. We do not need to assume that Plato agreed with the lines, mostly ridiculous, he wrote for Cratylus, to the effect that spoken sounds imitate the things they name. There are hints in the *Cratylus* of a different sort of imitation than onomatopoeia, in which it is not individual words but their grammatical and semantic relations to one another that matter, and where the imitated are not perceptible things but are a system of ideal types, the Platonic Forms that perceptible things imperfectly exemplify. The imitation, then, is diagrammatic, wherein relations among words mirror relations among Forms. This is of a piece with the idea of dialectic that we find in many of Plato's dialogues, prominently in the *Sophist* and in Book VI of the *Republic*. According to it, the discovery (or 'recollection') of truth is through the verbal art of building a coherent system of distinctions and connections among ideas. The test of coherence is the deduction of consequences: inconsistent conclusions mark incoherence.

Aristotle, as we saw, had a different view. But it should not be supposed that his conventionalist account of word-meaning was a reversion to sophistry. The very point of his insisting that words symbolize thoughts (which are 'the same for all' and are 'likenesses' of things) would seem to be to counter what would otherwise be the implication of conventionalism in linguistics. However, language could not be an organon of discovery for Aristotle, as it had been for Plato; hence his empiricism, his argument that concepts are formed by sorting experiences and abstracting from them. That is as may be; Aristotle's influence on the subsequent history of sign theory was remarkable in another respect.

When he spoke of words as symbols (*symbola*, in *De Int.* 1) and said that a name is a spoken sound significant (*sēmantikon*) by convention (*De Int.* 2), Aristotle did not define either *symbolon* or *sēmeion*. In another work, the *Prior Analytics*, when he did define *sēmeion* or 'sign', it was in a way that excluded words: 'anything such that when it is another is, or when it has come into being the other has come into being before or after, is a sign of the other's being' (*An. Pr.* II, 27:70a7, Jenkinson trans.). This implies that signs are either effects that signify their causes or causes that signify their effects. The passage occurs as part of a definition of a type of syllogism (cf. *Rhetoric*, I, 2): it assimilated significance to inference. To say that *X* signifies *Y* is to say that *Y* is inferable from *X*. And thus the theory of signs was made to fall under the logical study of inference.

That appears to have been decisive for the first development of a formal theory of signs, by the Stoics of the Early Stoa, third century B.C., and it had continuing influence through to the end of the Middle Ages (Markus 1957, pp. 60–4). Prior to St. Augustine, only the Epicureans and, following them, Sextus Empiricus took exception to the equation of significance with inference. The latter criticized the Stoics for being unable to account for the nonintellectual apprehension of signs. 'For, in fact, the dog, when he tracks a beast by its footprints, is interpreting signs; but he does not therefore derive an impression of the judgment "if this is a footprint, a beast is here"' (*Adv. Math.* 8. 271, Bury trans.).

The Aristotelian/Stoic theory tailored the concept of a sign to the analysis of natural signs; for words are not related to their objects inferentially. However, it is not clear what the early Stoics' intention was. Our knowledge of them is through later reports by Diogenes Laërtius and Sextus Empiricus. The latter cited a word, the name, 'Dion', when he wrote that the Stoics distinguished among 'the thing signifying and the thing signified and the thing existing' (*Adv. Math.* 8. 11–12). The first and the last are corporeal, for example, the sound, 'Dion' and the man, Dion, but the 'thing signified' exists, they said, 'in dependence on our intellect', and thus it is not corporeal.

The incorporeality of the signified was a difficult point for the Stoics, who were materialists. They gave the 'thing signified' the technical name *lekton*, from the verb *legein*, 'to say', and accorded it a dependent, shadowy sort of existence. It is the said or the sayable. That *lekta* must mediate between words and things is a direct consequence of Aristotle's doctrine that words are symbols of thoughts through which they signify things. Sextus' major attack on the doctrine was not ontological but epistemological: according to it, he said, it is not Dion himself, but only Dion *as* he is thought of, which the word 'Dion' signifies. And from this, he said, it follows that we can never apprehend the real man.

This has been a spotty tour, as promised. But all of the ideas and problems cited here are reflected in Peirce's writings on signs. There is reason to believe that he had all of these antecedents in mind, though his references to them are sparse. However, he made no reference that I know of to the most notable anticipation of his views, which occurred in a brief though influential passage in St. Augustine's *De doctrina Christiana*.

As foundation for his theory of scriptural interpretation, wherein meaning is not limited to the literal, Augustine brought words under the general category of signs, after having declared that signs are that

from which we learn about things.[7] Thus he needed a definition of 'sign', *signum*, that applied equally to words and to signs normally so-called. A sign, he said, is 'a thing that causes us to think of something beyond the impression that the thing itself makes upon the senses' (II.1: Augustine 1958, p. 34). This formula was new; it succeeded in encompassing both words and natural signs. And, like Peirce's conception of a sign, it is triadic. A sign is anything that causes one to think (such a thought is an interpretant) of something else (the sign's object).

Because it brought words and natural signs under one analytic category (earlier references to words as signs, from Aristotle onward, were casual), this statement has been hailed as inaugurating a new science: 'It was Augustine who first proposed a "general semiotics"' (Eco et al. 1986, p. 65). But that claim is wildly hyperbolic. Augustine made no such proposal. His definition of *signum* was wholly subordinate to his theological intent; he had no interest in founding a secular science. That the definition became 'classical throughout the Middle Ages' (Markus 1957, p. 71), is a fact apart from its author's intentions. Too much can also be made of the definition's triadicity, which was not a feature of it that drew attention before Peirce's writings on signs began to be published, posthumously, in the 1930s.

The more startling anticipation of Peirce's semeiotic occurs, rather, in the next two paragraphs of Augustine's text, in his division of *signa* into *signa naturalia* and *signa data*, where the latter is not limited to human languages or to conventional signs. Augustine's concept of *signa data* has been misconstrued by recent commentators, who have apparently been blinded, by the early Greek distinction between nature and convention, to his clear meaning. They have taken it as axiomatic that language differs from natural signs by being conventional in meaning, and they have read that idea, contrary to all the evidence, into Augustine's broader distinction. *Signum datum* has been regularly mistranslated as 'conventional sign'.[8] That translation is not justified by Latin usage nor by Augustine's definition of *signum datum* nor by the examples he gives.

Augustine called the crowing of cocks and the cooing of doves *signa data*, but surely he did not think that their crowing and cooing signifies by convention. He said that *signa naturalia* involve no intention or desire

[7] I.2 'things are learned by signs' (Robertson's trans., Augustine 1958, p. 8); but Augustine's view is much subtler than that statement would seem to indicate – see the extended dialogue on this topic in his *De magistro*.

[8] E.g., by Robertson in Augustine 1958, but so also in other translations I have found; a French translation of 1873, *artificiel*, is only marginally better – it is still a mistranslation.

of signifying, while *signa data* are those signs 'which living creatures show to one another for the purpose of conveying . . . the motion of their spirits or something which they have sensed or understood' (II.2.3: Augustine 1958, pp. 34–5). (Clearly, his definition of *signum* in terms of causing a thought has, then, to be generalized: feelings and actions may also be interpretants.) In the case of the lower animals, we think here of courtship displays such as the dance of a male grouse ('the motion of their spirits'), alarms sounded (of something sensed), and the like. Signs thus given – and 'signs given' is the literal translation of *signa data* – are given in order to signify. They are *given to* some other creature or creatures, actual or anticipated, for the sake of eliciting a specific type of response.

The 'type/token' distinction, derived from Peirce but now standard in analytic philosophy, applies not to words alone but to *signa data* generally. Each *signum datum* is either a replicable type or an individual replication (token) of such a type. In all cases, and not in language only, a token is formed in order to signify and the type has come into being in order to be so used. The existence of creatures who will respond to tokens of a given type in specific ways is presupposed. In fact, their potential responses determine the meanings that *signa data* have. For there is no other relation those signs have to their objects. The connection of leaps and flutterings to virility is arbitrary; nature could have fitted the male grouse with a swagger and mustaches to twirl, instead, but it didn't. The meaning of his dance is all in how the female is programmed to respond.[9] Replicable types come about in time, in the development of two sets of correlative dispositions, to replicate and to respond in specific ways to replicas. Those dispositions can be conventional, customary, habitual (think of how a horse is trained to understand certain commands), or instinctual. In the case of human language, dispositions to replicate and to respond (often in further replications) exist in the same individuals. We are both speakers and auditors, and thus can discourse with ourselves.

In Augustine's time as in ours, nature was ambiguously conceived of as one part of any of various oppositions. Four centuries earlier, in *De rerum natura*, Lucretius opposed nature to purpose. That would seem to be the opposition that Augustine had in mind when he said that *signa naturalia* involve no intention to signify. Smoke does not occur in order to draw attention to fire, but animal signals are meant to elicit specific

[9] I do not mean to deny that relative vigor in dancing correlates with vigor in other departments: thus a biological purpose is served by male competition for female favor. Yet the particular form of this competition is arbitrary.

sorts of response. Augustine thus redrew the fundamental division of signs: instead of being between the natural and the conventional, it was between the natural and the purposeful, with conventional signs being a subset of those signs that occur for a purpose. Peirce's idea of 'legisigns' and their replicas (chapter 8, section 1) is exactly Augustine's of *signa data*, and it applies to the same range of examples. Peirce, however, did not conceive of nature as that to which the purposeful is opposed.

2

The Development of Peirce's Semeiotic

This chapter serves several purposes. The first is to show that Peirce's 1868–9 doctrine of thought-signs was deeply flawed, that its flaws were apparent to him, and that he corrected them in divers steps over many years, the process not being completed until 1907. Although similar in form, his mature semeiotic is different in conception from his early semeiotic. Prevailing interpretations of Peirce's semeiotic have resulted from inattention to, or in some cases from heroic denial of, contradictions among writings of different dates.

A second purpose is to demonstrate that Peirce's semeiotic was developed with an eye to theories of knowledge and of mind; it was intended to be a department of philosophy, not a general science of human culture, à la Saussure's semiology. Its context, initially, was Kantian, and the problem that most exercised Peirce was posed by Kant's doctrine that objects not constituted by thought (i.e., not a function of or determined by thought's essential structure) are unknowable. Within the world constituted by thought, we can distinguish the mental and the physical; hence, both are knowable. But behind them, Kant said, there are 'things in themselves', inaccessible to science. Throughout his career, Peirce consistently rejected Kant's claim that there is an unknowable. However, there are two ways of doing that: one is to deny that there is anything not constituted by thought; the other is to assert that we can know things not constituted by thought. Peirce began by adopting the first alternative, which he called 'idealist', but he appears never

to have been satisfied by it. He gradually worked toward the opposite view.[1]

A third purpose of this chapter is to show that the struggle against idealism was the motor that drove the subsequent development of Peirce's philosophy, including his semeiotic.[2] Pragmatism was born of that struggle and helped to bring about the changes in his theory of signs that he made from 1881 to 1907. A fourth and final purpose of the chapter, then, is to elucidate the intimate connections between Peirce's pragmatism and his semeiotic.

1. 1865–1866: Thoughts as Representations

Peirce's earliest remarks about signs, or, as he then said, 'representations', occurred in his twenties. In 1865, at age twenty-five, he denied that he used 'representation' as a translation of the term *Vorstellung*, which Kant used as generic for any mental content; this, on the ground that, unlike Kant, he did not limit his term to mental contents (W1:257). The clear implication is that he applied the term to mental contents as well as to other things.

In the same passage, however, he made representation relative to the mind 'which could truly understand it'. Hence, when he concluded, 'Thus our whole world – that which we can comprehend – is a world of representations', that conclusion was more Kantian than it might have seemed. For it still made the comprehensible world relative to mind. But

[1] The transition is obscured by Peirce's later practice of using mentalistic language to designate realities external to (i.e., independent of) individual minds, so far as they are like minds in being lawful or purposeful. Peirce sometimes named this view 'objective idealism', but it is distinct not only from Berkeley's subjective idealism but also from the Kantian idealism of Peirce's early period. See also note 12, below; chapter 5, note 12; and chapter 11, note 2.

[2] Thus I take issue with the two best-known accounts of Peirce's development. Murray Murphey saw Peirce's logical discoveries as *necessitating* the major shifts in his doctrine (1961, p. 3 and passim); I see these discoveries, rather, as having been *welcomed* because they enabled Peirce to solve philosophical problems he had not yet been able to solve. Max Fisch (1986, ch. 10) described Peirce's 'progress' as one from nominalism to realism; but Peirce, I argue, was a realist (with respect to universals) from the beginning, and his progress was toward the recovery of a doctrine that he associated with nominalism, while remaining a realist. That, however, adds up to what Fisch called Peirce's 'three category realism', so we agree about Peirce's destination. (After completing this chapter and having occasion to reread Krausser 1977, I find that he anticipated my view of Peirce's development in a very concise manner: pp. 192–3.)

even without that Kantian or mentalistic gloss, the idea that the world is a world of representations is idealistic in spirit and in recent years has come to be called 'semiotic idealism'.

Vorstellung can be translated, in this context, either as 'presentation' or as 'representation', but the literal meaning of the verb *vorstellen* is 'to place before,' that is, to place something before someone or something – hence, to present it. As 'representation' conveys the suggestion that its object has been presented before, it would seem that 'presentation' is the better translation of Kant's term and that it is also the more accurate term for Peirce to have used. For, very often, thoughts, percepts, and the like are not re-presentations of something presented before (although, as we shall shortly see, Peirce in 1868–9 denied exactly that). However, 'representation' accords better with ordinary English usage, in which 'presentation' is limited to rather formal occasions. And so we shall adopt Peirce's usage and simply train ourselves to ignore its misleading suggestion. What is represented is not necessarily something that had previously been presented.

There is a problem that even 'presentation' faces and that is faced also by 'representation', when those terms are used to characterize thoughts. It is that what is either presented or represented is presented to or represented to someone. But is not a thought the reception, in a person, of such of presentation or representation? Is not thought what happens in a person when something is presented or represented successfully to him? It is, one might say, a person's comprehension of what is represented to him. It would seem, then, that it is a mere confusion to suppose that thoughts and other mental contents are themselves presentations or representations. It is to reverse their true role. We encountered that problem already, in respect to Locke's suggestion that ideas are signs (chapter 1, section 1). It will arise with respect to any attempt to construe thoughts or other mental contents as signs, representations, or presentations.

I suspect that it was this problem that accounted for Peirce's next move, which became his best-known contribution to sign theory. In 1866, after defining a representation as something that stands for something to someone, he wrote that it would be 'a little more precise' to say that it stands to an 'interpretant', a 'mental equivalent' of the representation (W1:466). The interpretant, then, is a second representation of the same thing. Having replaced 'to a person' by 'to an interpretant', Peirce could then make a person's thought a representation – not to the person himself,

as that would make no sense – but to another thought within that same person.[3] A thought may be the comprehension of what something represents, but it is so, Peirce implied, by being a second representation of the same thing. (He did not remark in 1866 that this entails an infinite series of thoughts interpreting thoughts, but that became a major theme in 1868–9.)

This also explains why Peirce spoke of 'interpretants' rather than of 'interpretations'. For he needed something particular to which a given thought is addressed. An interpretation is general, not particular: we may identify it, in Fregean style, as an equivalence class of interpretants or, more in Peirce's manner, as something embodied in or expressed by any of a number of interpretants (which then are, in that respect, equivalent: chapter 1, note 5). 'Interpretation' may of course refer to the process of interpreting, but 'an interpretation' and 'interpretations' can only refer to the products thereof. That product is general but borne by a particular, the interpretant.

Here we have Peirce's unchanging conception of a sign, as being one of three relata of a single, triadic relation. The other two relata are the object and the interpretant. That makes interpretation essential to signhood. Significance is not a direct relation of sign to object; instead, the significance of a sign is to be found in the interpretant. The sign signifies its object only via being so interpreted (or, later, being so interpretable). Significance, therefore, is a triadic relation, wherein, in one respect, the sign mediates between object and interpretant and, in another respect, the interpretant mediates between sign and object. This is the formal scheme that unites his earlier and his later semeiotic.

In the 1866 passage, Peirce spoke of the interpretant as an idea, in one's memory, which the representation 'addresses'. That makes significance to depend on an already existing thought, much as in Aristotle's account of how words signify. Within two years, by 1868, he had made the interpretant 'addressed' to be a thought subsequent to the sign (see below). Significance is past-dependent on the one view, future-directed on the other. That is a major change, foreshadowing pragmatism and characteristic of Peirce's distinctive intellectual cast.

[3] Persons, thus demoted from epistemological primordiality, may be seen as functions of thought interpreting thought. The process requires a body, but a person is not his body. Peirce drew such a conclusion, beginning in 1869. We, however, do not pick up this theme until chapter 11, section 5.

But in this early formulation of the triadic nature of signification, we also have a thesis that Peirce later rejected, that every interpretant is another sign of the same object.

2. 1867: The 'New List'

In May 1867, half a year after the passage last cited was written, Peirce delivered a paper to the American Academy of Arts and Sciences, 'On a New List of Categories' (W2:49–59), which has often been accorded a central, even a 'keystone' place in his philosophy (editor's note, EP1:1). That, I conjecture, is for three reasons: it is the first published statement of his famous, or infamous, list of categories, later named 1stness, 2ndness, and 3rdness; it purports to demonstrate (as Peirce never later did) the necessity of those categories a priori; and it was praised in what appears to be the highest terms by Peirce himself – 'my one contribution to philosophy' – in a letter circa 1905 (8.213).[4]

I suggest, however, that Peirce later abandoned the argument of the piece: beginning in the 1890s, he restated the categories on an entirely different basis (chapter 3) and explicitly forswore any a priori demonstration of their necessity. Around 1894 he wrote that 'each category has to justify itself by an inductive examination which will result in assigning to it only a limited and approximate validity' (1.301). It follows that he must have taken his 'one contribution' to be the list of categories first adumbrated in 1867, and not the argument for the list that he then made.

What was that argument? Kant is not mentioned, yet Peirce's opening sentence leaves no doubt that he was consciously adopting that philosopher's 'transcendental deduction' of the categories:

> This paper is based upon the theory already established, that the function of conceptions is to reduce the manifold of sensuous impressions to unity, and that the validity of a conception consists in the impossibility of reducing the content of consciousness to unity without the introduction of it. (W2:49)

Readers of Kant's first *Critique* will see immediately where this comes from and what theory it is that Peirce thought 'already established'. Kant proposed a table of 'logical functions of judgment'; it is supposed that each simplest judgment must have one of the combinations of logical forms

[4] Cf. 1.561, ca. 1905, and 2.340, ca. 1895. There are several other places as well where Peirce claimed that the 'New List' was his greatest achievement.

represented in this table. From this table, he derived a corresponding table of metaphysical categories. The idea is that to make any judgment at all, one must apprehend the sensuous manifold under a concept that corresponds to that form of judgment. Such concepts are the categories. To this 'metaphysical deduction', the transcendental deduction adds that, because there is no other way of apprehending anything than through judgment, then the metaphysical categories necessarily apply to the world insofar as it can be experienced, known, and understood. Thus those categories are justified.

Peirce, like many others, found Kant's table of logical functions of judgment haphazard and his metaphysical deduction unpersuasive. But otherwise, in this paper of 1867, he accepted the general idea of a metaphysical deduction, together with a Kantian transcendental deduction justifying the categories metaphysically deduced. He merely provided a different logical analysis with which the metaphysical deduction is to begin. But that analysis differed from Kant's not only in detail but in strategy: it described a series of five stages through which the sensuous manifold is reduced to the unity of a judgment. The three intermediate stages anticipate his later list of three relational categories. At this time, he did not question the adequacy of the Aristotelian logic Kant had assumed, though his analysis, as it emphasized relations, did point toward a logic of relations – one of the major advances in modern logic over Aristotle's, to which he was soon to contribute.

For our purposes, we do not need to examine Peirce's highly condensed and obscure account of judgment formation. Contrary to the importance so often accorded it, the 'New List' is a stepping stone, not a keystone. Furthermore, it was a stepping stone for Peirce, not for us. It is not required for mastery of his later thought. It has, indeed, been a stumbling block to those who have tried to understand his later thought in its terms.[5]

3. 1868–1869: Thought-signs

Peirce's first clear statement of a theory of signs was in 1868–9, in three articles published in the *Journal of Speculative Philosophy*: 'Questions Concerning Certain Faculties Claimed for Man', 'Some Consequences of Four Incapacities', and 'Grounds of Validity of the Laws of Logic' (W2:193–272). Therein, he continued his engagement with the problem

[5] See, e.g., my comments on Murphey and Apel, chapter 3, section 1.

of knowledge, but less dependently on Kant. Descartes was made a foil and an attack was launched against all of modern philosophy, the paternity claim for which is conventionally lodged against Descartes. For that philosopher proposed a clean sweep of the stores of traditional belief and received opinion, allowing nothing that could conceivably come to be doubted or that could not be deduced in careful steps from indubitable premisses. Indubitable to whom? The rock-bottom test of truth, in Descartes' view, is the individual's own consciousness. Thus the two sides of his program: on the one hand, dismantling tradition and social authority and, on the other, erecting an edifice of knowledge on new and supposedly solid foundations in individual consciousness. Peirce opposed both aspects of the Cartesian program – both its radicalism and its foundationalism (not a term Peirce used: see chapter 12, section 1) – an opposition he refined and deepened and expanded throughout the rest of his career. His argument in 1868–9, however, was narrowly focused.

Cartesianism entails not only that we have a capacity for certainty – objective certainty, the impossibility of being mistaken – but also that we have a capacity, equally certain, for distinguishing mental states that enjoy certainty from those that do not. For certainty will do us no good if we do not know, with certainty, when we are certain. Peirce therefore opened his attack on Cartesianism by asking, 'Whether by the simple contemplation of a cognition . . . we are enabled rightly to judge whether that cognition has been determined by a previous cognition or whether it refers immediately to its object' (W2:193). A cognition determined immediately by its object is one, presumably, that cannot be in error. Traditionally, such cognitions are called 'intuitions' (W2:193 and 193n1); hence, intuitive knowledge is certain.

Peirce adduced a number of empirical facts in support of the conclusion that we cannot know intuitively whether a cognition is an intuition (W2:194–9). The foundationalist might see this as deeply wrongheaded; for he maintains that a method has to be established first, before the facts obtained by its means are legitimated. But what if the facts Peirce adduced are ones that the (empirical) methods Descartes eventually justified (on a rationalist basis) would establish? Then Peirce's stratagem works as a *reductio ad absurdum*. He did not point this out. Was he therefore unaware of the real nature of his argument? I think, rather, that he delighted in baroque archness. In the present instance, his use of empirical findings is a way of deriding the foundationalist presumption that a methodology should be established from a position of factual ignorance, real or feigned.

Consistently with the position he was developing, Peirce did not claim to have shown anything with certainty; he suggested, instead, that the consequences of his conclusion, elaborated in the remainder of the essay and in its sequel, add to its support (W2:200). They do so, presumably, by forming a coherent account of human knowledge. And, in fact, Peirce's argument is much less convincing than his conclusions are interesting. Among those conclusions are these: that there is no intuitive knowledge at all, that is, that every cognition is determined by a preceding cognition, and that all thought is in signs. Peirce presented these two conclusions as equivalent, in effect adopting the Aristotelian/Stoic identification of significance with inference (chapter 1, section 6).

That all thought is in signs hardly depends on the narrow train of reasoning Peirce developed in these essays. We have seen that it is a doctrine he held from the beginning of his philosophical studies and that it is rooted in Kant's conception of thoughts as *Vorstellungen*. In notes of 1868, leading up to the three published essays, Peirce traced the idea back to Plato:

> Thought, says Plato, is silent speech of the soul with itself. If this be admitted immense consequences follow; quite unrecognized, I believe, hitherto. . . . From this proposition that every *thought* is a *sign* it follows that every thought must address itself to some other, must determine some other, since that is the essence of a sign. (W2:172–3)

In the last sentence quoted (repeated verbatim in 'Questions' at W2:207), we have the transition to an Augustinian conception of signs, given the doctrine of the 1868–9 essays, that the interpreting thought must be subsequent to the thought interpreted. But it is unclear in what sense one thought 'determines' another.[6]

The 'immense consequences' begin with the proposition that each thought occurs in an infinite sequence of thoughts. If every thought is a sign and every sign determines a subsequent thought, then we have an infinite *progressus*. For, the subsequent thought, being also a sign, must determine yet another thought, and so on, ad infinitum. But Peirce also proclaimed an infinite *regressus* on the ground that every cognition is determined by a preceding cognition. Thus his doctrine of thought-signs: every thought is both a sign and an interpretant. It is a sign that interprets the preceding thought-sign that determines it, and it is a sign

[6] That has remained a question, both for Peirce's exegetes today and for Peirce himself as late as 1909: see chapter 6, section 7, where the question is explored and I propose an answer.

interpreted in the succeeding thought-sign that it determines. Thought stretches infinitely in both directions, toward the past and toward the future.

This doctrine is most fully stated in the second of the three articles, 'Some Consequences of Four Incapacities' (W2:211–42). Therein we learn that it is through its being determined by a prior thought that a thought has a referent or object:

> For what does the thought-sign stand – what does it name – what is its *suppositum*? The outward thing, undoubtedly, when a real outward thing is thought of. But still, as the thought is determined by a previous thought of the same object, it only refers to the thing through denoting this previous thought. (W2:224)

And it is through being interpreted in a subsequent thought that a thought has a meaning or content:

> no present actual thought (which is a mere feeling) has any meaning, any intellectual value; for this lies not in what is actually thought, but in what this thought may be connected with in representation by subsequent thoughts; so that the meaning of a thought is altogether something virtual. (W2:227)

'The cat wants to be fed', think I. To what does that thought refer? It is determined by preceding thoughts, of the cat rubbing itself against my leg, of its meowing plaintively, and those thoughts, in turn, are determined by a growing awareness of auditory and other sensations that direct attention to the cat. What does it mean that the cat wants to be fed? That thought is interpreted in subsequent thoughts: an expectation of continued meowing if it is not fed or a resolution to feed the cat.

The sequence must extend infinitely in both directions; otherwise, we come to an intuition at one end or to a self-explanatory thought at the other end; or so Peirce at that time appears to have thought. Yet it is a fact that thinking takes place in a finite time. Thus Peirce held that thought forms a continuum: as an infinity of real numbers are packed into the finite interval between 0 and 1, so also an infinity of thoughts may be packed into a finite period of consciousness, say, between being distracted by the cat and feeding it.[7]

[7] I do not mean to imply that Peirce identified continuity with the real number system. Contemporary mathematicians accept the real number system as the arithmetical analysis of continuity. The rationale for this is the calculus, on which some additional light was thrown by Cantor's proofs that the rational numbers, taken out of order of magnitude, are countable, whereas the real numbers are uncountable. Peirce read those proofs in 1884. Even then, he argued that the real number system falls short, that the number of points in a continuum exceeds that of any infinite cardinal. In the end, Peirce seems to

There is, however, a problem. If thinking forms a continuum, then there are no individual thoughts:

> All the arguments of Zeno depend on supposing that a *continuum* has ultimate parts. But a *continuum* is precisely that, every part of which has parts, in the same sense. . . . In ordinary and mathematical language, we allow ourselves to speak of such parts – *points* – and whenever we are led into contradiction thereby, we have simply to express ourselves more accurately. . . . (W2:256)

Thus, any given thought, if it is real, is itself a continuum, a movement in time, arbitrarily distinguished from the larger continuum of which it is a part. And any given thought, if it is a fictional point-thought, has no immediate predecessor or successor; for, between any two points in a continuum, there is an infinity of others. But then Peirce's talk of thoughts preceding and following one another in a series is a fiction or, at best, an analytic device not to be taken in every respect literally. How, then, should it be taken?

4. 1859–1877: Nominalism versus Realism

There are further difficulties in this doctrine of 1868–9, to which we turn later (section 5). Separately, there was a tension within Peirce's philosophy at this time, and it is that – and not problems in his semeiotic – that accounts for the development of his thought during his second two decades.

In 'Some Consequences', in one and the same paragraph, Peirce wrote, with apparent symmetry, of thought's two extremes:

> At any moment we are in possession of certain information, that is, of cognitions which have been logically derived by induction and hypothesis from previous cognitions which are less general, less distinct, and of which we have a less lively consciousness. These in their turn have been derived from others still less general, less distinct, and less vivid; and so on back to the ideal first, which is quite singular, and quite out of consciousness. The ideal first is the particular thing-in-itself. It does not exist *as such.* (W2:238)

And later in the same paragraph:

> And what do we mean by the real? It is a conception which we must first have had when we discovered that there was an unreal, an illusion; that is, when we first corrected ourselves. Now the distinction for which alone this fact logically called,

have taken the concept of continuity to be based on the experience of things extended in space and time, and not on mathematical analysis (Potter and Shields 1977; Dauben 1981; but see also Murphey 1961, pp. 119–22; Hilary Putnam's discussion of Peirce's theory of infinitesimals, in RLT: pp. 37–54 and Putnam 1995; and below, chapter 3, section 6).

was between an *ens* relative to private inward determinations, to the negations belonging to idiosyncrasy, and an *ens* such as would stand in the long run. The real, then, is that which, sooner or later, information and reasoning would finally result in, and which is therefore independent of the vagaries of me and you. Thus, the very origin of the conception of reality shows that this conception essentially involves the notion of a COMMUNITY, without definite limits. (W2:239)

At one end, there is the thing-in-itself; at the other, there is reality. Thought proceeds from the one to the other. It looks, then, as if this is one process extending in both directions, identical with the process earlier described as extending from reference to meaning. But meaning and reality are not the same thing, and reference, one hopes, is to things real. Besides, there are two respects in which Peirce's account is asymmetric.

First, the thing-in-itself, which Peirce referred to in various passages (in 'Some Consequences' or in related writings of 1868) as 'outward', 'out of consciousness', and 'external', is external to the series it limits. An external limit of a series is that which the series approaches but never reaches, as, for example, 1 is the limit of the series, 1/2, 3/4, 7/8, ... In this case, the external limit, approached but never reached as we trace each thought-sign back to the thought-sign it interprets, is the supposed object to which the whole series refers; hence, it is that object that is the thing-in-itself. Reality, by contrast, is what 'reasoning would finally result in' – that is, result in a representation of – 'sooner or later'. Such a representation, thus, is not an external limit; it stands within the flow of thought, at least potentially.

Second, the series of thoughts by which reality is apprehended is infinite, or potentially infinite, in time; the 'community', that is, of inquirers, is 'without definite limits'. But given an infinite time, the thoughts that eventuate in a representation of reality can be infinite in number without forming a continuum. If they are continuous, that must be for some other reason. By contrast, as we work back toward the thing-in-itself, we are reconstructing a passage of thought that occupies a finite span of time, since thought, even if it never ends, did definitely begin at some time.

We are forced to conclude that Peirce was not describing one process that is infinite and continuous in both directions. Rather, he was describing two processes: one by which judgment is formed, and the other by which judgments eventuate in knowledge of the truth. More fully, his doctrine appears to be: (a) that there is one continuous process in which a thought-sign occurs, extending both backward (toward reference) and forward (toward meaning), all of which takes place within a finite time

and one thinker, and (b) that there is another process, perhaps not continuous and certainly not limited (unless by unfortunate accident) to a finite period or to one thinker, in which thought-signs approach the truth and actually achieve it piecemeal but never completely. Presumably, the latter process weaves together innumerable processes of the former kind, although how that is possible is not stated.

The first asymmetry, that the thing-in-itself is an external limit,[8] while reality is not, has an important anti-Kantian implication, namely, that the thing-in-itself is unreal. Hence, the object of thought, the object to which our thought-signs refer, is unreal. A further implication is that the individual thing, also, is unreal. For, at this time, as we shall see below, Peirce saw all of cognition as being general, never particular, in content; hence, if reality is as it is eventually represented to be in the course of thought, and only as it is represented to be, then it must be general, and general only. The individual, as something other than or more than the universals true of it, is the thing-in-itself; hence, it is unreal. These are, of course, disastrous implications. It turns out that all of our thought is about – nothing real. And what could be more absurd than to deny that there are individuals? Of what then would universals be true? And what can 'general' mean if there is nothing that is not general? Peirce was aware of these implications, as is evident from his attempt to evade them.

So far as the denial of individual existence is concerned, the attempted evasion consisted in an argument (found in 'Some Consequences' at W2:233; more clearly at W2:180–1, from an 1868 preliminary draft; in an 1870 article on the logic of relatives, at W2:390 and 390n8; and elsewhere), that an individual, such as Philip of Macedon, is itself general. The individual is alleged to be general on two grounds that Peirce did not clearly distinguish. The first is that Philip, notoriously sometimes drunk and sometimes sober, is the same person over an extent of space and time in which his properties vary. The 'absolute individual' or 'singular', by contrast, is fully determinate: each possible predicate is either true or false of it and none is both true and false of it. If none is both true and false of it, that is because it has no parts; for, wherever there are parts, there must be some predicate true of one part and false of the other. The singular, then, is unreal, as it occupies no space or time; it is an infinitesimal point,

[8] This use of the spatial metaphor of externality, for the Kantian thing in itself, should not be confused with an earlier use of it for the physical and for other minds: 'what we commonly call the external world' (W2:213, echoing W2:200–7 from the preceding essay).

'a mere ideal boundary of cognition' (W2:180). The (nonabsolute) individual, by contrast, is real but only because general: 'What does not exist at any time, however short, does not exist at all. All, therefore, that we perceive or think, or that exists, is general' (W2:390n8).

The second ground for asserting that individuals are general is that the individual can be known only through the predicates that are true of it: 'every cognition we are in possession of is a judgment both whose subject and predicate are general terms' (W2:180). (That the subject, e.g., 'Philip', is always a general term of course begs the question; but the argument can be restated without begging the question.) The individual is known only through the properties predicated of it; it is distinguished from its properties, Peirce said, only by being always further determinable, that is, by further predications (W2:180, 398n8).

But how is that which is thus infinitely determinable to be identified? That fact that we can always imagine a further determination, to be added to those already thought, does not itself mean that there is *something* thus *determinable*. The answer is indicated by the former account of the generality of the individual: the individual occupies some extensive space in an extensive time. It is that thing – over there, now – the gray donkey standing by the fence – that can be further described. Thus Peirce defined the individual as that 'which can be but in one place at one time' (W2:233).

The upshot is that the individual that is unreal is the absolute individual, the singular or ideal boundary of cognition. Whereas, the individual that we want to count as real, for example, Philip, is retained; but it is general.

The trouble with this evasion, in either formulation, is two-fold. First, no individual is a universal. As a matter of grammar alone, being Philip of Macedon is not something that could be true of more than one. The same applies to individuals as Peirce defined them: nothing that is at but one place at a time could be said with truth of more than one. (The individual is not said of its parts: my toes, for example, are not me; what is said of them is that they are mine.) To be sure, the individual (the real individual, such as Philip) is general in the sense of being one thing that has many parts; but that does not mean that it has the generality of a universal. It is distinct from all the universals that are true either of it or of its parts. Second, places and times are themselves particulars and cannot be specified wholly by universals. Peirce made this last point himself, in 1885 (see below, section 7).

To make sense of our experience of things as existing in space and time, we need to admit a dimension of experience other than conceptual

thought, a dimension consisting of that which is not at all general. Peirce introduced that dimension later, but in this earlier period, he lacked that resource. And thus he argued for the patently false conclusion that real individuals are general – general in a sense indistinguishable from the generality of predicates or universals.

So far as the denial of the existence of an object signified or thought about is concerned, the attempted evasion is the same. For that object, in Peirce's account, is the individual object. Thus, the claim is that the object of thought is real *so far* as it can be represented by general predicates. It is real as determinable, but it is nothing apart from its general determinations. And that evasion, being the same as the foregoing, fails for the same reason.

The failure of Peirce's attempted evasion of these two, linked problems is manifest in his never being able to set them aside. He persistently bit down on them, as on a tooth that pained him. He did so, for the most part, in terms of the issue between nominalism and realism. Again and again, he denied nominalism, on the ground that it posits an individual existence beyond all our experience of it, hence, on the ground that it consists in a meaningless assertion of an incognizable thing-in-itself. And just as often, he then proceeded to argue that we really do know this very same entity – *so far* as it is experienced and subjected to predications – and, hence, that it is in that way real, after all. So, is it or isn't it?

The discussion is complicated by Peirce's habit of characterizing nominalism and realism, not as their respective medieval proponents themselves understood those doctrines, but in terms of what he, Peirce, understood the arguments for them to entail. Nominalists held that individuals, such as Philip, are alone real; Aristotelian universals, such as man, they said, exist in name only. They did not claim that individuals exist outside of experience and are unknowable; quite the contrary. Nor did they suppose that individuals are infinitely determinate singulars occupying spatio-temporal points. Similarly, the realists did not deny individual existence; they merely insisted that universals are real also. But Peirce supposed that the argument for realism (the underlying argument that he discerned, and not an argument found in his Scholastic sources) denies the existence of anything not general and that the argument for nominalism (again, one he discerned) makes the individual to be the cause, external to experience, of those sensations that make up experience.

Peirce's argument for realism, at this time, was that all of cognition is general in content and that the real is what is knowable (and can only be what is knowable, as the incognizable is inconceivable). In the next

paragraph after the long one from 'Some Consequences' quoted above, he wrote:

> since no cognition of ours is absolutely determinate, generals must have a real existence. Now this scholastic realism is usually set down as a belief in metaphysical fictions. But, in fact, a realist is simply one who knows no more recondite reality than that which is represented in a true representation. Since, therefore, the word 'man' is true of something, that which 'man' means is real. The nominalist . . . believes that there is beneath this a thing in itself, an incognizable reality. His is the metaphysical figment. (W2:239)

Again, in his 1871 review of Fraser's edition of the works of Berkeley, he contrasted the 'nominalist' theory of reality, that it is the 'thing out of the mind, which directly influences sensation', to the 'realist' theory, that it is the object represented in 'the final opinion' toward which inquiry 'gravitates' (W2:467–70).

Peirce was clear on one point: these two theories of reality are not different definitions of 'reality'. Rather, they are two theories about what it is that meets a commonly agreed-on definition. The latter, in Peirce's words, is that 'The real is that which is not whatever we happen to think it, but is unaffected by what we may think of it' (W2:467; cf. W2:239 and 5.430 (1905)). Notice that this makes the real to be independent of particular thoughts but not of thought or its tendency in general; thus it is consistent with idealism. It also makes the real to be independent of what you or I think *about* it, which allows for there to be real thoughts, dreams, mistakes, and so on, that is, things that are not independent of what is *thought*.

Thus, he was able to entertain the possibility that both theories have something of truth in them. While defending the reality of the general and, hence, adopting realism and rejecting nominalism, he nevertheless tried to reconcile (what he took to be) the realists' and nominalists' diverse theories of reality. That is evident in the Berkeley review, though less obviously than it might have been. In later manuscripts, of 1872–3, Peirce explicitly affirmed both theories, for example:

> Here then are two opposite modes of conceiving reality. . . . I do not think the two views are absolutely irreconcilable. . . . The realist view emphasizes particularly the permanence and fixity of reality; the nominalistic view emphasizes its externality. (W3:29)

Notice that Peirce here used 'external' in the sense of being outside of all experience, or not constituted by thought, *and* that he was trying to

retain the view, which he takes to be nominalist, that the external in this sense is the real.

Peirce tried to reconcile the two theories of reality in various ways, but all of them along the line already indicated, that the individual is real *so far* as it is knowable and, hence, *so far* as it is general. But such attempts fail. For, on the accounts offered, the individual is never known as an individual. We have only a series of determinations never completed, and no sense of there being something of which they are determinations – no sense, that is, of which Peirce made sense. Peirce could *say* – he did say – that it is the individual that is thus determinable; but he gave no satisfactory account of how that individual is identified. This failure should be no surprise. It derives from the assumptions that caused the problem in the first place, namely, (a) that representation consists of none but general terms and (b) that each representation is of an object only through its interpreting a preceding representation of that same object. Thus, we never get back to the individual object, except as a posited external limit to the series of thought-signs. While those assumptions reign, every attempt to show the external to be real must fail.

From 1859 to 1877, Peirce tossed on the horns of this dilemma (W1:39–40, 60–1, 152–5, 307, 313, W2:180–1, 209–11, 238–9, 390 and 390n8, 467–70, W3:29–30, 34–5, 44–7, 56–9 (cf. 506–7), 80–1, 235–7). That is a nineteen-year period, taking Peirce into his thirty-eighth year – a long time for a major philosopher and man of genius to be trying to make contradictories agree. Nor can it be stated too strongly that that is what he was trying to do. The passages cited are replete with affirmations of externality and also with denials thereof. Peirce was famous for his feats of ambidexterity, such as simultaneously writing a question with one hand and its answer with the other. In this instance, one hand persisted in snatching away what the other persisted in proffering.

5. Three Flaws in the 1868–1869 Doctrine of Thought-signs

Many commentators assume that Peirce never subsequently abandoned his 1868–9 doctrine of thought-signs. However, that doctrine faced at least three problems, any one of which was sufficient reason for major revision. The problem that compelled the first modification of the doctrine is the one just mentioned, the idealistic implication that thought lacks objects not constituted by thinking. It derives from the doctrines that every thought-sign interprets a preceding sign and that all thought-signs are general.

The other two problems both derive from the doctrine that every thought-sign is interpreted in a subsequent thought-sign. The infinite *progressus* thus engendered is not itself a problem, implausible though it be. Problem the second is that, if a sign's significance depends on its actually being interpreted, then those interpretants cannot be mistaken. As a sign signifies what they say it signifies, significance may be assigned arbitrarily. Problem the third is that if significance depends on interpretants, actual or potential, being signs, then we have no noncircular account of what significance is.

As to the problem of arbitrariness: surely, Peirce did not intend that meaning is arbitrarily assigned, but the question is whether he provided any hint of a basis on which meaning is assigned nonarbitrarily. It might seem as if he meant that a sign's significance consists in how it would be interpreted by one who understands the sign, or that it is the sign itself, because of the meaning it already has, that causes it to be interpreted as it actually is interpreted; in either case, the interpretant, actual or potential, would only reveal the sign's significance and not create it. But that is not what he said. His stated doctrine (quoted earlier) –

> no present actual thought (which is a mere feeling) has any meaning, any intellectual value; for this lies not in what is actually thought, but in what this thought may be connected with in representation by subsequent thoughts; so that the meaning of a thought is altogether something virtual (W2:227)

– entails arbitrariness, be his intentions what they may. To be sure, it was always his view that meaning is not contained in a moment but is future-directed, and that view is expressed in this passage. But this same passage invites, and contains nothing to forestall, the charge of arbitrariness.[9]

As to the problem of failure to explain how thoughts signify: Peirce made significance to depend on interpretation, but then explained interpretation as consisting in signs: X signifies O because it is so interpreted in Y, which, to perform this service, must itself be a sign of O. But, on that account, Y can signify O only because it is so interpreted in yet another sign, Z. And so on, ad infinitum. Thus, the problem of accounting for significance is not solved but is merely handed on, from one sign to the next (Gentry 1952; Alston 1956–7; Short 1981a). Peirce attempted to meet the difficulty:

> It may be objected, that if no thought has any meaning [i.e., in itself, apart from its being interpreted], all thought is without meaning. But this is a fallacy similar

9 Criticism by Joseph Ransdell has caused me to refine my formulation of this point.

to saying, that, if in no one of the successive spaces which a body fills there is room for motion, there is no room for motion throughout the whole. At no one instant in my state of mind is there cognition or representation, but in the relation of my states of mind at different instants there is. (W2:227)

In short, since meaning cannot be located in any thought-sign, it must be found in the very process by which one thought interprets another. Here again we see why Peirce had to insist that every thought is actually interpreted, as it is the process as a whole that gives its parts significance.

That shift, from thing to process and from part to whole, does not explain how the translation of one thought-sign into another produces significance. Merely asserting that it does is not enough. Peirce's assertion that meaning is to be found in the movement of thought rather than in individual thoughts is unsupported by argument. For the same reason, his analogy to physical motion is questionable. Hence, that analogy is of no help whatsoever.

Peirce's failure to account for significance results in the inability of his doctrine to explain how significance can be distinguished from insignificance. The fact that the sentence, 'Possibilities pander to prodigious plentitude', can be translated into other languages or into other English phrases, and those into still others, and so on, ad infinitum, does not prove that it says anything. In fact, it says nothing. And since thought can be formed in such words, it follows that thoughts, like speech, can be nonsensical. But on Peirce's theory, this nonsense, being translatable, is not nonsense. Peirce's theory entails that certain phrases that signify nothing signify something; and, therefore, his theory is mistaken.

Much has been made of the 'translation theory of meaning', and for good reason; but the good reason has been mixed up with bad reasons. The good reason is that this theory blocks the attempt to construe meaning as an entity, as if words were scrip and their meanings are the gold that the scrip can be exchanged for. Meaning does not exist apart from its expressions. Peirce stated this view in various ways at various times (e.g., at W2:53–4, quoted above, and, in 1906 and 1893, 4.6, 127, 132). What the translation theory of meaning does not do, however, is to explain what meaning is. The bad reason for celebrating the theory consists in thinking that it does tell us what meaning is, namely, that it *is* translation, that meaning is achieved only in the endless substitution of one symbol for another. Clearly, meaning is no such thing.

Those are the three problems with Peirce's early semeiotic: it makes the object signified to disappear; it makes significance to be arbitrary; and it fails to tell us what significance is.

6. Derrida *et cie*

Some of the most prominent commentators on or appropriators of Peirce's semeiotic have not seen these three features as being problematic; or perhaps they have, and have embraced them precisely because doing so is paradoxical. Paradox has it charms, especially if one desires *éclat*. Consider Jacques Derrida in his 1967 *De la grammatologie*. I quote his ecstatic prose for the sake of its flavor:

> Peirce goes very far in the direction that I have called the de-construction of the transcendental signified, which, at one time or another, would place a reassuring end to the reference from sign to sign. I have identified logocentrism and the metaphysics of presence as the exigent, powerful, systematic, and irrepressible desire for such a signified. Now Peirce considers the indefiniteness of reference as the criterion that allows us to recognize that we are indeed dealing with a system of signs. . . . The representamen functions only by giving rise to an interpretant that itself becomes a sign and so on to infinity. The self-identity of the signified conceals itself unceasingly and is always on the move. The property of the representamen is to be itself and another. . . . (Derrida 1976, p. 49)

And so on. Derrida identifies this process as 'limitlessness of play' (ibid., p. 50).

The 'irrepressible desire' for a signified – that is, for something real, which signs might signify and which our words might represent or misrepresent – is here presented as a weakness. Only the weak would desire a 'reassuring end' to reference, fleeing the limitless freedom of semiotic play. Derrida later suggests that anyone who would insist that there is a reality beyond play manifests a totalitarian impulse to impose his arbitrary semiotic constructions, tendentiously named 'reality', on others. I would suggest, to the contrary, that the denial of unambiguous reference is a perfect cover for someone fearful of facing reality, and that the idea that there is only play invites totalitarianism. For if there is no reality, then there is no reason why one should not impose his vision on the rest of us: 'One view is as good as another, so I'm going to make you accept mine!' Truth's denial leaves a vacuum: the will to power fills it.

Derrida's ideas reverberate in the writings of Umberto Eco. 'Unlimited semiosis' is Eco's oft-repeated watchword, by which he means the endless process of signs interpreting signs, within which only is there meaning. He, too, attributes the idea to Peirce – not wrongly, but wrongly read into Peirce's later writings. 'Semiosis', Eco says helpfully, 'explains itself by itself; this continual circularity is the normal condition of signification . . .' (1976, p. 71). Equally helpfully: 'language . . . is clarified by successive

systems of conventions that explain each other' (ibid., pp. 68–9). Here we have both Peirce's early failure to explain what significance is and the unhappy implication that significance is imposed arbitrarily.[10]

7. 1877–1885: The First Flaw Corrected

Peirce, in successive steps, eventually corrected each of the three paradoxical features of his early theory. We begin with the paradox of the nonexistent object: the object that disappears into the signs that represent it, the individual that is nothing but universals.

As indicated above, from 1859 to 1877, Peirce struggled to make his idealism yield more than it could. At last, in 1877, a dramatic but little-noticed step was taken. In a manuscript written in the summer of 1877, we read his last restatement of idealism:

> It is usually admitted that there are two classes of mental representations, Immediate Representations or Sensations and Mediate Representations or Conceptions. The former are completely determinate or individual objects of thought; the latter are partially indeterminate or general objects. . . . Occam denies that any general objects of thought exist, which implies that no objects of thought have any resemblances, differences, or relations of any kind. I on the other hand have undertaken to show that just the reverse of this is the case. That no object is individual but that the things the most concrete have still a certain amount of indeterminacy. [The Philip of Macedon example follows.] . . . you do not get down to anything completely determinate till you specify an indivisible instant of time, which is an ideal limit not attained in thought or *in re*.
>
> It follows from this doctrine that we have no pure sensations, but only sensational elements of thought. (W3:235)

Then, in November 1877, Peirce published one of the papers for which he is most famous, 'The Fixation of Belief', and in it he wrote:

> To satisfy our doubts, therefore, it is necessary that a method should be found by which our beliefs may be caused by nothing human, but by some external permanency – by something on which our thinking has no effect. (W3:253)

[10] Eco is supposed to have recanted this view in a lecture of 1989 (Eco 1995), in which he charmingly cites his own earlier view as 'the most outrageous example' of the idea that 'texts can be interpreted in infinite ways' and proceeds to criticize the passage from Derrida I have quoted above. Unfortunately, the only way he can find to evade infinite interpretability is by identifying a text's meaning, 'if not as objective, at least as intersubjective', with a community's 'agreement' as to what it means – which still allows meaning to be arbitrarily imposed and, hence, to be infinitely alterable. For a more detailed criticism of the earlier Eco, see Colapietro 1989, ch. 2.

The use of the word 'external' to mean what is not at all constituted by thought – and yet without any qualification of its being an ideal limit of cognition or an unknowable thing-in-itself – is stunning. Peirce continued,

> Such is the method of science. Its fundamental hypothesis . . . is this: There are real things, whose characters are entirely independent of our opinions about them; those realities affect our senses according to regular laws. . . . (W3:254)

Equally stunning, this is the view of reality Peirce had associated with nominalism, that it is the thing out of consciousness – that which he had just called 'external' – that causes our sensations. Such causes can only be individual things and events; the sensations caused are equally individual. Consistently with Peirce's idea of nominalism, the existence of such causes is not itself given in sensation; it is spoken of here as a hypothesis.

The nominalistic theory of the real was thus adopted without the qualification made heretofore, that the individual is real only insofar as it is general. It was nevertheless joined with the other theory of reality, that Peirce had associated with realism, which makes the real to be the knowable. For he proceeded, immediately after the words last quoted, to claim that 'by taking advantage of the laws of perception, we can ascertain by reasoning how things really are'.[11] The hypothesis may therefore be to increasing degrees confirmed, by progressive success in filling in its detail. Thus, in 'How to Make Our Ideas Clear', the 1878 sequel to 'Fixation', Peirce wrote:

> Different minds may set out with the most antagonistic views, but the progress of investigation carries them by a force outside of themselves [Peirce's nominalism] to one and the same conclusion [his realism]. . . . The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth, and the object represented in this opinion is the real. (W3:273)

The external force explains the fated agreement and in that agreement the external is known. The upshot of this marriage of 'nominalist' and 'realist' theories of reality is that the external is knowable. Sometime in late summer or early autumn 1877, Peirce had ceased to be an idealist of the sort he initially was.[12]

[11] Of course, the laws of perception have themselves to be discovered by empirical investigation. All empirical inquiry proceeds on the basis of theories already accepted. How objectivity is nonetheless possible is explained in chapter 12.

[12] As he continued to maintain that the real is knowable and that the knowable, even when an individual, has general features and thus is of the nature of the mind, Peirce was

To maintain the view that the individual cause of sensation can be an object of knowledge, Peirce had to modify his doctrines that every sign is general and that each sign interprets a prior sign. That is, he had to identify a class of signs that signify individual objects directly, without the mediation of rules or other general relations. From early on, he had mentioned such signs, designated either as 'signs' or as 'indices', and given examples: proper names, pointings by fingers or weathervanes, and so on. The trouble was that he did not conceive of these as designating their objects without the mediation of general ideas.[13]

From 1879 through 1884, Peirce wrote little in a philosophic vein, his time being taken up by scientific work. With his appointment in 1879 as a part-time Lecturer in Logic at the newly formed Johns Hopkins University, however, he perforce turned a portion of his attention back to formal logic. In a manuscript of 1881, he wrote that 'it is an important theorem of logic that no proposition whatever can be completely and fully expressed in general terms alone' (W4:249). For, as he proceeded to explain, that about which a proposition is, whether it is particular or general and whether it is real or fictional, has somehow to be designated as belonging to *this* world or *this* fiction, and so on, and that 'cannot be expressed by any general description' (W4:249–50). That is a flat-out contradiction of the basic premiss of his earlier idealism, that all knowledge is wholly general and that the individual can be known *only* through a series, never completed, of general determinations.

In logical papers of this period Peirce and his student, O. H. Mitchell, independently of Frege, introduced quantifiers into predicate logic. Mitchell's paper, 'On a New Algebra of Logic', was published in an 1883 book Peirce edited, *Studies in Logic*, of papers by his students and himself. Two years later, in a published paper, he wrote:

> generality is essential to reasoning. . . . But [general terms] alone do not state what is the subject of discourse; and this can, in fact, not be described in general

still able to describe his philosophy as idealistic – e.g., in 1907 he called it 'conditional idealism' (5.494) – though he rarely did so except in the early 1890s when, mostly in relation to his cosmology, he called it an 'objective idealism' (6.24–5,102, 163): see chapter 5, note 12.

[13] The 1867 ('New List') definition of an index, as consisting in a 'correspondence in fact' (W2:56), was not only cryptic, but, as Murray Murphey pointed out, since the 'New List' based cognition wholly on general relations, it reduced indices to nothing more than an 'it', a 'concept of the present, in general' (Murphey 1961, pp. 298–300). Murphey's entire chapter 15 should be studied for its penetrating account of the present topic. I believe that Murphey is right about Peirce's early conception of the index, despite Peirce himself never admitting the change: see below, notes 14 and 15.

terms; it can only be indicated. The actual world cannot be distinguished from a world of imagination by any description. Hence the need of pronouns and indices. . . . The introduction of indices into the algebra of logic is the greatest merit of Mr. Mitchell's system. (W5:163–4)

A recurring variable functions as a relative pronoun, its quantification with reference to a domain of individuals is indexical, like a demonstrative pronoun. It is possible, then, to attribute the new conception[14] of the index to the introduction of quantifiers to formal logic. But Peirce needed the conception in any case and would probably have come to it one way or another. He was close to it in 1881.

That conception must make indexical reference to depend on relations that can obtain only between pairs of particulars, such that, by it, the one particular may serve to pick out or indicate the other. In the manuscript of 1881, he wrote that all words are conventional and, thus, general, and yet that some of them may be employed in relation to an 'avenue of sense' or other 'real relation' to particular referents. And

so far as they refer us to some living experience or to something with which we have been made familiar by its action on us or ours on it, they signify their objects, not by virtue of habitual association merely, but by the force of a real causal connection. . . . "Here," "now," "this," are rather like finger-pointings which forcibly direct the mind to the object denoted. (W4:250)

And in 1885:

The index asserts nothing; it only says "There!" It takes hold of our eyes, as it were, and forcibly directs them to a particular object, and there it stops. Demonstrative and relative pronouns are nearly pure indices, because they denote things without describing them. . . . (W5:163)

Of course, Peirce was not claiming that indexical reference is a form of knowledge. If it were, it would be intuitive and infallible. Knowledge requires description of its objects, hence, general terms, and, with the introduction of these, we get conjecture and fallibility. The point is only that there must be an indexical component of knowledge, an immediate connection of the particular thought to its particular object, via which general concepts can be predicated of particulars.

This idea of indexicality required a reconception of the category that in the 'New List' was known as 'relation'. That reconception, under the

[14] Earlier in the same paragraph, Peirce cited the 'New List' as the source of the trichotomy of signs of which the index is one (W5:163n1) and did so without mentioning any change of conception. The change is clear, nonetheless.

new rubric, '2ndness', was provided in Peirce's phenomenological reinterpretation of the categories, beginning in 1885 (W5:235 ff.).[15] In it, he identified phenomenological 2ndness as a two-sided awareness of inseparable polarities, for example, force and resistance, action and reaction. To feel oneself acting, one must feel a resistance thereto; to feel a resistance, one must feel oneself exerting a force. Thus we experience something as being other than ('external to') our experience of it; both are particular. This experience is directional, hence, spatio-temporal; self and other are by it experienced as located relative to one another. In experience, 2ndness is itself indexical and accounts for our ability to think about particulars, albeit thinking imports general terms. A particular thus experienced may be an index in turn of some further particular to which it is related dyadically. However, this topic is a large and difficult one; further discussion of 2ndness is deferred to the next chapter, and for a more rounded account of indices, see chapter 8.

The new conceptions of 2ndness and of indices made it possible for Peirce to modify his conception of the individual. It is no longer problematic that the individual be defined as that which is at but one place at a time, since we now understand how places and times may be identified; for indices indicate spatio-temporally. But then the individual may also be defined in another way, as that which can enter into dyadic relations *in re*. Thus, in 1888, Peirce adopted Duns Scotus' term *haecceity* to name that character, which he defined as 'hereness and nowness', Scotus' *hic et nunc* (W6:205, 455).[16] *Haecceity* is irreducibly particular: 'Why IT, independently of its general characters, comes to have any definite place in the world, is not a question to be asked; it is simply an ultimate fact' (W6:205). This *haecceity* Peirce identified as 'pure secondness', a fact 'not calling for and not capable of explanation' (W6:206). *Haecceity* is primary; spatio-temporal location is secondary (W6:455). *Haecceity* or 2ndness is the root of individuality.

[15] In 1.567, part of a ca. 1899 fragment, Peirce wrote of an error in the 'New List' regarding the category of relation; the error appears to be that of not including among dyadic relations those that are 'mutual', i.e., symmetric, and *in re* – exactly those, as of mutual opposition, later made paradigmatic of 2ndness. But indexicality depends on 2ndness. Here we have an implicit admission by Peirce that in the 'New List' his conception of the index was deficient in the way Murphey alleged it was. Joseph Ransdell called this passage to my attention, in this connection.

[16] Murray Murphey (1961, p. 131) and Max Fisch (1986, p. 195) said that Peirce adopted Duns' term in 1890; but that was due to a mistaken dating of the manuscript, 'A Guess at the Riddle': see Nathan Houser's introduction to W6: lxxxi.

But as the poles of a relation of 2ndness are spatio-temporal points, this is not yet a full account of a real individual occupying some extensive place and time. The full account involves also the category of 3rdness, which includes law and continuity. Philip is a continuum of spatio-temporal points whose changes in quality, however abrupt, conform to general laws of his being (he did not become drunk without drinking nor pacific until sober). Here again, a more complete account must be deferred to the next chapter. The main point, however, is that a continuum of reactions or relations of 2ndness reduces neither to the laws thereof nor to the reactions therein; each is an irreducible aspect of individual existence.

8. After 1885: Consequences of the Foregoing

Peirce's discovery of the (truly) indexical sign resulted in a cascade of further discoveries and consequent revisions of his early semeiotic. We may review these advances in rapid order, since they all belong to his mature semeiotic and are developed systematically in chapters 6–9.

a. The discovery of the index enabled Peirce to relinquish the thesis that every cognition must be preceded by a cognition, ad infinitum. A cognition combines indices and concepts and, by its indexical component, it may relate either to a prior cognition or other prior sign of its object – or to its object directly. The earliest indication of Peirce's having grasped this implication, that I have found, occurs in 1901, when he mentioned the 'first' judgments we make concerning percepts (7.198). The long wait may be due to his having written little on relevant topics, over the twenty years from 1881 to 1901. A passage almost identical to his 1901 statement occurs, more prominently, in the 1903 Harvard 'Lectures on Pragmatism', where he described the 'perceptual judgment' as 'the *first* judgment of a person as to what is before his senses' (5.115, emphasis added). This is as clear a rejection of his earlier view – that there is no first cognition but that every cognition interprets a preceding cognition – as one could wish.

It does not follow that such judgments are infallible intuitions, entirely determined by their objects; for the object must still be apprehended conceptually, and the union of concept with index is fallible because conjectural. In these same lectures of 1903, Peirce likened perceptual judgments to hypotheses, from which they differ only in occurring uncontrollably (5.181). They are subject to correction when found not to cohere with other such judgments (Bernstein 1964). Peirce's fallibilism, his

anti-Cartesian denial of intuitive knowledge, was thus preserved even while the doctrine of an infinite regress of judgments, on which it was originally based, was jettisoned. The denial of Cartesianism that was at first idealistic became realistic.

b. The discovery of the index also ushered in a vast extension of semeiotic: to new classes of sign, of interpretant, and of interpreters. For although the index was discovered as playing an essential role within cognition, it is by its nature not limited to cognition. Anything that compels attention or that channels it in a particular direction is an index. Thus, natural effects function as indices and must be added to cognition and language as belonging within semeiotic's purview. That was explicitly and unmistakably stated in 1885 (W5:163) and 1886 (W5:379).

At the same time, it was recognized that anything that signifies by resemblance is an icon (W5:380). Icons, too, occur outside of thought and language; in 1888, Peirce noted that iconic signification is characteristic of the fine arts (W6:212–13). Thus, cognition, art, and natural signs are all brought into one semeiotic framework.

Now, even within cognition, an index is interpreted by an act, namely, an act of attention directed by the index. Thus, any act of attention interprets an index; it does not have to be an act that is a component of a thought. A seasoned driver who brakes a car automatically, without thinking, on seeing a stop light, interprets that light as a command to stop. Hence, interpretants are not always thoughts. In addition to act-interpretants, Peirce implied in 1886 that icons are interpreted in feelings: 'The icon represents its object by virtue of resembling it. It thus depends on simple feeling. Mental association has nothing to do with it' (W5:380). However, it was not until 1904 that he explicitly allowed interpretants to be actions or feelings as well as thoughts (8.332). In 1906, he spoke of a sign as 'determining . . . [an interpreter] to a feeling, to an exertion, or to a Sign, which determination is the Interpretant' (4.536). Later, he called these the emotional, energetic, and logical interpretants, respectively (chapter 7, section 5).

The final extension of semeiotic, to nonhuman interpreters, Peirce hardly mentioned at all. The most famous passage to that effect is in a letter of 1908 to Lady Welby. There, he said that the interpretant is 'an effect upon a person', but then added that this was a 'sop to Cerberus, because I despair of making my own broader conception understood' (LW:80–1). However, if there are signs that can be interpreted nonconceptually, in actions or feelings, then there is no reason why human beings alone may be sign-interpreters. In fact, we ordinarily speak of the lower animals as

responding to a variety of natural signs, of predators or of prey, and also as signaling one another.

The effect of these changes was to transform Peirce's semeiotic into a vehicle for a naturalistic account of the mind as having developed out of more primitive semeiotic processes.

c. Significance is on Peirce's theory triadic. But an icon is related to its object monadically and an index, dyadically. These relations must therefore be distinct from signification, although iconic and indexical significance depend on them. We have therefore to distinguish between significance, which is a triadic, mediated relation of sign to object, and another relation on which significance depends and which may be less than triadic. This 'prior relation' (as I shall call it) was not named by Peirce. However, it is implicit in the account he later gave of the icon/index/symbol trichotomy (chapter 8, section 2). And it is implicit, also, in his elimination of the two remaining paradoxes of the early theory.

9. 1903: The Second Flaw Corrected

What a sign signifies depends on a prior relation and that relation obtains independently of the sign's actually being interpreted. It follows that significance may now be distinguished from actual interpretation. A sign's significance is how it *would be* interpreted on the basis of a prior relation. Significance is therefore a type of possible interpretant, not an actual interpretant. In other words, significance is grounded interpretability.

Significance is no less triadic on this analysis. To be sure, X actually signifies Y even in the absence of any actual interpretant, Z. Nevertheless, to say that X signifies Y is to imply that there is a type of interpretant of X that would be grounded. And Y is not an object (in the sense of being an object of a sign), nor is X a sign, in virtue of their prior relation *simpliciter*. They are properly so named only because that relation grounds a possible interpretant.

To maintain this view, Peirce had to add at least two theories to his armory. One of these is that subjunctive conditionals and counterfactual conditionals represent realities irreducible to actual occurrences. For if there are signs without actual interpretants and significance is nevertheless a triadic relation of interpretability, then that potentiality – interpretability – must be real independently of its being actualized. What would be or would have been must be irreducible to what was, is, and will be. The reality of potentiality is sometimes implied in Peirce's early writings, but on at least one infamous occasion in 1878 it was

denied.[17] In any case, he had no formal account of it before 1885, when he began to reformulate his three categories phenomenologically (W5:235–47). As we see in some detail in the next chapter, the category of 3rdness comprises that of real possibility or potentiality, in connection especially with continua and laws. In 1903, Peirce argued on factual grounds for the reality of law, as implicating, but also as irreducible to, actual events (5.93–107).

The other theory that had to be added is that of final causation. If an interpretation can be grounded, then it must have a purpose; for a ground is something that justifies with respect to a purpose. It follows that interpretation is purposeful and, hence, that significance exists relatively to a possible purpose. But sign interpretation is not limited to human consciousness in Peirce's mature semeiotic. Thus, there must be at least the possibility of purposeful action without consciousness of purpose. That is often named 'final causation'. Here again, there is evidence in Peirce's early writings of a positive view of final causation but no formal theory thereof. And a formal theory is very much needed if one is to adopt an idea so widely disparaged in the modern period. It was not until 1902 that Peirce provided such a theory (chapter 5, section 5).

Peirce did not begin to relinquish the idea that a sign must actually be interpreted until after his phenomenology was well developed and his theory of final causation was stated. In a much-quoted definition of 'sign' for Baldwin's *Dictionary*, he still assumed that every sign requires an actual interpretant. A sign, he said, is

> anything which determines something else (its *interpretant*) to refer to an object to which itself refers (its *object*) in the same way, the interpretant becoming in turn a sign, and so on *ad infinitum*. . . . If the series of successive interpretants comes to an end, the sign is thereby rendered imperfect, at least. (2.303)

That passage, published early in 1902, must have been written the preceding year. In 1902, he wrote similarly but with this qualifying addendum:

> It is not necessary that the Interpretant should actually exist. A being *in futuro* will suffice. (2.92)

Now, if the future consists of what *will* happen, then this marks no advance over the preceding; but it is more likely that by '*in futuro*' Peirce meant what *may* happen. In 1903, he was clearer, claiming that a mere *capacity*

[17] That was when Peirce wrote that there was no difference between a hard thing and a soft one before they are brought to the test (W3:266). In 1905, exhibiting some embarrassment, he corrected this misstep, saying that 'it is the reality of some possibilities that pragmaticism is most concerned to insist upon' (5.453–7).

to determine an interpretant makes something a sign. That capacity for determining an interpretant (presumably of a specific type) must consist in the sign's prior relation to its object. Here is the passage, in which 'representamen'[18] replaces 'sign':

> while no Representamen actually functions as such until it actually determines an Interpretant, yet it becomes a Representamen as soon as it is fully capable of doing this; and its Representative Quality is not necessarily dependent upon its ever actually determining an Interpretant. . . . (2.275).[19]

In another passage of 1903, he wrote similarly, referring to a '*possible* Third Correlate being termed its Interpretant' (2.242, emphasis added). We observe Peirce inching forward.

In yet another passage of 1903, after giving a definition patterned on his 1902 definition, he wrote,

> It follows at once that this relation [of determining an interpretant] cannot consist in any actual event; for in that case there would be another actual event connecting the interpretant to an interpretant of its own of which the same would be true; and thus there would be an endless series of events which could have actually occurred, which is absurd. For the same reason the interpretant cannot be a *definite* individual object. The relation must therefore consist in a *power* of the representamen to determine *some* interpretant to being a representamen of the same object. (1.542, emphasis in original)

Peirce himself thus declared in 1903 that his famous doctrine of 1868–9, reiterated as late as 1902, is 'absurd'. So we have his word for it.

The reason he thought it absurd, to judge by this passage, is that there cannot be an endless series of actual interpretants in a finite time. Apparently, he recognized that a series of actual interpretants must be discontinuous. This argument assumes that the interpretant of a sign must be a sign; for otherwise an infinite *progressus* of interpretants could be forestalled even if every interpretant is actual. Whatever the reason, Peirce at this time identified that interpretant that a sign must have as being one that is merely potential. In 1906 and thereafter, Peirce referred to this potential interpretant as a sign's 'immediate interpretant' (chapter 7, section 2). In 1909: 'The Immediate Interpretant is an abstraction, consisting in a Possibility' and 'My Immediate Interpretant is implied in

[18] In 1903, Peirce used 'representamen' as the generic term, and explained that 'A *Sign* is a Representamen with a mental Interpretant' (2.274). In 1905, in a draft of a letter to Lady Welby, he spoke of 'representamen' as a former usage: 'But there is no need of this horrid long word' (LW:193). It follows that 'sign' was then the generic term, hence, that not all signs require mental interpretants.

[19] In the *Collected Papers*, this passage is dated ca. 1902. The Peirce Edition Project now says 1903.

the fact that each Sign must have its peculiar Interpretability before it gets any Interpreter' (LW:111). To be sure, he continued to insist that every sign has an interpretant, but *that* interpretant is a specific potentiality, a 'peculiar interpretability'.

But then there is something against which to measure any interpretant actually formed. Is it or is it not properly grounded? Does it or does it not actualize an immediate interpretant? Does it attribute a significance to something that, independently of its being interpreted, has that significance? If not, it is a misinterpretation. Thus, misinterpretation is possible.

By this means, the paradox of arbitrariness was eliminated. As an added benefit, the infinite *progressus* of interpretants is eliminated. Signs *may be* interpreted and reinterpreted *ad infinitum*, but their actually being interpreted, even once, is not necessary to their having significance, hence, to their being signs.

10. 1907: The Last Flaw Corrected

There is one last problem to solve, one last paradox to eliminate. In the period just examined, 1902–6, Peirce had yet to say what it is for one thing to signify another; that is, he had yet to do so without making signifying to depend on further signifying, *ad infinitum*. It makes no difference that the interpreting sign is now supposed to be potential rather than actual or that the unending sequence of sign-interpretants is *in posse* not *in esse*. The question is still begged, if it is supposed that a sign's signifying depends on its interpretation by further signs.

After 1904, the doctrine that every interpretant is a sign was no longer standing firm; but neither had it fallen. It was like a decayed mansion that is in the process of tumbling down. Icons and indices were admitted to have interpretants that are not signs, at least sometimes or in part. But Peirce continued for many years to affirm the 1868–9 view that all thoughts must be interpreted, if interpreted fully, in further thought-signs (e.g., at 8.332). It was not until 1907, at age sixty-eight, seven years before his death, that he made the final decisive change in his semeiotic. That was within a manuscript, MS318, that consists of labyrinthine multiple drafts of an unfinished article; it has never been published in its entirety.[20]

[20] The parts published in the *Collected Papers* are in 1.560–2, 5.11–13, and 5.464–96. Other parts are in NEMIII/1:489–94 and EP2:ch. 28. But not all the parts have as yet been published, much less published together.

MS318's topic is cognition, or what Peirce therein named 'intellectual signs' or 'mental signs'; it is a reformulation of his pragmatism. Pragmatism is not a general theory of meaning but pertains only to the meanings of those signs – words, concepts, statements, beliefs – of which cognition consists, and only to such of their meaning as belongs to cognition. The interpretants proper to such signs are neither energetic nor emotional but are what Peirce called 'logical'. Until 1907, he had supposed that a logical interpretant (not so named except in MS318) is always another intellectual sign. But in MS318:

> I do not deny that a concept, or general mental sign, may be a logical interpretant; only, it cannot be the ultimate logical interpretant, precisely because, being a sign, it has itself a logical interpretant. (NEMIII/1:493–4)

Again, in a sentence from MS318 that is still unpublished (except for my having quoted it in an article or two), but which states the fundamental question that the manuscript purports to answer:

> far from holding that a sign can be the "naked," that is, the ultimate meaning of a sign, I was just about to insist that it cannot be so; and the great enigma that leads up to pragmatism – at least to my form of the doctrine – is, "What can this naked or ultimate meaning be?"

That is a stunning reversal of his earlier view. Meaning is not an endless translation of sign into sign. There must always, in every case, be an interpretant that is ultimate in the sense of not being yet another sign.

The ultimate interpretant is one form among others in which the immediate interpretant – the sign's interpretability – can be realized. An English-speaking person not fluent in French reads *chat* and thinks 'cat'. One word is interpreted by another. As it is done so correctly, *chat*'s immediate interpretant is realized, but not in its ultimate form. For we can ask, what does 'cat' mean? Presumably, that is known by our Anglophone reader without further translation. But in what does that terminal understanding consist? That is what the doctrine of the ultimate interpretant is intended to answer.

There is a fundamental difference between ultimate interpretants and final, or ideal, interpretants.[21] In the case of cognition, the final interpretant is the truth, as full as is sought, about the object signified and is itself of the nature of a statement: it is what Peirce elsewhere called 'the final opinion'. But any statement, true or false, final or provisional,

[21] There has been much confusion about the divisions Peirce made among interpretants: those confusions are dispelled in chapter 7.

must be meaningful, hence, it must have an ultimate interpretant. Final interpretants, when verbal, must themselves have ultimate interpretants.

Now, we already know, in a way, what Peirce's pragmatic view of meaning entails. Take his formulation of the pragmatic maxim, in 1878:

> Consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have. Then our conception of these effects is the whole of our conception of the object. (W3:266)

One supposed problem with this maxim is that the effects we conceive an object to have depend on the contexts in which we imagine the object to be placed and on our beliefs about the things forming those contexts; as our stock of germane beliefs grows, so does the list of conceivable effects. There is no end to the effects that we might conceive injustice, say, or electromagnetism to have. But this consequence is not a problem at all, if Peirce's intent was to show that conception is a function of knowledge, that meaning is inexhaustible, and that its explication is never complete (chapter 10, section 6).

There is a quite different problem of endlessness that the 1878 formulation raises. As Peirce himself pointed out in 1906, the emphasis throughout that formulation was on conception, in order, he said, 'to avoid all danger of being understood as attempting to explain a concept by percepts, images, schemata, or anything but concepts' (5.402n3). As concepts are signs, that was consistent with the early semeiotic doctrine that signs are to be interpreted by signs. It is not a problem, if we think of the pragmatic maxim as applying to explication only. For explication is a verbal activity; hence, it must be in terms of words and the concepts they express. But, as Peirce's 1906 gloss itself indicates, more than a maxim of explication is at issue: at least by that date he was looking to 'explain a concept', apparently in the sense of saying what it is or what possessing a concept is. And for that purpose, an infinite *progressus* of signs interpreting signs will not do.

If possessing a concept means having certain expectations about the effects any exemplar of it would have under such-and-so conditions, then that possession is dispositional. Peirce had already associated belief with habits of action (in 1877, W3:247; cf. 5.12). The fundamental change in doctrine that occurred in 1907 was to have recognized that it is the habit itself, and not a concept of it, that is the ultimate interpretant of a concept. Verbal interpretants and verbal definitions, Peirce said in MS318, are 'very inferior to the living definition that grows up in the habit' (NEMIII/I:494). Again,

> The real and living logical conclusion *is* that habit; the verbal formulation merely expresses it. . . . The concept which is a logical interpretant is only imperfectly so. It partakes somewhat of the nature of a verbal definition, and is very inferior to the living definition that grows up in the habit. (5.491)

In 1907, Peirce broke out of the hermetic circle of words interpreting words and thoughts interpreting thoughts. It is only through the medium of purposeful action, even if only a potential action for a possible purpose, that words and thoughts relate to a world beyond themselves and acquire objects of or about which they are.

3

Phaneroscopy

Our systematic exposition of Peirce's mature semeiotic begins, in this and the next two chapters, with its nonsemeiotic foundations. The chronological development of Peirce's thought is behind us; a few historical comments may nevertheless introduce our new approach.

The year 1902 marked a fresh start for Peirce. In that year, he presented an 'architectonic' ordering of the sciences, within which he announced a new science, of phenomenology. His phenomenological reworking of the three intermediate categories of the 'New List' had begun as early as 1885 (W5:235–47) and in succeeding years grew in intensity and refinement. But it was not until 1902 that he described what he was doing as a new form of inquiry, with a distinctive method and subject. This was the moment of a reordering and reconception of his entire philosophical enterprise.[1] In the course of working out this architectonic, Peirce examined, at some length, the idea of final causation. Those discussions, too, were anticipated in various ways in earlier writings, but it was at this point that the idea of final causation assumed explicit central importance in Peirce's philosophy. All in one year, we have a new architectonic, a new science of phenomenology, and a new emphasis on final causation.

The word 'phenomenology' was introduced by J. H. Lambert in 1764 and was used by Kant and then by Hegel. Husserl first used the word in print in the second volume of his 1900–1 *Logische Untersuchungen*, thus preceding Peirce by a year. But it is doubtful that Peirce had read

[1] An earlier stab at an architectonic was made in 1891 (6.7–34), but that ordering of the sciences was entirely different and based on a different principle.

that work. There is no evidence of any influence by either philosopher on the other. When Peirce first used 'phenomenology', he referred to Hegel. That was in 1902; in 1904 he switched to 'phaneroscopy'. It will be convenient to use the term 'phaneroscopy' to distinguish Peirce's inquiry from the versions of phenomenology due to Husserl and his successors.[2]

There are many important differences between Continental phenomenology and Peirce's phaneroscopy, but there is also a point of comparison: like Husserl, Peirce was concerned to examine appearance without judging the reality of what appears.[3] That posed a problem for both philosophers, as both recognized that ordinary language – the descriptive language already available to us – is profoundly realistic. We cannot use it without conveying judgments about what is real. In what language, then, can phenomenological description be carried out? Peirce's solution to the problem of phenomenological description lies at the heart of his phaneroscopy.

That solution presupposes a particular ordering of the sciences. Thus it may explain Peirce's choice of one ordering principle rather than another. We do not have room to consider the 1902 architectonic (1.180–202, 238–82, 5.5–10, 34–40, 120–50) in depth, but a summary of it is necessary.

1. The 1902 Architectonic

The word for and the idea of architectonic derive from Kant, who supposed that knowledge has an architecture that is to be found in an 'affinity of its parts' manifesting 'their derivation from a single supreme and inner end' (*KdrV*, B802, Kemp Smith, trans.). The idea is that the sciences are self-organizing; architectonic ordering is not externally imposed. So also Peirce, though he, unlike Kant, emphasized (a) the social aspect of knowledge's architecture and (b) its evolution over time: 'a great building . . . is

[2] See Spiegelberg 1957 and 1965, pp. 17–19, and Krausser 1977 for discussion of the relations of phaneroscopy and Continental phenomenology.

[3] But whereas Husserl held that intuitive knowledge can be attained of the contents of phenomenological description, Peirce did not exempt phaneroscopic observation from his general denial that intuitive knowledge is possible. Karl-Otto Apel, following Spiegelberg 1957, writes that 'evidently' phaneroscopic description is meant to be intuitive, since it is meant to be presuppositionless (Apel 1981, p. 111). However, the two are not the same: there is still a possibility of error in fitting presuppositionless concepts to the data conceptualized. Moreover, in section 2 I argue that Peirce's phaneroscopic method has presuppositions, that is, makes tentative assumptions tested in its application, even though the descriptive language it uses is presuppositionless.

meant for the whole people, and is erected by the exertions of an army representative of the whole people' (1.176). The affinity of the parts emerges from out of those parts in the course of their development; it is not given to us a priori. A 'supreme and inner end' is not denied, but emerges only gradually.

Also unlike Kant, Peirce claimed that every science, including mathematics and philosophy, is observational (1.238–42). They are to be distinguished by the kinds of observations on which they are severally based. Mathematics is based, Peirce many times argued, on observation of the results of experiments on diagrams (1.240, 3.560); philosophy is based on observations open to anyone to make in the ordinary course of life (1.241); and the special sciences are distinguished from the rest and from one another by the various special means of observation they employ. Peirce said that he borrowed the rule, by which all these sciences are to be ordered, from Auguste Comte: where 'one science depends upon another for fundamental principles, but does not furnish such principles to that other', then the latter is basic to the former (1.180). But he did not explain, at least not at any length or very clearly, what such principles might be.

The architectonic ordering of the sciences, in broadest outline, is: sciences of discovery, first, of review, second, and the practical sciences, third. The sciences of discovery and review are, together, theoretical, as opposed to practical (1.239). They are concerned to establish general truths and are not guides to conduct. The sciences of review summarize and order the results of discovery. Architectonic itself is a science of review (obviously, but see also 1.182, 256; cf. Kent 1987, pp. 48–50). That is a point of fundamental importance. For it shows that the architectonic idea is not foundationalistic: the right ordering of the sciences is not given to us once and for all but grows out of discoveries actually made, in which are spawned new questions, new techniques, new specialties, and new connections among existing sciences. Thus, architectonic, though its usefulness depends on its being relatively enduring, is subject to revision as inquiry proceeds.

The order within the sciences of discovery is: mathematics, philosophy, and the special sciences, which are divided into the physical and the 'psychical or human'. The psychical sciences include both the psychological and the social sciences, also historiography. As observations in mathematics are of diagrams representing mere possibilities, mathematics cannot inform us about what is actual (1.184, 240). Philosophy and the special sciences, by contrast, are positive: they represent the actual world. The

order within philosophy is: phaneroscopy, the normative sciences, and metaphysics. Metaphysics divides in twain, the two parts providing principles, not themselves based on specialist observations, for the physical and psychical sciences, respectively.

The order within the normative sciences is: aesthetics, ethics, and logic. Logic thus depends on ethics and ethics on aesthetics. By 'aesthetics' Peirce did not intend the philosophy of art, but rather a science that attempts to show that certain ends are ultimate, by their appeal to feelings of unqualified admiration (1.191, 280, 573–615, 5.36, 130). Given the ends we conclude are ultimate, ethics then attempts to determine rules for our conduct. Ethics is a science of discovery, and hence is theoretical, not practical; it provides knowledge about how we ought to behave but is not designed to promote good behavior. Logic is the ethics of thinking. It, too, is a theoretical science of what ought to be done, not a handbook for the doing of it. Nevertheless, it is normative and is not a study of pure form: logic in the latter sense, for example, the logic of relations, is part of pure mathematics (1.247). Logic as the third normative science, after aesthetics and ethics, is sometimes named 'semeiotic'. It divides into three parts, of which semeiotic in our sense of the term is the first, logic more narrowly construed as the theory of inference is the second, and methodology is the third. Semeiotic therefore depends, in some sense, on ethics, on aesthetics, on phaneroscopy, and on pure mathematics.

There is nothing in Peirce's idea of architectonic that denies that an architectonically more basic science may have developed after a science to which it is basic, precisely in order to provide it the principles it needs, much as the development of mechanics required the invention of the calculus. The less basic science may provide the point of or the motive for the more basic science; and in other ways, too, the less basic may have to be learned first, at least to a degree, if the more basic is to be comprehensible or more easily comprehended. A less basic science may also in other ways explain or rationalize one that is more basic. An obvious example is that the normative study of the rules of valid inference can be applied to the mathematical reasoning that that study itself presupposes (Peirce frequently distinguished between *logica utens* and *logica docens*).

I have emphasized what architectonic does not proscribe, in order to counter what might otherwise seem a fatal objection to a main theme of this book, that Peirce's mature semeiotic depends on his theory of final causation. For that theory belongs to metaphysics, a science

architectonically posterior to the normative sciences and, hence, to semeiotic. To be sure, one could develop the mature semeiotic without mentioning final causation. But it would be a phenomenological description of sign-interpretation that avoids explanation and thus does not answer such questions as whether consciousness is constituted by interpretative acts, or conversely. For the same reason, there could be no explanation of the phenomenon of significance itself – not, that is, until we had advanced well beyond semeiotic.

The idea of architectonic is itself teleological. That was obvious already from Kant's reference to 'a single supreme and inner end'. But Peirce's notion, that that end is not known a priori but can only emerge in time, requires a different idea of final causation than Kant's. It is noteworthy that most of Peirce's discussions of final causation occur in the context of his architectonic writings. A doctrine that was fated to have a central place in his thought first gained prominence only in one corner of it.

But what explains why he turned to architectonic in the first place, some decades after his initial infatuation with Kant had cooled? I suggest that it was the problem posed by his new science of phaneroscopy – that is, the science he was already deeply engaged in before he conceived of it as a new science. Phaneroscopy itself was called into existence by the epistemological problems detailed in the preceding chapter.[4] The need to show that the hypothesis of an 'external' reality is meaningful required Peirce to expose the experiential roots of the ideas of externality and of the 'would be'. That required, in turn, a new mode of analysis of the elements that constitute experience, and that new analysis – phaneroscopy – could not employ realist conceptions. If it employed such conceptions, whether from physics or from common sense, it would beg the questions it was meant to address. From what, then, was it to draw its descriptive vocabulary? All that is left is pure mathematics, for it alone of the sciences does not purport to represent anything actual. If mathematical ideas can be used to classify appearances, by their purely formal aspects, then that would be description that does not judge the reality of that which appears. Here we have the initial two parts of Peirce's architectonic – mathematics, then phaneroscopy – and also a motive for putting the sciences into an architectonic order.

[4] Apel 1981, pp. 115ff., offers a different account of Peirce's motive for inventing phaneroscopy. That account is more appealing than mine, since it invokes deep questions of philosophical method; but it is not for that reason more accurate.

The part of mathematics Peirce chose for this purpose is the logic of relations, a new study to which he himself had recently made major contributions. (To distinguish this study from logic properly so-called, we might better designate it as the algebra of relations or the formal study of relations.) Why this branch of mathematics and not some other? There is no great mystery here: the three intermediate categories of the five given in the 1867 'New List' were characterized in relational terms, and in writings of the late 1880s and the 1890s (W5:292–308, W6:166–215, 1.354–520), in which we see phaneroscopy beginning to be practiced *avant la lettre*, those same categories were reworked even more markedly in terms of their relational character. When, in 1902, Peirce formally declared that such analysis must take its principle from mathematics, there was no doubt about what part of mathematics would serve. His thought was always thoroughly relational.

Nevertheless, none of that justifies the choice. It does not show that phaneroscopy must rely on the algebra of relations rather than some other part of mathematics. The specific way in which Peirce developed phaneroscopy was without rationale: it was a post hoc systematization of what was already being done. That is why Murray Murphey complains,

> It is impossible to regard Peirce's phenomenological treatment of the categories as anything more than a quite unsuccessful sleight of hand. . . . Now this algebra [of relations] could certainly be used to classify the elements of the phaneron. . . . But what the phenomenology does not show is why it [such a classification] should be made. There are certainly other ways of classifying the elements of the phaneron. (Murphey 1961, p. 368)

Karl-Otto Apel, who also assumes that the categories need to be justified a priori, claims that 'the real basis' for these later developments in Peirce's thought 'is provided by the fact that Peirce in 1867 performed a 'transcendental deduction' of them (Apel 1980, p. 84). But Murphey is surely right that in 1902 and after, 'his method of classifying the sciences has now made it impossible for him to prove that his categories are either necessary or particularly important' (op. cit.).

Peirce's choice of how to proceed with phaneroscopic analysis can be explained but it has no a priori justification. Its explanation is that he knew already what its outcome would be; he knew that the analysis of experience in relational terms was fruitful, that it permits a defense of his realism and in other ways supports a coherent philosophy. That does not preclude other categorial schemes, which might be illuminating in their own way. Peirce's phaneroscopic theses are like hypotheses in natural

science: based on observation, they are nonetheless not fully justified by observations already made but must prove themselves by their fruitfulness for further research, in competition with alternative hypotheses.

To suppose, as Murphey does, that this is a defect, is to fail to realize how much Peirce's philosophy had changed since those days of Kantian inspiration when he wrote his 'New List'. Apel, too, does not recognize the degree to which Peirce embraced a fresh beginning for philosophy, at once empirical and conjectural, in lieu of the Germanic tradition of a priori philosophizing.[5] Peirce rejected a priori philosophizing; his later thought registered an openness to unanticipated experience; in it, there is a sense of adventure.

2. The Phaneron and Phaneroscopic Method

In writings of 1904 and later, Peirce substituted the neologisms 'phaneron' and 'phaneroscopy' for 'phenomenon' and 'phenomenology'. The phaneron is something like what Locke meant by 'idea': it is that which forms the immediate content of awareness. However, Locke and the other British empiricists built a number of assumptions into their conception of ideas, whereas Peirce wished to avoid making any assumptions, so far as that is possible:

> English philosophers have commonly used the word *idea* in a sense approaching to that which I have given to *phaneron*. But in various ways they have restricted the meaning of it too much to cover my conception . . . besides giving a psychological connotation to their word which I am careful to exclude. The fact that they have the habit of saying that "there is no such idea" of this or that, in the very same breath in which they definitely describe the phaneron in question, renders their term fatally inapt for my purpose. (1.285)

The various restrictions alluded to are four in total, as Locke took ideas to be (a) passive and (b) atomistic episodes (c) contained within (d) individual minds. A momentary image, blue and triangular, happening to occur in one's mind simultaneously with a loud clanging, would be an example of a juxtaposition of two Lockean ideas or of a Lockean idea compounded of two.

The assumed passivity of ideas is mitigated by a distinction made by Hume, between ideas impressed on us in sensation or reflection, which are 'forceful' or 'vivacious', and the 'fainter' ideas that are their images

[5] I should like to say, however, that I believe Apel's interpretation of Peirce's philosophy (1980, 1981) to be the most profound and important yet made and, moreover, correct but for its apriorism, its hankering for transcendental deductions.

and that linger in memory or are called up from memory's storehouse and combined in new ways in imagination (*Treatise*, Bk. I, Pt. I, § I). Mitigated but not banished: Hume still assumes that the idea is merely there, in a merely receptive medium, like a thumb-print in clay.

By the assumed atomicity of ideas I mean their occurring in discrete series. Only so, may ideas be individuated; individuation permitted the empiricists to analyze experience in terms of such relations among ideas as resemblance and contiguity. But ideas of process, and of relations that are involved in processes, are then made more difficult to account for. If ideas occur in discrete series, how can we have any idea at all of the continuous passage of time or of a causal process in which one thing, over time, gradually becomes another or makes another?

By the 'psychological connotation' of the empiricist theory of ideas, Peirce meant their assumption that ideas are merely psychological, that they are wholly contained within a consciousness of them. Peirce was careful to exclude that assumption, and its contrary, from phaneroscopy. The phaneron is what appears, as it appears, even if, as it appears, it appears to be more than mere appearance. For example, objects are seen as if they are at a distance from ourselves and as if they exist independently of their being perceived. It is phaneroscopy's business to note such appearances, but not to judge their veracity. It neither assumes nor denies that the contents of consciousness are wholly within consciousness.

The fourth restriction is that ideas belong to individual minds, hence, that no idea in one mind can be identical to any idea in another mind, no matter how much alike those ideas may be (though whether, on this assumption, it makes sense to speak of different persons' ideas as being similar or dissimilar, is questionable). As far as phaneroscopy is concerned, we do not know that that is so. The phaneron may be one thing, albeit the phaneroscopists are many.

In consequence of rejecting these four restrictions, Peirce's conception of the phaneron is extremely vague (as he himself stressed: EP2:362). One cannot even say whether the plural form 'phanerons' is permissible. And if it is, how phanerons might be individuated is unclear: by the persons to whom they appear? by the moments in which they appear to a given person? in some other way? Thus we have this extraordinary statement:

> Phaneroscopy is the description of the *phaneron*; and by the *phaneron* I mean the collective total of all that is in any way or in any sense present to the mind, quite regardless of whether it corresponds to any real thing or not. If you ask *when*, and to *whose* mind, I reply that I leave these questions unanswered. . . . (1.284)

That was in the summer of 1905. A little earlier, Peirce spoke of phanerons in the plural (1.286). But that usage gave way, apparently, to his realization that there is no basis, at least at the outset of phaneroscopic analysis, for distinguishing distinct phanerons. The shift from the plural to a noncommittal singular was anticipated in 1904, before Peirce adopted his new terminology, when he wrote William James, 'My "phenomenon" for which I must invent a new word is very near your "pure experience" but not quite since I do not exclude time and also speak of only *one* "phenomenon"' (8.301).

Phaneroscopy observes, describes, and analyses the phaneron. But even to speak of observing the phaneron, as Peirce did, is perplexing. For the phaneron includes observations. If I observe a wren building its nest, that observation is part of the phaneron. Does it make any sense to speak of observing an observation? Certainly, I cannot peer at it or listen to it, the way I can peer at the wren and listen to its twitterings. However, just as I can notice things about the wren, I can, if I turn my attention to it, perhaps a moment later, notice things about my observing the wren: for example, the way the visual image had a sharp focus and a fuzzy periphery, the fact that I can remember features of which I was not conscious at the time, and that there are differences of kind between the visual image of and judgments made about the wren. Phaneroscopic observation is attending to and noticing things about experience itself, without judging whether the items experienced are real.

Peirce wrote, 'There is nothing quite so directly open to observation as phanerons' (1.286), but also, 'These observations escape the untrained eye precisely because they permeate our whole lives' (1.241). Now, how can one be trained to observe the phaneron? That is the same as to ask, how can the judgments formed in phaneroscopic observation be checked for accuracy? For, in general, one is trained to observe by comparing different observations (judgments[6]) of the same thing. They may be one's own observations or, better, one's own and those of other observers. One judgment is checked against another, and when discrepancies are found, we try again, until consistency is attained. Uncoerced consistency is the

[6] Philosophers imbued with the tradition of British Empiricism tend to think of an observation as something other than a judgment, such as an image, but in ordinary speech, 'an observation' normally refers to a judgment, albeit one formed as an effect of sensation on us. Avoiding misunderstanding, Peirce sometimes referred to observations as 'perceptual judgments' (chapter 12), but in the present context he conforms to ordinary usage, and so do I. Agreement of observations, then, is a logical relation among judgments, not a similarity of images (how could different observers compare their respective images?).

mark of accuracy and competence in observing. It is a fallible mark, as mutually consistent observations may fail to be consistent with additional observations yet to be made. There is no check on observation beyond more observations. The passage from 1.286 continues,

> since I shall have no need of referring to any [phaneron] but those which (or the like of which) are perfectly familiar to everybody, every reader can control the accuracy of what I am going to say about them. (1.286)
>
> The reader, upon his side, must repeat the author's observations for himself, and decide from his own observations whether the author's account of the appearances is correct or not. (1.287)

However, this raises several further questions.[7]

When observations are checked against one another, it is assumed that they are observations of the same thing. Disagreement may be taken to show error, in one judgment or the other, only if it is not instead taken to show that different objects were observed or that no such objects exist at all or that we lack means to observe them accurately, and so on. Agreement, conversely, is evidence of several things at once: that it is the same thing that is being observed in several different observations, that all of those observations are accurate, that the observers have attained competency in making observations of that type, and that such things do exist. In short, agreement of observations not only confirms the observations themselves but is evidence in favor of the assumptions made in making those observations and in comparing them to one another. But all of this depends on an initial, problematic or probational, identification of the objects of distinct observations. There must be some way in which that identification, correct or incorrect, is made.

In the observation of physical bodies, events, or processes, the key is spatio-temporal location, together with assumptions about the way in which things of the kind observed change location. I can look a second time at the place where the wren is, or I can direct your attention to that same place, keeping in mind that wrens hop about, from twig to adjacent twig. What plays the role of location and movement in reidentifying the objects of phaneroscopic observation? How can one even suppose (knowing is not at issue) that it is the same thing that is being observed a second time? And what assumptions are made in this process?

[7] It also shows that Apel is mistaken when he writes that Peirce's phenomenology involves 'a way of knowing . . . that relies . . . solely upon intuition in individual consciousness' (Apel 1981, p. 114).

The answers to these questions are implicit in Peirce's phaneroscopic practice. He bids us, for example, to think of a train's whistle at night or of being poked in the back by a ladder that a man is carrying down the street. Ordinary words are used in an ordinary way to draw one another's attention to objects and events of familiar kinds. There is no difficulty of communication in that, physical reality being assumed. Attention is then shifted from the object to one's experience of it. Once attention is thus redirected, the experience itself can be examined without supposing anything about the reality of what is being experienced. In comparing observations about that experience – or about such experiences – we do not assume a physical world. In Husserl's phrase, such assumptions are 'bracketed'.

However, phaneroscopy is not presuppositionless. In making phaneroscopic observations – a fortiori in a language shared with others – we assume that such experiences have universal features noticeable by anyone and that there are other persons with whom we can communicate. That Peirce was aware that phaneroscopy has such presuppositions, and in that way fits into his general account of the social nature of inquiry, can scarcely be doubted. He wrote nothing to indicate that he felt that phaneroscopy was an exception to what was otherwise his theory of knowledge. Now, these presuppositions do not invalidate phaneroscopy. Just as the presupposition of physical reality is tested in making observations of what we suppose to be physically real, so also the presuppositions of phaneroscopy are tested in making and comparing phaneroscopic observations.

So much for identifying the object of phaneroscopic observation – an identification that remains problematic until agreement about the character of the object is achieved. The remaining problem is to establish a common language in which the phaneron may be described – a language that makes no distinction between appearance and reality. In the next section, we examine the purely formal language Peirce used for this purpose. It is a language in which elements of the phaneron may be distinguished and classified.

One last remark on phaneroscopic method. In identifying elements of the phaneron in purely formal, metaphysically neutral terms, the phaneroscopist is giving content to those forms. In that way, a list of phaneroscopic categories is generated. It is of the essence of this method, however, that the reader cannot understand what is being said if he does not reproduce the phaneroscopic investigation himself:

> [the reader] must actually repeat my observations and experiments for himself, or else I shall more utterly fail to convey my meaning than if I were to discourse of effects of chromatic decoration to a man congenitally blind. (1.286)

The same is true of Husserlian phenomenology and of Descartes' *Meditations*. These are not forms of philosophy that consist in verbal argumentation; rather, words are used to direct the reader to his own experience, and it is that experience, and not the words themselves, which carries the burden.

3. The Language of Phaneroscopy

The purely formal vocabulary that Peirce drew from the algebra of relations is simple. A relation, in any one instance of it, has a certain number of relata, that is, items that it relates in that instance. Its 'adicity' or 'order' is that number of relata. A relation can be monadic or dyadic or triadic, and so on; in general, it is *n*-adic if it has *n* relata.

A monadic relation *R* has but one relatum, *x*, at a time. We can represent this by '*Rx*'. If *R* is being blue, then *Rx* is the fact that *x* is blue. Monadic relations are more usually called properties, qualities, or attributes.

R is dyadic if it has two relata, *x* and *y*, at a time, hence, *Rxy*. '*Rxy*' is to be read, '*x* is *R* to *y*', though if *R* is identified with a specific relation, it may be read idiomatically in various ways: '*x* is larger than *y*', '*x* loves *y*', and so on. The order in which *x* and *y* are written matters in the case of some dyadic relations, not others. If *R* is the relation of being equal to, then, for a given *x* and *y*, *Rxy* if and only if *Ryx*. The relation is, logicians say, symmetric. The same fact is stated with *x* and *y* written in either order. But if *R* is the relation of loving, then we all know from unhappy experience that *Rxy* does not entail *Ryx*. Even when love is returned, the relation itself is not symmetric. In that case, for a given *x* and *y*, *Rxy* and *Ryx* are distinct facts. Notice, also, that the same individual can sometimes be more than one relatum of a given relation. Some people we know love themselves, *Rxx*.

R is triadic if it has three relata at a time, *Rxyz*. An example would be giving: '*Rxyz*' could be used to represent the fact that *x* gives *y* to *z*.

So much for the basic language, which improves on a corresponding strategy in Peirce's 'New List' of 1867 (W3:49–59). In the earlier work, he wrote of an asymmetric act of abstraction – of attending to one thing while neglecting another – named 'prescision' or 'prescinding', wherein

one discovers that, for example, one can attend to space while neglecting color but not conversely. He was then able to build an ordered series of categories, by successive steps of prescision. Prescinding still plays a part in phaneroscopy, but its results are to be confirmed by formal analysis. The question is whether a given relation can be 'reduced' to (defined by) others. That idea needs to be stated with more care than Peirce did, as follows.

We shall say that a predicate '*Rx*' or '*Rxy*', and so on, is 'complete' when substituting names of individuals for '*x*', '*y*', and so on, results in a complete sentence. Thus, '*x* is larger' is incomplete. 'Larger than what?' we want to ask. But '*x* is larger than *y*' is complete; substituting names of individuals for '*x*' and '*y*' forms a complete sentence. Now, the adicity of a relation is the adicity of the predicate, if complete, that represents it. Thus, being larger than is dyadic. However, we can distinguish between a relation as it is represented and as it is in fact. Being a gift is monadic: 'This horse is a gift' is a complete sentence. But nothing is a gift unless there is someone who gives it and someone to whom it is given. So, '*x* is a gift' is elliptical for something more complex that, as it happens, involves a triadic relation: that there exists someone and there exists someone (necessarily someone else?) such that the first gives *x* to the second. Being a gift is indeed monadic; but its reality lies in a distinct, triadic relation.

Notice that such an analysis does not work in reverse. That *x* gives *y* to *z* cannot be analyzed as the conjunction, *x* is a giver and *y* is a gift and *z* is a recipient. For that conjunction could be true even though *x* did not give *y* to *z*. In general, an analysis is a judgment about what constitutes the relation analyzed, whether that judgment is based on concepts merely or also on empirical knowledge, physical theory, religious revelation, or anything else. And that same judgment contradicts any reverse analysis.

An analysis of a relation is an equation of the form 'if and only if' (hereafter 'iff'), where one side is a statement of the relation analyzed and the other side is a statement of one or more relations different from the one analyzed, plus such additional structure, strictly logical, as is required. By 'strictly logical' I mean that it consists of nothing more than the apparatus of deductive logic: sentential connectives, for example, 'and' (hereafter '&') and 'not', and quantification plus variables, for example, 'something' and 'anyone'. Quantification presupposes a range of entities alluded to. For 'some' we shall substitute the so-called existential quantifier, '$\exists x$'. '$\exists$x(. . . x . . .)' is to be read 'there exists (in the specified domain) at least one thing, *x*, such that . . . x . . . '. Thus: *x* is a gift iff $\exists y \exists z$(*y* gives x to z). But not: x gives y to *z* iff *x* is a giver & *y* is a gift & *z* is a recipient.

Consider another example: being a grandfather is elliptical for the conjunction of two dyadic relations: *x* is a grandfather iff $\exists y \exists z$(*x* is the father of *y* & *y* is a parent of *z*). Again, the analysis cannot be reversed, defining parenthood in terms of grandfatherhood. For one thing, it is conceivable that the human race had ended in the contretemps between Cain and Abel, in which case there would have been a pair of lamentable parents and no grandfather.

Peirce tended to use the term 'reduction' only for those analyses in which a relation is shown to be a logical compound of relations of lower order. For example, the triadic relation of benefitting reduces to a logical compound of two dyadic relations: *x* benefits *y* by *z* iff *x* did *z* & *z* benefits *y*. Giving, Peirce argued, is not similarly reducible, since *x*'s laying *y* down and *z*'s picking *y* up does not constitute giving. Giving, unlike benefiting, is irreducibly triadic. Another example of reduction, though problematic, is one that Peirce often cited, though not in these formal terms: *x* is similar to *y* iff $\exists R(Rx \,\&\, Ry)$. Here, quantification is over relations; hence, we are employing second-order predicate logic, and it is arguable that we are making a realist assumption about relations. A nominalist might hold that similarity relations are irreducible or that they are elliptical for triadic relations of comparison.

As our examples show, judgments of reducibility or irreducibility typically take us far beyond the limits of phaneroscopy. Presumably, Peirce gave such examples merely to illustrate the idea. Within phaneroscopy, we are to count a feature of experience as reducible, or not, only so far as either pure mathematics or phaneroscopic analysis itself dictates. Physical explanation and metaphysical argument are not germane. From this point forward, we refer to reducibility in phaneroscopy only. However, that turns out to have metaphysical implications.

A relation that is irreducible is, in the word Peirce sometimes used, indecomposable. It may have a certain complexity, indeed, exactly that complexity represented by the number of its relata. But it cannot be taken apart, as if it were a sum of pieces. It is of the utmost importance to note that what Peirce is engaged in here is not logical analysis. He uses the language of the algebra of relations, applicable in the first instance to predicates, to analyze not predicates or propositions but elements of experience. He is showing that *they* have a structure corresponding to that of the language in which they are described.

Each relatum of a given instance of an irreducible relation may be a relatum of numerous other relations as well. Thus, an irreducible relation, while indecomposable, has an 'external structure', as Peirce

expressed it, by which it may connect to other parts of the phaneron. Peirce, whose formal scientific training was in chemistry, likened this to the theory of the chemical elements (EP2:362–3).

Phaneroscopy identifies the most general features of the phaneron that are irreducibly monadic or dyadic or triadic. The phaneroscopic categories, therefore, are reasonably if unromantically named 1stness, 2ndness, and 3rdness. Something that is irreducibly monadic belongs to the category of 1stness and may be referred to as a 1st; a 2nd is irreducibly dyadic; a 3rd is irreducibly triadic.[8] Peirce claimed that there are 1sts, 2nds, and 3rds, hence, that the categories of 1stness, 2ndness, and 3rdness are not empty. That is, he claimed that there are instances of monadic, of dyadic, and of triadic relations that are irreducible, at least as far as phaneroscopic analysis is concerned.

Why not 4thness? Peirce claimed to have a proof that all relations of higher than triadic order reduce to one or another compound of lower-order relations. But the only arguments to that effect that survive in his manuscripts are brief and merely suggestive; for example, 'that which combines two will by repetition combine any number. Nothing could be simpler; nothing in philosophy is more important' (1.298). It is unlikely that he had a rigorous proof; for contemporary efforts to construct such a proof are complex and depend on technical assumptions (Herzberger 1981; Burch 1991; cf. Kerr-Lawson 1992).

Consistently with the view advanced above, that Peirce's phaneroscopic method needs no a priori justification but will be justified, if at all, by its results, I suggest that the lack of a formal proof of the completeness of its categorial scheme is of no decisive importance (Bernstein states the point well: 1965, p. 70). As we shall see, its three categories illuminate every issue. The burden of proof therefore lies on those who think the categories incomplete. Let them cite an example of irreducible 4thness. Or let them cite a philosophical problem whose solution requires 4thness.

Peirce provided a metaphysical gloss of the three phaneroscopic categories, suggesting that the concepts of possibility, actuality, and law or reality are derived from those of 1stness, 2ndness, and 3rdness, respectively. By his architectonic, that gloss is problematical until a metaphysics is established; yet he did not carefully separate the metaphysical

[8] Peirce sometimes used '1st', '2nd', '3rd' ordinally, to designate relative positions of items within a complex, regardless of their respective categories. We shall avoid that usage. For us, a 1st is always an instance of 1stness, and so on.

interpretation of the phaneroscopic categories from phaneroscopy itself. He moved rapidly from one to the other. And so shall we, though with warning flags flying.

We proceed now to a description of the phaneron in formal, relational terms. I shall make no attempt to reconcile Peirce's various accounts of 1stness, 2ndness, and 3rdness, but will draw most heavily on his later statements (1.284–353, 5.41–92, 7.524–52, 8.264–9, 327–31, EP2:267–72) and less on earlier statements (W5:235–47, 292–308, W6:166–215, 1.354–520). For here, as everywhere else in his writings, we witness thought in flux but also growing in clarity and cogency.

4. 1stness and 2ndness

The whistle of a train in the night derails the train of my thought. We are to distinguish the irreducible parts of this experience, and see what order of relations they answer to. What we can distinguish thus is not claimed to be capable of existing separately or of being experienced separately. It is claimed only that it is an element of experience that does not reduce to the other elements with which it occurs.

I can, on reflection, distinguish the quality of the whistle – as it is heard, not the physical stimulus – from the fact of its occurring here and now, and also from its being produced by a train. Hence I can distinguish it from its relation to anything else. 'Firstness is the mode of being of that which is such as it is, positively and without reference to anything else' (8.328). 'The typical ideas of firstness are qualities of feeling, or mere appearances. The scarlet of your royal liveries' – Peirce was writing to Lady Welby – 'the quality itself, independently of its being perceived or remembered, is an example' (8.329). Any such quality has to be prescinded, Peirce said, from the fact of its occurring or being embodied. For the fact – that the scarlet is in the tunic or that the whistle occurred now, breaking in on my thought – involves a relation to something else.

We call the color seen 'red', but redness, even as a visual quality and not as a physical property, is not the exact red that is seen. For unlike a color seen, redness is a continuum of shades. The color seen belongs to that continuum, and to call it 'red' is to recognize that fact. Conceived of as loud or piercing or a whistle, as bright or red or a color, the quality experienced involves relations of comparison to other qualities and of inclusion in types. Its conceptual apprehension is in terms of such comparisons. But what is compared – the exact quality felt – can be abstracted from the comparison and recognized to be, in itself, without continuity.

'It cannot be articulately thought. . . . every description of it must be false to it' (1.357). That, strictly speaking, is itself false; for assigning unique things to types is what description does, and a description is true when it does it well; Peirce was himself engaged in describing 1sts in that very statement. But his point was that a description does not capture the uniqueness of a 1st and, hence, that the idea of 1stness – the idea of qualities as experienced, not as conceptualized – cannot be conveyed by descriptions alone. Instead, words must be used to direct us to attend to our experience in a certain way: if we obey those instructions, then we will grasp the idea. 'Phenomenology can only tell the reader which way to look and to see what he shall see' (2.197). 'Go out under the blue dome of heaven and look at what is present . . . [as with an] artist's eye' (5.44). 'With an artist's eye': that is, attending to the immediate quality independent of classification and explanation.

An experience that lasts some time may have a great complexity and involve many feelings and yet also have an overall feeling: 'Every operation of the mind, however complex, has its absolutely simple feeling, the emotion of the *tout ensemble*' (1.311). In various places, Peirce gave such examples as 'the quality of the emotion on contemplating a fine mathematical demonstration' (1.304), 'the tragedy of King Lear has . . . its flavor *sui generis*' (1.531), and being 'heartrending' or 'noble' (1.418). Think, for example, of the feeling invoked in us on contemplating an action of noble self-sacrifice: the feeling is simple but the action is complex. There is an echo here of a view he had maintained much earlier, that in an emotion a number of feelings are brought to unity, as if under one predicate (W2:228–9, 1868).

'Secondness is the mode of being of that which is such as it is, with respect to a second but regardless of any third' (8.328). 'The next category that I find, the next simplest feature common to all that comes before the mind, is the element of struggle' (1.322). 'The type of an idea of Secondness is the experience of effort, prescinded from the idea of a purpose. . . . The experience of effort cannot exist without the experience of resistance. Effort is only effort by virtue of its being opposed' (8.330).

We note in these passages a conflation of phaneroscopic description with two other types of study: metaphysics – the definition of 1stness and 2ndness as 'modes of being' – and conceptual analysis – the talk of 'ideas' and the claim, seemingly put forward as a tautology, that 'effort is only effort' in being opposed. Nevertheless, we can ferret out the phaneroscopic content. Whatever our idea of effort may be, what we experience when we experience what we call 'making an effort' is something that

involves equally an experience of something else that we call 'resistance'. By reflecting on this experience, we recognize its two-sidedness, that an experience either of effort or of resistance is equally and at the same time an experience of the other; it is an experience of opposition between the two. And thus it is dyadic.

It is irreducibly dyadic, at least as far as phaneroscopic analysis is concerned. For, being two-sided, this aspect of experience is irreducible to any conjunction of two consciousnesses. It is, Peirce elsewhere said, 'a fact of complexity' not itself complex: 'it is not a compound of two facts. It is a single fact about two objects' (1.526; take 'object' here broadly). Locke also spoke of an idea, that of solidity, which 'arises from the resistance which we find' (*Essay*, Bk. II, ch. IV). However, he did not recognize a formal difference between this idea and others; from his discussion, all ideas would appear to be on a par, as if all were monadic.

Peirce found the dyadic element throughout all experience, even perceptual (perception, then, is not wholly passive):

> There can be no resistance without effort; there can be no effort without resistance. They are only two ways of describing the same experience. It is a double consciousness. . . . The waking state is a consciousness of reaction; and as the consciousness *itself* is two-sided, so it has two varieties; namely, action . . . and perception. . . . (1.324)

Thus, 2ndness is not limited to such experiences as those that we would say involve muscular effort; it also includes that effort of attention that can be opposed and overcome by the insistence of sensation. The train's whistle forced itself on my attention, disrupting my thought. There is no perception without that kind of forcefulness.

Contrary to what Hume seemed to imply, memory and imagination do not lack all forcefulness. Peirce said that *all* of experience is 'a consciousness of reaction'. Perception, I suggest, is like a door forced open against our resistance to it (we were thinking about something else or expected something else), while imagination is like a door that we force open against the flood of current sensation (preferring to think of something other than what we are seeing). Sensation resists our desire to think of something else, but occasionally we overcome that resistance. And memory is sometimes like the one (those memories that rush upon us when we would rather they didn't) and sometimes like the other (the memories that we with some effort deliberately call up). Every quality of feeling that occurs, whether in perception, memory, or imagination, occurs, then, with a degree of force that elbows out its competitors.

That is what constitutes its occurring. Occurrence involves forceful relations of action and reaction, hence, 2ndness. Relations of 2ndness constitute the here-and-now, the *hic et nunc* of what occurs or, if it endures, of what exists. '[E]xistence lies in opposition merely' (1.458). Peirce borrowed Duns Scotus' term *haecceity* to name this aspect of experience (1.405). It follows that the quality of feeling is not in itself an occurrence; it is that which occurs; its occurring is something added to it.

> [By qualities of feeling] I do not mean the sense of actually experiencing these feelings, whether primarily or in any memory or imagination. That is something that involves these qualities as an element of it. But I mean the qualities themselves, as mere may-bes, not necessarily realized. (1.304)

Here we have the quick transition from phaneroscopy to metaphysics. 1sts are possibilities, 2nds, actualities. The transition is misleading, as possibility comprises more than the qualitative 1sts we have so far considered; but more on that below.

Possibility is an element in the actual. Anything actual is some possibility actualized. We never encounter quality except as occurring, yet it is not reducible to its occurring. For, in itself, a quality remains a possibility even when actualized. When we abstract the quality from its occurrence – from its being experienced or remembered or imagined – we abstract the possibility from its actualization.

> A quality is a mere abstract potentiality; and the error of those [nominalistic] schools lies in holding that the potential, or possible, is nothing but what the actual makes it to be. It is the error of holding that the whole alone is something, and its components, however essential to it, are nothing. (1.422)

5. Two Forms of Generality

The identification of 1sts with possibilities poses a problem. Any possibility we can state is general in the sense of being indeterminate; that is because the verbal representation of possibilities is by description, and every description is general. The possible man sitting where you are now may be of any age or weight. A more specific description is still indeterminate: the possible old fat man sitting there also comprises continuous variations albeit in narrower ranges. The possibilities contained in a narrower range, if continuous, are no less multitudinous than those contained in a wider range. A single quality, such as redness, is, we noted earlier, a continuum of shades, and, hence, it is indeterminate with respect to shade. But 1sts, like actualities, are fully determinate: the

qualitative contents of consciousness are of precise hues or tones, and so on. If being indeterminate is what distinguishes the possible from the actual, then 1sts cannot be possibilities.[9]

However, at least by 1888, Peirce did not distinguish the actual by its being determinate but, rather, by its *haecceity* (chapter 2, section 7). Being abstracted from occurrence, 1sts lack *haecceity* and thus are possibilities. We can say that they have a negative sort of generality (1.427), of the maybe (1.304), in the sense that they could occur any number of times, and in different places (however unlikely that may be). But they are not general in the positive sense of comprising variations; they are not general in themselves. The positively general are continua of the negatively general.

But what of the positively general? Does our apprehension of it fall under the category either of 1stness or of 2ndness? It would seem not, as neither 1sts nor 2nds are positively general. Before proceeding further with the question, how positive generality (hereafter, generality *simpliciter*) is apprehended, let us see whether we can secure a firmer conception of what it is.

Peirce claimed that 'True generality is, in fact, nothing but a rudimentary form of true continuity' (6.172); for 'the idea of a general involves the idea of possible variations which no multitude of existing things could exhaust' (5.103; as to 'no multitude', see chapter 2, note 7). A general term, such as 'red' or 'man', is not only predicable of many actual instances but covers a continuum of possible variations: the generality it represents is comprised of those innumerable possibilities.[10] Conversely, 'Continuity is nothing but perfect generality of a law of relationship' (6.172). For any continuum is defined by a law of gradation or of divisibility that applies to all that it contains. Generality, then, entails law. Again conversely, any law entails a continuum, namely, of the possible instances to which it would apply. We recognize this (though the point is controversial; see below, sections 7 and 9) in admitting that any statement of law, whether a law of nature or a law instituted by men, entails subjunctive and counterfactual conditionals: *were* a stone to be released here, it *would* fall; *had* a dozen benighted souls sworn the solemn oaths, our club *would have had* that many more members. Each of generality, continuity, and law implicates the other two, and each entails inexhaustible possibility.

[9] Charles Hartshorne (1952) raised this as an objection; what I write here is by way of an answer.

[10] On the entire, quite complicated subject of Peirce's conception of generality, see the invaluable Boler 1963.

Generality, continuity, law, and unactualized possibilities are found not only in types of quality, types of thing, and laws and rules of all kinds, but also in actual individuals. A stone, star, organism, or person is continuous spatio-temporally: it is undivided though divisible at any number of points (that a body consists of discrete particles does not contradict this, and would not even were they like little marbles at determinate distances from one another; for the spaces between are also parts of the body). Further, such individuals possess dispositions and capacities that endure, unexercised, over some continuous period, sometimes changing gradually in that period. The lump of sugar is soluble at every moment of its existence, though dissolved in none. John is irritable and growing more so. Since dispositions and capacities entail subjunctive and counterfactual conditionals, they are laws, though limited to the behavior of certain individuals.

Much the same applies to social entities. Some, for example, crowds and countries, are spatially defined, though the borders may change over time. Others – presbyteries, chess clubs, business cartels – are not spatially defined, though they may occupy buildings and though their members have bodies. However, they exist continuously for some period of time and their rules apply to any possible member during that time. A club, a board of directors, a church, a tribe, a crowd is never simply the members it happens to have.

As the categories of 1stness and 2ndness do not capture that aspect of experience in which we apprehend continuity, generality, and law, they do not account for our knowledge or seeming knowledge of physical bodies, dispositions, living bodies, organic functions, persons, personalities, and social organizations. But how are these entities experienced? Are they experienced at all, or only constructed by thinking? For they are comprised, in part, by unactualized possibilities, and thus it would seem that they cannot be present, as such, to sense. We do not perceive the merely possible. Or do we?

6. The Experience of Continuity

In 1887–8, Peirce wrote as follows, evidently about the experience of something continuous:

> Kant gives the erroneous view that ideas [in the Lockean sense] are presented separated and then thought together by the mind. . . . What really happens is that something is presented which in itself has no parts, but which is nevertheless

analyzed by the mind, that is to say, its having parts consists in this, that the mind afterward recognizes those parts in it. (W6:449)

Peirce wrote here of 'parts' without qualification, but he was referring to parts of continuous wholes only and not, say, to people in a crowd. For, of an extensive quantity, only if it is continuous could one say, plausibly, that it 'in itself has no parts' or that its parts 'consist' in being recognized.

In this passage we notice two things. One is that, like Kant and unlike the British empiricists, Peirce held that thinking enters into basic forms of experience. Things as understood, analyzed, or explained are parts of the phaneron. Phaneroscopy does not endorse an explanation, analysis, or understanding; it merely recognizes it. The other is that, unlike both Kant[11] and the empiricists, Peirce held that the content of sense experience, prior to analysis, is itself continuous. For only so could he hold that the role of thinking is in this case analytic rather than synthetic. The continuous whole has to be given first, before it can be analyzed as divisible. It *is* continuous prior to analysis, but is not recognized to be continuous except in analysis.[12]

The potential parts into which a perceived continuum is analyzed are apprehended only through thinking. '[A] mere dull staring at a superficies does not involve the positive apprehension of continuity. . . . Thus, all apprehension of continuity involves a consciousness of learning' (7.536).

[11] In both the A edition and the B edition of the Transcendental Deduction, Kant spoke of separate impressions that must be combined: 'the sequence of one impression upon another' must be 'run through and held together' by a 'synthesis' (A99), and 'the combination . . . of a manifold in general can never come to us through the senses. . . . [A]ll combination . . . is an act of the understanding' (*KdrV*, B129–30, Kemp Smith, trans.). Kant was Lockean in assuming that sensuous intuitions are passive and atomistic.

[12] At the conclusion of W6:499, Peirce said, 'The whole conception of time belongs to genuine synthesis.' But I take that as referring to the conception of time *as a whole*. That conception, then, is the result of putting together the analyses of perceived durations, so as to form an idea of a single, uniform time. Peirce made a similar point about space in a letter of October 27, 1887, to James, that 'objective space' is built up by the 'synthesis of fragmentary spaces', accounting for its 'unity & uniformity'. He then proceeded to suggest that 'the unity of objective time is due to a synthesis of fragmentary times' (L224).

On this topic, Peirce was preceded and influenced by James. Many times before 1886, he espoused the Kantian view that space and time are known by synthesis (1868, W2:199, 199n4, 1878, W3:317, and as late as 1885, W5:246); the analytic view he expressed many times after 1887 (e.g., in 1892, 6.102–63). In 1886–7 James published papers, reprinted in his 1890 *Principles of Psychology*, on the perception of space and time, expressing the analytic view (the 'specious present' is a short span of time perceived as present). Peirce's letter was written to acknowledge the receipt from James of a copy of one of those articles. Cf. Girel 2003, on Peirce's reading of James' *Psychology*.

Why 'consciousness of learning'? Because the apprehension of continuity is not in the conclusion, that there are parts, potentially, but is in the passage to that conclusion from an initial sense of undifferentiated wholeness. The experience of continuity combines sensation and thought: a continuum is present to sense as undivided but thought as divisible. That is why the parts are regarded not as actual but as potential. And yet, those possible parts are present in what is sensed. Ergo, we do perceive what is merely possible, the counterfactually possible. For it is unactualized potentialities that comprise continua, and we perceive continua.

If continua, whether spatial or temporal, are objects of sense experience prior to or independently of the thinking in which their continuity is recognized, then that experience must itself be continuous. In 1893, Peirce wrote: 'consciousness is not limited to a single instant but . . . immediately and objectively extends over a lapse of time' (7.466). What we experience as the present contains the dying past and an aborning future, and thus we have a direct experience of the passage of time. Although not without precedent, this view departs from the dominant tradition.

7. The Experience of Causing

Many continua are not present to sense as such. We observe instances of gravitational attraction, but the universal sway of gravity over empty space, covering a continuum of counterfactual possibilities, is not part of what we see. Such continua are introduced into experience by thought alone: they are not found by analysis of what is sensed but are conjectured. Conjecture, however, presupposes conception. From whence did the concepts of law and continuity arise? We have already seen what Peirce's answer was: they are formed by generalization from those spatial and temporal continuities (real or apparent) that *are* present to sense. So also in the case of causation and causal laws.

You lean your shoulder against the heavy door and feel it slowly open: continuous effort is met by continuous resistance, with a continuous change in position. Thus you experience *opening* the door, that is, *making* it to open. It is an experience of a kind of necessitation short of logical necessity. For it is not logically necessary that the door opens. The experience is lawful in the sense that any continuum is; in this case, you can reflect, 'While pushing, the door opened; the harder I pushed, the faster it opened'. But that is not a law known beforehand (you might have been doubtful whether you could open the door). Nor does it explain

the event; it is experienced in it. It is a lawfulness directly experienced. Any law that explains the event has to be learned inductively from like cases, not necessarily ones of direct experience. Such a law is a law of making-to-happen, of a causing such as that which was directly felt.

We may speak, here, of direct experience of causing, but it is not direct in the sense of being intuitive; it is fallible. It is possible that one is tricked or suffering an illusion (think of a parent opening a door while a toddler, bursting with pride at his prowess, is pushing on it). Further experience may attest to the efficacy of our efforts, or not. Often, conjecture of causal relations not directly felt is better founded than are judgments based on direct experience. The direct experience is nonetheless fundamental to our concept of causing and causation. Note that such experience is also of being caused, for example, to be made to go faster and faster downhill, sledding or skiing.

It is no surprise that this analysis of the idea of causality contradicts Hume's (*Treatise*, Bk. I, pt. III, §XIV); for we saw from the first that phaneroscopy rejects the British empiricists' assumption of sensory atomism. Hume could not have recognized the experience of nonlogical necessity, even when he cited examples of such. For he understood those examples in terms of congeries of separable impressions or ideas. Whereas the primordial experience of causality is of efforts inextricably connected to resistances, and, moreover, it is of these 2nds united in an undivided continuum, lawful in itself: something of indecomposable complexity. (Inconsistently, Hume traced the idea of causal necessity back to our 'inner sense' of the mind's being 'determined by custom' to anticipate an effect upon perceiving the cause.)

Obviously, there is no room here to develop Peirce's idea of causality. Some later philosophers and psychologists, including James, also discussed the perception of causality (see the references in Harré and Madden 1975). But that runs counter to the dominant strand, of regularity analyses and the like, rooted in British empiricism. That strand is not so dominant as it was; G. E. M. Anscombe's well-known 1975 contains many aperçus consistent with the foregoing.[13] Nevertheless, a major contemporary project is to explicate counterfactual conditionals consistently

[13] Ironically, Anscombe's lecture was primarily devoted to disconnecting causality from determinism (1975), as if no one had done so previously; she admitted in a footnote that Ian Hacking had told her this had been anticipated by Peirce. Hacking's allusion was probably to the famous *Monist* article of 1892, 'The Doctrine of Necessity Examined': 6.28–45.

with the assumption that observation is of mere regularities (more on this below, section 9).

At the conclusion of his 1903 lectures on pragmatism, Peirce famously said,

> The elements of every concept enter into logical thought at the gate of perception and make their exit at the gate of purposive action; and whatever cannot show its passports at both of those two gates is to be arrested as unauthorized by reason. (5.212)

This passage has often been misunderstood to conform to the theory of the logical positivists, that every legitimate concept is a logical construct built out of something like Lockean ideas. Peirce's phaneroscopy, which forms a major part of those same lectures, indicates otherwise. The gate of perception opens not only to qualitative contents, but also to forceful oppositions and to law-governed continua. And the gate of purposeful action turns on hinges of subjunctive and counterfactual conditionals. For otherwise purpose could not issue in action. Why would I bother to move out of the way of an oncoming vehicle, if I did not believe that were I not to do so I would be struck down?

8. 3rdness

A combination of two things is triadic, the whole being the third relatum. As a combination of two can be combined with another, it would seem that combinations of more than two can be reduced to a sequence of combinations of things taken two at a time. If so, all combination is triadic. Continuity is the perfection of combination, in which the parts are potential merely. The experience of continuity therefore falls under the category of 3rdness. 'Thirdness is the mode of being of that which is such as it is, in bringing a second and third into relation to each other' (8.328).

The words 'such as it is' are more important than it might at first seem. Here is an alternative formulation: 'The third is that which is as it is owing to things between which it mediates and which it brings into relation' (W5:304). A nail through two boards brings them into relation, but the nail is not 'owing to' the boards it connects. The triadic fact, that nail A connects boards B and C, reduces to a pair of dyadic facts, that A is in B and A is in C. But the whole that is formed by nailing these boards together – and is owing to the things it relates – is irreducible. For it is more than the boards it combines: it is also their being in that

relation. The relation is general and thus the whole can be modified while remaining the same (one board might be rotated relative to the other or additions to the structure might be made). So also, if some additional persons wandered into the crowd of red-headed men choking Fleet Street, it would be the same crowd though larger and perhaps no longer exclusively red-headed. The relation that defines the whole comprehends a continuum of possible modifications. The only binding of parts that is irreducibly triadic is a binding by a general relationship or law of connection.

In these examples, thought predominates: the actual parts, the boards or the people, are present to sense, but the continuum of possibilities that the structure or crowd entails is not present to sense. Thought's role is synthetic. By contrast, when we see a shooting star, thought's role in the perception of movement is analytic; for the continuous streak across the sky is present to sense independently of thought. Either way, thought is essential to the experience.

In a letter of 1903 to William James, Peirce wrote, 'The third element of the phenomenon is that we perceive it to be intelligible, that is, to be subject to law, or capable of being represented by a general sign or Symbol' (8.268). There is a lot packed into that sentence: a distinct element of the phaneron is the intelligible; the intelligible is the lawful; the lawful can be apprehended only in a symbol. A symbol (chapter 8, section 2) is a sign that is a law determining what may count as its instances and also determining what they are to be interpreted as signifying. The intelligible is therefore that which can be apprehended in a sign whose law mirrors its own law:

> the meaning of a word really lies in the way in which it might, in a proper position in a proposition believed, tend to mould the conduct of a person into conformity to that to which it is itself moulded. Not only will meaning always, more or less, in the long run, mould reactions to itself, but it is only in doing so that its own being consists. For this reason I call this element of the phenomenon or object of thought the element of Thirdness. It is that which is what it is by virtue of imparting a quality to reactions in the future. (1.343)

Again, 'every triadic relation involves meaning' (1.345). Verbal meanings are like the laws they represent: conversely, Peirce saw the latter as constituting the meanings of things – as, in ordinary parlance, we might say that today's heavy rain means flooding tomorrow.

In all three categories, Peirce's phaneroscopy revives the Presocratic doctrine, 'like is known by like'. In 1stness, the relation of experience to its

object is one of identity: the quality we feel is the quality of our feeling.[14] In 2ndness, the force experienced is correlative with one's resistance to it. The two are alike in being opposed, and neither could occur without the other. In 3rdness, experience is a step further removed from its object; in it alone is error possible, for in it alone is there judgment. But, when accurate, the experience of 3rdness mirrors its object: thought's conditional expectations diagram the laws they represent.[15]

9. The Categories Interpreted Metaphysically

Concepts acquire additional content in being applied (chapter 10). Peirce's three categories begin life as mathematical ideas of orders of relation, then, applied to the analysis of experience, acquire phaneroscopic meaning, which, in turn, yields an analysis of the modal concepts of possibility, actuality, and law or necessity (logical or nonlogical). Ontological categories of types of being are closely related to the modal categories: corresponding to possibility, we have possibilities, and to actuality, actualities. But what is the metaphysical interpretation of 3rdness? Recall (chapter 2, section 4) that, for Peirce, reality (in its 'realist' definition) is the object represented in that opinion toward which an ideal inquiry would lead in the long run. It is therefore represented by theory, the language of which is explicable, in accordance with the pragmatic maxim, in terms of conditional expectations. Reality is also (in its 'nominalist' definition) that which constrains opinion, shaping it toward its own representation. Either way, reality is lawful; it is law and that which law governs. 'Reality consists in regularity. Real regularity is active law' (5.121). What we judge 'unreal' – the contents of dreams, illusions, fictions, lies, and errors – are things not part of, and often incompatible with, the objects and laws of which we have had sustained and coherent experience. None of this is to deny that laws may be probabilistic or govern imperfectly and that what is lawful in some respects may be chaotic in others. We recognize the reality of chaos when it occurs among entities in other respects lawful and thus real.

[14] Notice that by making likeness in this instance to be identity, Peirce avoids Wittgenstein's critique of images as 'inner pictures' resembling 'outer' pictures (1967 [1953], p. 196). We lack room to explore the implication that perception is 'immediate' (but see 7.639). Wittgenstein's argument does not affect the doctrine that habit mirrors law, as that relation is publicly observable, not private and 'inner'.

[15] Richard Rorty, that famous smasher of mirrors of nature, is right to exclude Peirce from the pantheon of 'pragmatist' philosophers he admires (Rorty 1982, pp. 160–1); nor would Peirce have admired Rorty (Haack 1998).

Peirce distinguished between reality and actuality, or existence. The existing is the reacting; reaction is instantaneous, here and now. Reality is enjoyed by laws that have no here-and-now existence and that are not reactants. At the same time, there would be no reality without existence. For we call no represented law real if it does not govern actual reactions. It is by observation of what actually occurs that we distinguish real laws from mistakenly supposed laws. But law is not reducible to the reactions it governs, for it is not exhausted by the regularities that have occurred. If it is real, it makes subjunctive and counterfactual conditionals true; conduct may reliably be based on it (see below).

Some kinds are real, being defined by laws governing actualities. Other kinds are nominal, defined by our arbitrary choice of attributes. Biological taxa and words are real kinds that come into existence and go out of existence at given times and places. Before 1942, the word 'radar' did not exist; the dodo no longer exists. They come into existence through the exemplars they have then and there; and, through these exemplars, more exemplars are propagated, in conformity to the law exemplified (chapter 5, note 8; chapter 8, section 1). Other kinds, such as chemical elements (even those not yet occurring but predicted by the Periodic Table), are so closely tied to laws by which actual instances may be generated from existing materials that they may be said to be real and not nominal, although there is no time at which they came into existence.

Individual things, such as this cat, this house, this person, this nation, are law-governed continua (spatial and temporal) of instantaneous reactions. Therefore we speak of them with equal correctness as real and as existing. But we tend to use 'exist' more with reference to time and place (dodos exist now only in museums), 'reality' more when it is a question of what or whether something is (Was that really a cat? Was that cat real?). Processes are real and exist (or occur) in the same way as do individual things, for greater or lesser periods, shading when shorter into events; they exhibit lawfulness. Events, even if nearly instantaneous, are judged real if they can be located in a web of cause and effect, more or less lawful. Much of this is a matter of ordinary usage, which Peirce's metaphysical distinctions explain.

Ordinary usage, usefully supple, foils pedantry. The word 'exists' and its cognates are used freely wherever the locution 'there is' might be used, regardless of the category of being that is at issue: impossible creatures exist in fable, there exist solutions to mathematical equations, and so on. Peirce and the present author may also be allowed such syncategorematic usage, I hope, without being charged with inconsistency.

We lack room to compare Peirce's metaphysics to contemporary views; but perhaps I may briefly suggest the stance one might take. On the Stalnaker-Lewis theory of counterfactuals (Stalnaker 1968; Lewis 1973a, b), to see whether 'if *p* were, *q* would be' is true, one looks for a possible world (1) in which *p* and *q* are both true and (2) that is more similar to the actual world than is any possible world in which *p* is true and *q* is false. There are standard objections to this, concerning the individuation of possible worlds, the identity of individuals across possible worlds, and what 'similarity' might mean when applied to possible worlds. There is a further objection, one that I think shows Peirce's theory to occupy a relatively strong position. It is to be found by reflecting on why the late David Lewis put all possible worlds on an ontological par.

The only thing that makes our world actual, Lewis said, is that we are in it; intelligent denizens of other possible worlds doubtless think their world is actual, and do so on the same basis (Lewis 1973a, §4.1). He insisted that this is what the affirmation of possibility entails. One might instead describe it as the denial of possibility, as it allows for possibilities only if they are just like actualities. But I suspect that there is reason behind the madness. Since Lewis did not, so far as I know, state that reason, I shall. If possible worlds are merely possible, then the locution '*p is* true in possible world *W*' is a solecism, or at best elliptical, for '*p* would be true were *W* actual'. And that is a counterfactual. Hence, far from being explicated by possible world semantics, counterfactuals are presupposed therein – unless we adopt Lewis' absurd stratagem of treating each possibility as someone's actuality.[16]

Alvin Plantinga has in other ways raised the same specter of circularity: for example, he points out that worlds are similar in part by having the same physical laws, as Lewis admits; but counterfactuals are implicated by laws, and therefore trans-world similarity can hardly be used to explicate counterfactuals (1974, p. 178). (None of this is to deny that possible worlds semantics may be used, as by Kripke, as a model by which to

[16] One might also wonder about the peculiar use of the word 'in' when Lewis and others speak of a proposition *p* being true *in* a possible world *W*. 'Truth in' has become a common idiom among logicians, perhaps beginning with Tarski's talk of truth in one or another formal language. But in ordinary speech, propositions are not true *in* anything; they are true *of* things. And therefore we should reflect on what it might mean to speak of 'truth in', especially when that mode of speech is extended from languages to worlds. To say that *p* is true 'in' *W* seems to me to be a way of wriggling out of the uncomfortable choice between treating *W* as actual or treating *p* as what would be true were *W* actual. As the latter is a counterfactual, it thwarts the desire to use possible world semantics to explicate counterfactuality.

demonstrate the consistency of a logic of modal operators. Models typically are artificial.)

Now, Lewis admitted that one can take modal idioms (ditto counterfactuals) as unanalyzed primitives – that is, as having a logic but no analysis into logically simpler concepts – but that, he added, 'is not an alternative theory at all, but an abstinence from theorizing' (1973a, p. 85). A remarkable statement. He assumed that a theory of modality can only be a logical analysis of modal language into nonmodal language. And that would appear to be a false assumption, in light of Peirce's phaneroscopy. As we noted in section 3, Peirce used the tools of logic (more precisely, the algebra of relations, which is a branch of mathematics) to form a formal, relational analysis of the phaneron, not a logical analysis of concepts or language. If Peirce was right, that 1stness, 2ndness, and 3rdness are indecomposable elements of experience, in which such metaphysical concepts as those of possibility, actuality, and law are rooted, then an analysis of those concepts into simpler concepts is impossible. But that does not preclude phaneroscopic analysis, which then provides an alternative theory of modality.

10. The System of Categories

The formal structure of phaneroscopy augurs an elaborate system, and that in two respects. On the one hand, the categories apply to one another, and on the other, they subdivide endlessly.

Categories apply to one another in this wise. A real law is distinguished from a possible law by its actually applying to actual occurrences; that is the 2ndness of 3rdness. And actual occurrences and real laws are possibilities that have been realized; their being possibilities is the 1stness of 2ndness and the 1stness of 3rdness, respectively (1.530–4). That is expressing the matter ontologically, but it is also true phaneroscopically. The 1stness of 2ndness, as an element of the phaneron, is the feeling of occurrence; that of 3rdness is the feeling of continuity; and the 2ndness of 3rdness is the experience of not being able to violate a law or of its taking effort to divide a continuum into parts, and so on.

And categories subdivide as follows. A 1st's occurring – hence, not the 1st itself but the occurrence of it – is a 'degenerate' form of 2ndness in which one pole is a 1st: the 2ndness is added to it, it does not constitute it (its being in this sample does not make the color to be what it is). In 'genuine' 2ndness, both poles are 2nds, as in a physical action and reaction: neither is what it is apart from its opposition to the other

(1.527–8, cf. 1.462–70). Degeneracy is not reducibility: the fact that a 1st occurs is not reducible to relations other than dyadic (the color's being in this sample is not a matter of pure 1stness).[17]

3rdness, Peirce said, exhibits two degrees of degeneracy (1.473, 5.70); that works out into an unending series of trichotomous divisions, according as each of the three relata of a triad is either a 1st, a 2nd, or a 3rd, and as each component 2nd or 3rd is either genuine or degenerate in one or another degree (5.72). For example, causal laws obtain among realities, but the color spectrum is defined by a law of continuous gradation that holds among possibilities (and yet it is not a law of logic). The latter is monadically degenerate (1.473).

Other examples are harder to come by, apart from Peirce's taxonomy of signs. It is in that taxonomy (chapters 8 and 9) that the formal structure of phaneroscopy is best illustrated and proven to be fruitful. At least in my experience, the distinctions it entails often seem at first to be empty, but then the search for examples is rewarded, and each such discovery sheds an unexpected light.

[17] However, Peirce's discussions of these matters were fragmentary and marvelously inconsistent and often unpersuasive. Thus, he also used the genuine/degenerate terminology, borrowed from mathematics, to refer to irreducible/reducible instances of the categories (e.g., at 1.365). See Kruse 1991 for a careful discussion of Peirce's various uses of these terms; also below, chapter 8, sections 3 and 4.

4

A Preface to Final Causation

'Why is Jones running?' 'To get to work on time.' We speak of human actions as being done for the sake of an end or in order to bring it about. It is the end sought that explains the act. Aristotle extended the same way of thinking to natural processes, but without supposing that they are directed by anything like a human intelligence. In his view, an acorn's structure and stages of growth are for the sake of its becoming an oak: the end, without its having been consciously entertained by anyone, accounts for the means by which it tends to be achieved. We shall name such ends 'final causes'. By 'teleology' we shall mean such a doctrine as Aristotle's, that there are final causes in nature.

So understood, teleology is supposed to have been overthrown in the Scientific Revolution of the seventeenth century, which replaced it by mechanistic explanation. Organic life, which seems organized for the sake of its perpetuation, was an apparent exception, but Darwin's theory of natural selection is said to have expelled teleology from that last redoubt. Peirce nevertheless revived the doctrine, although in terms of a metaphysics different from Aristotle's. According to Peirce, final causation is not opposed to modern science but is implicit in some of its theories, Darwin's especially.

A similar view has been developed more recently by some biologists and philosophers, but only at a methodological level, eschewing metaphysical commitment to the reality of final causes. In the next chapter, I argue that the methodological thesis cannot be sustained by itself; it requires Peirce's deeper-going view. It will be seen that Peirce's account of teleological explanation removes the mystery from teleology and shows it to be a rationally acceptable part of natural science.

Peirce's teleology grounds his mature semeiotic – in particular, its analysis of what it is to signify. Thus intentionality (chapter 1, section 3) acquires a naturalistic explanation. Mind and freedom are shown to be an aspect of some natural processes, and that aspect is shown to be irreducible to the mechanical aspects of those same processes. But neither this chapter nor the next draws those implications, which are deferred to chapters 6 and 11. This chapter prepares the ground for the next, and the next confines itself to a systematic development and defense of Peirce's concept of final causation – a concept at once naturalistic and nonmechanistic.

A prefatory chapter is needed, because of the tangle of confusions and misunderstandings that have grown over the idea of teleology. Those errors have to be hacked away, one at a time, before we can get a clear view of the subject. Error, however, is not the whole problem; underlying the many errors there is a single theme: the profound resistance to teleology, or to anything that challenges the reign of materialism and mechanistic explanation.

1. Strange Objects of Desire

We begin with a merely grammatical point, about English usage: a purpose is always a type, hence, a universal. A purpose is always general; it is never a particular. A purpose has to be general, since, when it is a purpose, it is not yet attained: it is not yet actual, and only the actual or what was actual is particular. What is yet to be done, no matter how exactly we specify it, leaves room for infinite variations, even if minute and unimportant. Of course, normally, one seeks an outcome that will be a particular, but what one seeks is a particular – any particular – of a given type. It is the type alone that can be specified; before the type is achieved, there is no particular of that type that could be specified. It therefore makes no sense to suppose that a purpose is a particular.[1]

Types tend to get short shrift, as if all that there is to talk about, really, are particulars. That types are nothing but *façons de parler* is the doctrine known as nominalism, that universals exist in name only. Like the rejection of teleology, nominalism marks the modern period. But regardless of what position one takes on the question of nominalism, there is no

[1] 'Particular' is used here as a noun; used as an adjective, we can of course speak of particular purposes or of one purpose in particular. See the section 'A Note on Terminology' in the Preface.

getting around the fact that we cannot refer to a purpose without referring to a type.

Purposes are often confused with desires qua psychological states. That is impossible, as a psychological state is a particular, while a purpose is a type. At most, a purpose is a type of outcome that is desired. The confusion is perhaps due to the fact that the plural noun 'desires' is used in two senses: for states of desiring and for the things desired. But it also reflects a modern prejudice toward explaining purposes psychologically. But even if we were to adopt the theory that purpose always is a consequence of desire, we have to recognize that it is not a simple consequence. For desires can be resisted, and, thus, it can be one's purpose to resist certain desires. At most, then, a person's purposes always reflect his strongest desires or strongest set of compossible desires. But that is only a theory, not an analytic truth about purpose. Those inclined toward such a view are strongly recommended to read Dostoyevski.

As far as the word 'purpose' is concerned, it is open to us to consider the classical view, associated with Plato and Aristotle and St. Thomas Aquinas and other premodern thinkers, and also with Peirce and a few other moderns, that certain types of possible outcome, by their inherent nature, create a desire for them. Although they are general and are no more than as yet unactualized possibilities, purposes, on that view, may be objective. For they become our purposes because of what they are, and not because of our subjective constitution. In that case, our purposes are not, or need not always be, grounded, subjectively, in our desires; our desires might instead be sometimes grounded in what is objectively good, other times in what appears to be but is not good.

The considerations advanced here about the word 'purpose' apply equally to such other words as 'end' (not in the sense of terminus but in the sense of something sought) and 'goal'. In all of these cases, what might be said to be a final cause is inherently general, a type of possible outcome, and not a particular existent or anything actual. And the question can be raised, whether these possibilities are made to be ends or goals by desire or other subjective states of an organism or, alternatively, whether they have an objective mode of being qua end or goal.

However, there are in the literature of teleology other examples of final cause that appear to be actual or particular beings and not types. Exemplary individuals are said to inspire emulation and thus to bring about imitations of themselves (thus Aristotle's conception of God as final cause of cosmic order). But in the case of such a final cause, we can speak of an associated end that is general, namely, to be like the individual in

question. Many can be like the same individual in some respect, and can achieve that kind of likeness in different ways, under different conditions, at different times and places, and to various degrees. It is that general end, and not the individual, that is sought. Furthermore, it is the general character found in the individual, and not the individual per se, that inspires.

Similar considerations would apply to Plato's theory of Forms, discussed below, except for this consideration: the Forms are timeless, placeless ideals, yet they are not general but are concrete in the sense of being infinitely specific and therefore predicable of nothing but themselves. It may therefore be a distortion of Plato's meaning to suppose that what is inspiring about a Form is some general feature it possesses. Plato either lacked or chose not to employ a concept of generality. The end sought under a Form's inspiration is nonetheless general, a type of outcome. If we wish to include Plato's theory among teleologies, then we should have to adopt this formula, that a final cause may either be the type of outcome sought or an ideal archetype emulated. The type sought is of course something that ideally will be achieved, while the ideal emulated entails a type of outcome sought.

In either case, we have a doctrine that seems extraordinary, even preposterous, indeed, hardly conceivable from the modern point of view, namely, that mere possibilities have an influence over what actually transpires or, in other words, that the ideal influences the actual. The question thus raised about our purposes is the one famously formulated by Plato in his dialogue *Euthyphro*, about God. Is X good because God desires it or does God desire X because it is good? Is X good because we desire it, or do we desire X because it is, or appears to us to be, good? Is goodness a function of subjective states (whether human or divine) or is it objective? This question has enormous implications, not only for whether moral judgment is subjective or objective, but also for the views we take of selfhood, freedom, scientific objectivity, and so on. Politically, in words Abraham Lincoln used, at the conclusion of his Cooper Union address, the question is whether might makes right, or right makes might. Lincoln said, 'We must have faith that right makes might'. If right makes might, then the ideal influences the actual.

2. What Is Mechanical?

'Mechanical' has no standard unambiguous usage at present nor, indeed, any clear meaning. Nevertheless, contemporary philosophers tend to

believe that, at bottom, the world is mechanical. Perhaps the latter phenomenon depends on the former.

In a recent dictionary of philosophy, David Hull mentions a spectrum of definitions of 'mechanistic explanation', saying that at its most inclusive, the term means 'little more than a commitment to naturalistic explanations' (Audi 1999, pp. 550–1). What the little more is, he does not say. But naturalistic explanation itself does not exclude final causes as understood by Aristotle or Peirce; for neither thought that his doctrine invoked supernatural causes or that it depends on anything other than empirical investigation of the way the world actually is.

Historically, as Hull notes, mechanistic explanations were opposed to teleology. We shall use 'mechanical' and its cognates in that historical sense. Whatever we make 'mechanical' to mean and whatever we make 'final' to mean, we will keep them opposed. Opposed, that is, not in the sense that explanations of both types cannot be legitimate, but in the sense that they can never be the same or reduced one to the other. But that decision does not determine the precise meaning of 'mechanical', since the conceptions of teleological explanation and of final causation are yet to be determined. Our question is whether naturalistic explanation must always be exclusively mechanistic. To frame that question in a nontrivial way, we need definitions of 'mechanical' and of 'final' that are rooted in historical usage but that are also the broadest possible while still maintaining their mutual opposition.

The science of mechanics is too narrow to provide such a definition. Not all theories in modern science that have been opposed to teleology belong to mechanics. Hence, we shall have to form a more general idea of the mechanistic, of which mechanics will be but one example. But let us begin by reflecting on mechanics.

The conception of mechanics has undergone a remarkable evolution. At first, mechanics was the study of the transmission of motion through bodily contact, as with cogwheels and billiard balls. But gravity, whose law Newton formulated but did not explain by any mechanism, seemed to be a kind of 'action at a distance', that is, force acting instantaneously over distance without bodily contact. And with quantum mechanics, in which there are even stranger relations between distant particles, the assumption of determinism was replaced by probability. The standard definition of mechanics, as the science of the effects, either movement or equilibrium, of forces on bodies, abstracts from these variants (if we allow that 'effects' may be related only probabilistically to their 'causes'). But what is a body?

In the development of the wave theory of light and field theories of electromagnetism (in which, at any rate, there is no action at a distance, as they allot time to the propagation of energy through space), a continuous material medium was once supposed. At another time, wave and field phenomena were interpreted by a hypothetical interaction of discrete particles. Here we have two ideas of the body involved, one continuous and singular, the other discontinuous and plural. At a later stage in the development of these theories, no material medium at all was supposed; wave and field theories thus became independent of mechanics. In contemporary physics, with Schrödinger's wave mechanics, matter itself is seen to have wave characteristics. So, what is matter? The particles of microphysics do not behave in ways it was once thought proper for bodies to behave. Thus it is far from clear what philosophers today, who believe that matter and mechanical action is everything, really do believe.

The generalized idea of mechanics mentioned above is subject to a further generalization, by dropping its reference to bodies and thus extending its reach to wave and field theories. The result cannot be called 'mechanics', but we will adopt it as a definition of 'mechanistic'. Mechanics is then but one mechanistic science among others.

Let us say that an explanation of a particular, E (for effect), is *mechanistic* if and only if, by general laws or equations, deterministic or probabilistic, it relates E to particulars (forces, bodies, events, states, conditions, fields, or processes) that exist or occur or obtain not later than E. The laws cited will also be called 'mechanistic': *mechanistic laws* relate particulars to particulars – that is, they relate particulars of one type to particulars of other types. And something is *mechanical*, we shall say, if, and only so far as, it conforms to mechanistic laws: that is, if, and only so far as, facts about it are explicable mechanistically. A *mechanical cause* is a particular that is not later than its effect, to which it is related by a mechanistic law.

There are of course mechanistic explanations of laws or general phenomena as well as of particulars, and these are usually the explanations that are of most interest in science. Roughly, such explanations show the law or general phenomenon to be explained to be an instance of other laws, perhaps as applied to conditions of certain kinds. But these latter laws must also be mechanistic, that is, they relate particulars of one type to particulars of other types. That is fundamental to our ensuing argument, namely, that mechanistic explanation is always in terms of laws that relate particulars (of one type) to particulars (of other types).

In practice, the requirement of law may be greatly relaxed: a rough idea that this is a regular way that things go on may do, and that idea does

not have to be stated. 'The window was broken by being hit by a stone' is a mechanistic explanation. Also, we said nothing about whether laws must be universal or may be local, nor how causes are to be identified (e.g., as complete, partial, necessary, sufficient, independently controlled, salient practically). Our definitions take for granted that there are concepts of explanation, law, and cause; but they presuppose no specific such concepts. Thus we evade complex controversies. We have only stipulated what, in each of those categories, however they may be construed, we shall call 'mechanistic' or 'mechanical'.

There are philosophers who insist that a mechanistic explanation must cite particular mechanisms that 'bring about' the effects explained. We omit such a clause, since many scientists and philosophers have thought that subsumption under law suffices to explain phenomena nonteleologically. But the broth may be peppered with such clauses, according to your taste. Our definitions are deliberately broad, meaning only to exclude the teleological.

Final causation is excluded by our having made mechanistic explanation of particulars always to be *by* particulars. For a final cause is never a particular. Particulars are identified spatio-temporally.[2] Final causes – whether ends or ideals – have no spatio-temporal identity or particular existence.

By this definition of 'mechanistic', psychological, sociological, and economic explanations are also mechanistic to the extent that they explain particular outcomes as following by law, often probabilistic, from particular conditions. If that seems too much a stretch of the mechanistic idea, we may call these explanations 'nomological', noting that they also are opposed to teleological explanation.

It has been common among philosophers at least since Hume to suppose that mechanical causes are particular events that precede and determine their particular effects, which are also events. We have brought that idea into closer conformity with physical theory, wherein events are not always at issue, the processes described are often continuous, many equations relate coëxisting conditions, and laws may be probabilistic.[3]

[2] Spatio-temporal location may be complex. Two fields of force extend throughout space and in that sense coincide. Yet they are distinguishable spatio-temporally by the fact that they are identified with different magnitudes and directions at the same spatio-temporal points (in the simplest cases, they have distinct centers). The magnitude and direction of a force is of course revealed through actual effects on existing bodies.

[3] Many authors instead deny that causality has much to do with modern physics. It comes to the same. My choice is dictated by the convenience of causal language for our purpose.

Notice that I have used the adjective 'mechanical' to characterize phenomena, their causes, and so on, and the adjective 'mechanistic' to characterize theories and explanations. (By calling laws 'mechanistic', we take them *as stated*; that is not to deny their reality.) For our purposes, it helps to keep these two levels terminologically distinct. For example, to call a mechanistic explanation 'mechanical' is confusing, as that could mean that it was produced thoughtlessly. Philosophy presents a third level. The philosophical idea that all of the world operates mechanically and that everything can be explained mechanistically is conventionally named 'mechanism', but as that term also applies to particular mechanical systems, I suggest that we use 'mechanicalism' instead.[4] One can accept many mechanistic explanations and theories – one can be a physicist working exclusively in mechanics – without being a mechanicalist. I like my neologism 'mechanicalism' because it is as awkward as is, in my opinion, the view it denotes.

3. Teleology's *Locus Classicus*

The word 'teleology', derived from Greek roots, is not Greek. It appeared first in 1728 as a Latin coinage, *teleologia*, by Christian Wolff, to designate his Leibnizian, not to say Panglossian, view that everything is for the best and, in particular, for the good of human beings. It was doubtless Aristotle's use of the word *telos*, Greek for an end, completion, or fulfillment, that inspired Wolff's invention. And, yet, when Aristotle cited a *telos* to explain such a process as the growth of an acorn into an oak, he did not imply that it occurs for man's good. The acorn does not grow in order to provide us shade in the summer and firewood in the winter. Nor did Aristotle imply that things are designed by a divine craftsman for purposes of his own. If, then, we cite Aristotle's philosophy as the *locus classicus* of teleology, as we have good reason to do, it is with some irony, since he himself did not use the term and since those who coined it held views contrary to his own.

One prominent misunderstanding of teleology, possible only to those ignorant of Greek philosophy, is that it is a primitive doctrine, something like animism, that preceded mechanistic explanations and that must give way before the obvious superiority of the latter. The opposite is the truth.

[4] E. J. Dijksterhuis, in his magisterial historical study, *The Mechanization of the World Picture*, struggled with these same terminological problems, solving them differently (1986 [1950], p. 3).

Teleological ideas were introduced in the fifth century B.C., explicitly in reaction to the mechanistic theories of earlier thinkers and in an effort to explain what they had left unexplained.

Mechanistic explanations had been familiar to the ancient Greeks, from the Milesian philosophers of the sixth century through to the atomists of the fifth century. Admittedly, those explanations were crude and speculative (e.g., that air is rarefied water or that objects are made by atoms interlocking); they had no name equivalent to 'mechanistic'; they were not conceived of as we conceive of mechanistic explanations, in terms of laws linking cause to effect; and, in some cases, though definitely not in the case of the atomists, they were mixed up with nonmechanistic ideas (e.g., of love and hate as forces of combination or separation). Nevertheless, it will be convenient, and not for present purposes misleading, to call them mechanistic.

Now, mechanical pushes and pulls and buffets have predictable results, but there is no reason, in mechanistic explanation itself, to suppose that those results will add up to an orderly outcome. Take hammer and chisel to a block of marble, and we can predict that a series of blows will produce a series of chips. But what explains why a statue of Apollo emerges? Teleology was introduced to explain the emergence of order from out of chaos – that is, order among outcomes: a work of beauty, an orderly cosmos, a good society. For that was something that it seemed mechanistic explanation could not do.[5]

The first clear expression we have of teleological thinking is in Plato's *Phaedo*, in words given to Socrates, who attributes the idea to Anaxagoras, his older contemporary. Anaxagoras said that intelligence or *noûs* controls existence, making things to be for the best. So also, Socrates, in prison awaiting execution after having declined an opportunity to escape, is portrayed as ridiculing the thought that his bones and sinews alone can explain why he is there, and not his judgment that it is better to comply with Athens' law, preserving the order of his *polis*, than to violate it. Two forms of explanation are therein implied; for surely bones and sinews also explain what Socrates does, as well as his judgment about what is best.

[5] Nor does that prove any very great want of ability on the part of ancient philosophers. Newton himself thought that the formation of the solar system could not be explained on his principles but required an act of divine creation. However, I am not going to argue that every example of order requires a teleological explanation. See, e.g., the discussion of statistical mechanics in chapter 5.

In what is presumably a later dialogue, the *Timaeus*, Plato explicitly mentioned (at 48A) two kinds of cause (*aitia*), of which one is 'the errant cause' or necessity (*ananke*). *Ananke* relates to compulsion by torture as opposed to persuasion by reason, but it came to mean compulsion generally; in the *Timaeus*'s context, 'mechanical compulsion' is a reasonable gloss. Order is created by subordinating necessity – that is, mechanical compulsion – to reason. 'We must speak of both kinds of cause, but distinguish causes that work with intelligence to produce what is good and desirable, from those which, being destitute of reason, produce their sundry effects at random and without order' (46E, Cornford, trans.). Note that 'random' here means lack of purpose, not lack of mechanical necessity, and that order is alleged to be missing in the effects, not in the sequence of cause and effect.

It is clear, then, that teleological explanation was introduced not in ignorance of mechanistic explanations but deliberately to supplement those explanations and thus to account for the forms of order they leave unexplained, namely, order of outcomes, as opposed to order of sequence.

All the other misunderstandings of teleology derive from the mechanicalist assumption that only particulars can explain particulars. Thus, contemporary philosophers tend to identify explanatory types with one or another particular: God's acts of volition, human desires, or future effects (mistakenly supposed to be particular even before they occur). That anything else could be a cause is to them so absurd an idea that they do not recognize it even in those texts, of Plato and Aristotle, where it is asserted. It is necessary, therefore, to take a moment with those texts.

In brief, I shall argue that Aristotle identified final causes as types and that Plato arguably but not certainly held a similar view. The term 'final cause' is a great convenience here. However, that term, and also 'efficient cause', now standard, were derived from the thirteenth-century Latin translations of Aristotle and have no Greek equivalents. 'Final cause' corresponds to Aristotle's phrases 'the cause that is a *telos*' and 'what something is for', and 'efficient cause' corresponds to 'the primary source of the change or the staying unchanged' (*Phys.* II, 3:194b30–3, Charlton, trans.). Examples of efficient cause are surprisingly varied – the father of the child, the sculptor's art of the statue, the motive of the action, luck and the automatic – and Aristotle's conception of efficient causation is certainly not ours of mechanical causation. For one thing, some of the efficient causes he mentions are general, for example, the sculptor's art. Causes we would call mechanical, some of which Aristotle

mentions, appear to fall under the subcategory of the automatic or luck.

Final causes, as Aristotle conceived of them, are clearly universals or types, although not all universals are final causes. However, it is a distinguishing feature of Aristotle's teleology that a universal, to be a final cause, must be in something particular, whether as contained therein or as the form thereof.[6] It must be the form of the parent; or it must be present in the seed, in the immature organism, or in the mind of the artist; or it must be the form of something good and beautiful, for example, God, that is imitated. He adds that the final cause, in the parent or seed or mind of the artist (but not in God), is also the efficient cause. Presumably its powers qua final and qua efficient are distinct, but this has been left obscure.

Plato, as always, is more difficult to pin down. Socrates reports in the *Phaedo* that he was disappointed by Anaxagoras' failure to carry through the idea that intelligence governs existence; doubting his own capacity to complete the task, he says, he attempted a 'second sailing', namely, the theory of Forms. The suggestion is that that theory is second-best. Apart from the question of whether the theory of Forms was held by the historical Socrates, there are at least two other questions. Did Plato intend the reader to accept the claim that the theory of Forms is second-best? Socrates' modesty was famously ironic and Plato's literary method profoundly indirect. And whether second-best or not, was the theory of Forms intended to be a variant of Anaxagoras' view or a third alternative, opposed equally to Anaxagoras and Democritus? That is, is it properly described as a teleological theory?

In the *Timaeus*, order among existing things is in some passages attributed to a divine craftsman, or Demiurge, who shapes things into imitations of the Forms. Did Plato mean us to conclude that order requires the forceful actions of an agent? If so, then the existence of order testifies to the existence of a beneficent Creator. Such a view is sometimes called a 'theistic' or 'transcendent' teleology, as opposed to Aristotle's 'natural' or 'immanent' teleology, of which more below (section 4). And perhaps that is the sort of teleology proposed in Socrates' 'second sailing'.

However, there are other passages in the *Timaeus* where the Demiurge is not mentioned and the Forms appear to have an influence directly on what would otherwise be the chaos of matter banging around

[6] The major texts from which I draw my account are *Phys.* II, *De Part. An.* I, *Meta.* Δ, 2, 4, 5, Θ, 7, Λ, 6–10.

mechanically. Thus, the Demiurge may be no more than a literary conceit personifying the power of the Forms. If that is the case, then it is arguable that Plato had a teleology very much in the same sense that his pupil Aristotle had one. In both, the ideal has a direct influence on the actual, accounting for the order – the beauty and goodness – we find in nature and, sometimes, in human conduct.

But what of the idea that the ideal requires an active agent? If Plato introduced that idea, then either he was serious about it or he must also have intimated some sort of rebuttal to it. In several dialogues, he interposed an agent between the rational and the material: in the *Republic*, a warrior class is required to put the wisdom of the philosopher-rulers into practice, controlling the less wise populace, and in that dialogue and in the *Phaedrus*, a corresponding spirited part of the soul controls the soul's desiring part at the behest of the reasoning part. So, it may seem that he did indeed think an active agent is necessary. However, these very formulations make the following question obvious and inescapable: how are warriors (enforcing political order) or the spirited part of the soul (enforcing reason's rule) or the Demiurge (creating cosmic order) *themselves* persuaded by reason? To postulate such agents merely reproduces the problem without solving it. If the ideal, to have any power, requires to be enforced by an active agent, then there is no way to explain its power over that agent, except by postulating another such agent, and so on ad infinitum. We must conclude that, if there is any order, beauty, or goodness at all, then there is a direct influence of the ideal on the actual. And it seems to me that Plato must have intended us to see this. Recall the question raised in the *Euthyphro*, mentioned earlier. If, as the *Timaeus* suggests, the Demiurge creates order in imitation of the Forms *because* the latter are good, then the Forms, by their goodness, have a direct power over the Demiurge's actions. By the same token, it is unnecessary to postulate a Demiurge.[7]

Some scholars have drawn a different conclusion, namely, that a theory of Forms without a Demiurge is no teleology at all. They claim that the

[7] You might object that Forms can have this power only through a consciousness of them, and that is why cosmic order requires a Demiurge. But that does not explain why the reasoning part of the soul, which is a consciousness of the Forms, requires a spirited part. Either the reasoning part can act, in which case the spirited part is superfluous, or it cannot, in which case the spirited part is out of its control. More important, this response dumps the problem into the mysterious bucket named 'consciousness'. If there is to be an explanation of consciousness, and if consciousness can be moved to action by the ideals it is conscious of, then *that* capacity has to be explained. And, so, we are back to where we started. How can the ideal have any influence on the actual?

'second sailing' was indeed a second-best theory – one that gave up the attempt to explain the emergence of order. Not implausibly, they associate Plato's Forms with Aristotle's formal cause rather than with his final cause. A formal cause is that form that constitutes *what* a thing is, just as that same thing's matter or material cause accounts for *that* it is here not there, now not then. An individual is composed of form in matter or of matter formed. Just as the efficient and final causes, in Aristotle's philosophy, are dual aspects of becoming, the formal and material causes are dual aspects of being. And Plato, as he mentions only the Forms, gives only a formal analysis of being, they say, and not a teleological explanation of becoming.

However, in both the *Phaedo* and the *Timaeus*, the question addressed is that of becoming. The formula Plato employed at *Phaedo* 95E is repeated verbatim by Aristotle at the beginning of *Physics* II, 3, except for the addition of one word: what is the cause (Aristotle says, primary cause) of generation and destruction? If the theory of Forms is meant to answer that question, even if only in a second-best way, then the *eidē* must be causes not merely of being but of becoming.[8] But so far as it is intended to account for the occurrence of its copies, an *eidos* is a final cause in Aristotle's sense; that is, it does not act mechanically but, rather, has an influence by evoking desire and emulation. Of course, it is not such a final cause as Aristotle's metaphysics accommodates.[9]

4. A Budget of Errors

If we define 'teleology' by its *locus classicus* in the philosophies of Plato and Aristotle, then most of the teleologies we find in the history of thought are not teleologies at all.[10] Instead, they evoke one or another efficient cause. The Judeo-Christian-Islamic idea of a Creator, who fashions the world in

[8] So Aristotle understood the *Phaedo*: see *Meta.* A, 9:991b3–4. His complaint there, more fully at *De Gen. et Corr.* 335b9–16, is the neglect of an efficient cause to explain why a Form is copied at certain places and times and not at all others. But that lack is supplied in the *Timaeus*, which postulates a turbulent, grainy, recalcitrant 'Receptacle', of matter in mechanical motion, which is to be shaped into imitations of the Forms (49B-51B). A Form, we may then interpolate, will be copied where it can be, for as long as it can be, and not where or when the resistance is too great. Aristotle elsewhere said (*Meta.* A, 6:988a8–10) that Plato acknowledged only two causes, the formal and material; these, then, also serve, when applied to becoming, as final and efficient causes, respectively.

[9] The view of the *Phaedo* I have presented here is more or less the traditional one. It is also the one suited to Peirce's understanding of final causality. It has been much contested in recent years, most notably by Vlastos 1969; cf. Wiggins 1986. See Stern 1993, ch. 4, for references to the literature.

[10] See Lennox 1992 for a clear, succinct classification.

conformity to His will, has often been called a teleology, but God and His acts are particulars (albeit rather large ones), hence, efficient causes. So also the various vitalistic theories in biology, that have postulated a force within organisms that drives them toward certain goals of maturation, and so on: these are particular forces. Whether we wish to call such doctrines 'teleology' or not, they are the polar opposite of classical teleology.

The same applies to the many modern identifications of teleological explanation with one or another mechanistic explanation, for example, so-called cybernetic explanations in terms of feedback loops (chapter 5, section 6). The attraction of such analyses is that they reduce what seems not mechanistic to a mechanism. But, by that same token, these versions of teleological explanation are not teleological in the classical sense. So also when a modern author explains goal-directed behavior in terms of an 'intentional state', such as a desire, and then construes that state as operating as a force or efficient cause.[11] For Aristotle, when a person walks for his health (the Peripatetic Philosopher's favorite example), the final cause that explains that action is not a desire to be healthy, but is health itself. Health, by its attractiveness, explains the desire for it (*De An.* III, 10, esp. 433b13–20).

Another mistaken idea about teleology[12] is that it boils down to 'reverse causation', wherein the future determines the present or wherein mechanical effects are the causes of their mechanical causes. A causes B mechanically and B causes A finally, so B causes itself (but so, also, A causes itself!). A tidy theory, but I know of no one who held it. That which comes about through the influence of a *telos* or an *eidos* is not that *telos* or that *eidos*. The effect is a particular of the type and is not the type itself. For that reason alone, final causation cannot be backward causation. A final cause is a possibility, not a future actuality. Those who define final causation as reverse causation set up a straw man.

Even among professional philosophers, it is today the misunderstandings of teleology that have prevailed; teleology as classically understood has almost vanished from sight, submerged beneath waves of error. Take, for example, this summary statement of teleology in a 'handbook' on the philosophy of biology by one of its prominent practitioners:

> In the good old days, that is, in the days of Aristotle and his Christianized followers, teleology used to refer to life forces moving toward their goals, or to a Supreme

[11] Woodfield 1976 is a good example, but one might as easily cite Hobbes.
[12] E.g., Reichenbach 1951, pp. 192–3; Grünbaum 1963, p. 312.

Being's plans and to the world being directed toward Its ends, or (in a pinch) to causes somehow working backward out of the future. (Ruse 1988, p. 44)

There you have everything teleology is usually thought to be – and that Aristotle never asserted.

Nor did Aristotle's 'Christianized followers' (an odd description of Christians who, in the thirteenth century, adopted some parts of Aristotle's metaphysics previously unknown in the Christian West) misunderstand him on this point. According to Aristotle's theology, God, the Unmoved Mover, accounts for an ordered cosmos (*Meta.* Λ, 6–10). But, as this god moves without moving, it is unlike the God of *Genesis*. It does not fashion a world through its own acts. Its actuality consists in thought: it is thought thinking of itself. And its influence is through its being a model of self-sufficiency, a perfection emulated in their deficient way by material things – hence the circular motion of the heavenly bodies. St. Thomas Aquinas *limited* the role of final causation in order to make room for the Christian God – an efficient, not a final cause. For Aquinas, final causes explain only the desires of conscious agents. Thus he was able to take the appearance of finality in unthinking nature as a proof of God's existence: there being no final causation in nature, nature's order must be due to an intelligent Creator.[13] As others have pointed out, that was a step toward modern science, as it entailed a mechanistic view of the natural world.

5. Hume's Ghost

So much for the confusions that have hidden classical teleology away from contemporary eyes. But might the charge of confusion not be returned, like a shuttlecock, to us? For is it not confusing to call a so-called final cause a 'cause'? If what 'cause' now means is a mechanical cause, then 'cause' is a mistranslation of Plato's and Aristotle's *aitia*, at least when that term and its cognates are used to denote other than mechanical causes. That view has been argued vigorously by classicists who take the logical empiricists' essentially Humean analysis of causality to be *the* concept of cause in ordinary English, and who then point out that the original meaning of *aitia* in ancient Greek was that of 'charge' or

[13] St. Thomas' *quinta via*, in his *Summa theologica* and *Summa contra gentiles*. See, e.g., Copleston 1962, pp. 63–4. In the *Summa contra gentiles*, I, 75, St. Thomas wrote, 'The causality of an end consists in this, that other things are desired for its sake' (Gilby, trans., Aquinas 1951, p. 65).

'accusation'.[14] From that juridical root, the word came to stand for moral responsibility and then for any sense of being responsible. Their conclusion is that Aristotle's four *aitiai* are four ways of answering 'why' questions and that, as not all such questions require a cause in the modern sense for an answer, it would be better not always to translate *aitia* by 'cause', and especially not when the term is being used teleologically. Gregory Vlastos suggested 'reason' as the proper generic term.

In answer to the charge, we may begin by noticing that the Latin source of our word 'cause', *causa*, was rooted, similarly to the Greek *aitia*, in juridical proceedings; it is related to such words as *accusatio*, the act of accusing, and *causare*, to plead a cause or bring an action. And that range and relation of meanings is retained in contemporary English usage, as the use of 'cause' and 'accusing' in the preceding sentence attests. One pleads a cause at court, makes common cause with the like-minded, devotes himself to a hopeless cause, accuses others of sundry failings and excuses his own. It would seem that 'cause' and *aitia* were both adapted, from a moral or juridical usage, to refer to mechanical causes, and for the same reason: namely, their connotation of 'being responsible' for something.

Aitia was so adapted by Plato in the *Phaedo*, the term not having been used in that way by the Presocratics, who, instead, often called their explanatory principles *archai*. Now, the word *archē* is also nonmechanistic in its root meaning, an *archon* being a ruler, king, and so on. A mechanical cause seems better designated by a word suggesting responsibility than by a word suggesting rule. Following Aristotle, Theophrastus and subsequent commentators regularly spoke of the *archai* of the Presocratics as *aitiai*. It became the standard term for such causes, just as the descendants of *causa* in English and other modern languages have become the standard terms for mechanical causes. 'Cause' would therefore seem to be the ideal translation of *aitia*. Both terms are used to refer to mechanical causes while retaining a broader meaning, enabling us to ask whether mechanical causes are the only causes, that is, the only things responsible for what happens, the only objective determinants of phenomena.[15]

[14] Wicksteed and Cornford 1957, pp. 126–7; Owen 1965; Vlastos 1969; Charlton 1970, pp. 98–100; Frede 1980; and Wiggins 1986.

[15] Thus William Wallace, in his two-volume study of causal explanation, says, with reference to Bunge 1959 and the thought of Claude Bernard, that Bunge, like Bernard, seeks out 'all the determiners of phenomena. A determiner, in his understanding, is not really different from the *aitia* of the Greeks or the *causa* of the medievals. . . . [T]hese terms

But do not philosophers today use 'cause' in such a technical sense of their own as makes the notion of a final cause an oxymoron, breeding nothing but confusion? While admitting that philosophers are perhaps easily confused and that one should take special care when addressing them, we must also remember that words' meanings embody doctrines and that to allow a usage to prevail unchallenged is sometimes the same as to allow a dogma to reign unchallenged. The view Plato and Aristotle were challenging is that material and mechanical causes (as we would call them) suffice to explain everything – the same view that prevails among philosophers today. The issue was not whether there are other modes of understanding (or ways of answering 'Why?' questions); it was whether there are additional objective determinants of becoming. The English word for such a determinant is 'cause'. To translate *aitia* by a term such as 'reason', which has primary reference to the mind rather than to the world (its Latin root refers to calculation, keeping accounts, justifying one's conduct, making plans, and the like), is to obscure the issue.[16]

There have indeed been modern defenses of teleological explanation that associate it with reasons given to justify, as opposed to citation of objective determinants. Kant's was the first. In his first *Critique*, he had already made the fundamental structure of nature, more or less Newtonian, to be a function of the categories and principles necessary to any possible understanding; hence, he concluded, it does not appertain to 'things in themselves'. Nature itself, then, is merely phenomenal. In his third *Critique*, the *Critique of Judgment*, he made a teleological account of nature equally inescapable and yet not constitutive of nature, even as merely phenomenal. Teleological judgment, he said, invokes an analogy to the purposeful action of an intelligent being, in order to apprehend nature as a system. But that, he added, can be done without presuming that purpose actually explains phenomena of nature: the judgment is 'reflective', not 'determinative' (*KdU*, §61). We apprehend nature as a system, or cannot help but attempt to do so, even while knowing that this answers to a cognitive need of ours without corresponding to anything in nature itself. But is that not to say that we knowingly hoodwink ourselves? And is that plausible?

were roughly equivalent, and only in the contemporary period has causality generally been narrowed to its Humean understanding' (Wallace 1972–4, vol. II, pp. 245–6). That narrowing, however, has been confined to philosophers and has not affected ordinary usage.

[16] Ronna Burger 1984, p. 252n2, and Paul Stern 1993, p. 208n45, make a similar point.

A similar sophistication may be found among more recent authors, who also (a) defend teleological explanation as a way of satisfying one or another intellectual need while (b) denying that there is any phenomenon not due exclusively to mechanical causes (deterministic or probabilistic). But discussion of those and other alternatives to a Peircean account of final causation is best delayed until after the latter has been examined. My sole purpose in this section has been to argue not that there is a genuine issue about whether there are final causes, but that the possibility of such an issue should not be allowed to be foreclosed by the simple, and false, claim that 'cause' *means* mechanical cause, à la Hume. If, instead, the word 'cause' means an objective factor responsible for what happens, then it becomes possible to ask whether causes are limited to particular forces, events, conditions, and so on or may in some cases be something else, such as a type of possible outcome.

However, laws binding causes to effects pertain exclusively to mechanical or efficient causes; hence, 'causal law' always refers to laws of efficient causation. There are no laws of final causation.

6. Ordinary Purposes

In ordinary speech, we use the word 'purpose' broadly. Certain customs serve a social purpose, though no one had consciously designed them for that purpose. A driver braked his car on purpose, though he did so automatically, his mind being on other things. The purpose of bees collecting nectar is to feed their larvae, the purpose of eyes is to see. Most philosophers who have discussed the topic regard some or all such statements as either mistaken or figurative. But all of the types illustrated are run-of-the-mill, common expressions, not unusual, literary, or fanciful. That raises an interesting question: on what basis can one assert that the literal meaning of a word is narrower than its standard usage?

That basis can only be that one's ability to make literal sense of standard uses is limited to just some thereof. If the only literal meaning of term A that we can come up with is B, which does not apply to certain cases C to which A is applied, then A must apply to C figuratively (at best) and not literally. But, then, if a broader meaning can be suggested – one that assigns less of ordinary usage to the merely figurative – it would seem to be the more plausible analysis of that usage.

Now, we have already determined, as a matter merely of grammar, that a purpose is a type of outcome. We may add to this that it is a type regarded as explanatory; for that is how the word is used. Why did you

do that? Why do we have eyes? In either case, a purpose is cited as an answer. So far, we have no reason to suppose that the one use of the word 'purpose' is literal, the other figurative. In either case, a type of outcome is cited to explain the phenomenon, and that, perhaps, is all that the word 'purpose' means, namely, a type of outcome that is explanatory.

The reason this conclusion seems problematic to philosophers is that they cannot figure out how a type of outcome could be explanatory, except in cases where there is a conscious agent who is thinking of that type and desires, or otherwise chooses to seek or to create, something of that type. Their conclusion is thus that any talk of purpose where there is no such agent is either mistaken – the mistake being to suppose that there is such an agent where none exists – or figurative. One who says that we have eyes to see with either believes that God created us, and gave us eyes because He wanted us to see, or believes that we have eyes *as if* a beneficent creator had given us eyes to see with.

But there is another possibility. It is that ordinary speakers have no idea, in some cases, how a type of outcome is explanatory, and, not being philosophers, they do not care. They merely assume that it is, because otherwise much that seems organized to achieve ends – animal behavior, organic life, and so on – would be impossibly fortuitous. It is not necessary to suppose that ordinary speakers have a complete theory about how types of outcome explain phenomena. It is not necessary to suppose that they either believe in the theory of *Genesis* or speak figuratively. They may be speaking vaguely (in the sense of 'vague' defined in chapter 10, section 3) on the basis of presuppositions of which they are not explicitly aware and could not begin to fill out in detail or to justify.

Philosophers' business is theory, and, by a kind of *déformation professionel*, they insist on finding theory in the innocent utterances of their fellow citizens. But, truth to tell, philosophers themselves do not have a very good idea of how conscious purposefulness works. Nor is there agreement among physicists about how gravity works. That does not stop us from employing such notions. Why, then, expect ordinary folk to be more exacting in their modes of expression?

Darwin did not banish purpose from the organic world. Rather, he showed how types of outcome can be explanatory even without there being conscious selection of means to outcomes of those types. In place of conscious selection he proposed what he called 'natural selection', that is, selection that is not made by any conscious agent. Organic features are selected *in* a process that selects *for* certain types of outcome; the features selected thus tend to be of the types (or to have effects of the

types) selected-for (chapter 5, section 3). But the upshot is that the types selected-for are part of the explanation of the organic features selected, just as the aims of a human actor are part of the explanation of what he does. Darwin thereby justified a mode of discourse justified otherwise by the theory of *Genesis* and which, more importantly, has flourished and would continue to flourish even without our knowing how to justify it.

So, is that what 'purpose' means, and means literally: (1) a type of outcome (2) that explains why there are outcomes of that type? I think that that will not quite do. In the next chapter (section 1), we encounter types of outcome that are explanatory not because they are selected-for but because they encompass a preponderance of equiprobable alternatives, and these seem not properly described as 'purposes'. Thus: a purpose must be a type of outcome that is explanatory (3) because it is selected-for. But in the same chapter (section 4), we encounter types selected-for that we are also reluctant to call 'purposes'. The type for which they are selected is not a type of effect they have but is an organization that they themselves exemplify. Thus: a purpose (4) must be a type of effect for which something is selected as a means. However, we also speak of the purposes for which agents act – about which, more in a moment. Let us say, then, that a purpose is a type of outcome for which an agent acts or for which something was selected as a means.

That, I suggest, is the literal meaning of 'purpose' implicit in ordinary usage: it accounts for the breadth of ordinary usage without making any part of that breadth figurative or mistaken. But this definition is satisfied by the effects for which organic features are naturally selected. That explains why the wide acceptance of Darwin's theory has not altered our habit of speaking about organic features as having purposes.

If all of that is correct, then philosophers, including Peirce, are mistaken in thinking that 'purpose' literally means a type consciously sought. 'A purpose is merely that form of final cause that is most familiar to our experience' (1.211), he said. But that is simply the long-enduring 'folk etymology' of those very strange folk, the philosophers. The idea that every purpose must be conscious is perhaps the lingering death-grip on us of an outmoded Cartesian dogma, the division of being into a *res extensa* and a *res cogitans*.

The reason I am taking time to explicate this bit of ordinary language is two-fold. It is simpler and more natural to speak of purposes than of final causes, types of outcome, and the like, and, therefore, I wish to be allowed to do so. More importantly, ordinary usage of 'purpose' and its cognates brings with it some subtle distinctions that are of fundamental

importance to the argument of later chapters. Let us preview those distinctions now; they concern having a purpose, acting for a purpose (or acting purposefully), serving a purpose, and being used for a purpose.

'Having a purpose' applies broadly to existence, occurrence, performance, placement, and so on. Take placement: a rock, we suppose, exists for no purpose, yet it may be on my desk for a purpose, namely, to weigh down papers. Its placement was selected for that effect. The heart exists for a purpose, which is to pump blood, thus to circulate blood, thus to bring nutrients to cells, and so on: effects of one type were selected for those of the next. And thus the heart's pumping also has a purpose, which is to circulate blood, and so on. But that action is perfectly mechanical, and, for that reason, we deny that the heart acts purposefully (or for a purpose). Having a purpose and acting for a purpose are distinct.

Purposeful action may appear mechanical, but it is subject to modification if it fails its purpose. New trials must be made, perhaps randomly, until something that works is found. Potential selection for a type of effect is built into purposeful action. But, in addition, to be purposeful, an action must be of a mode selected for the type of effect it selects for. And that selection must either have been natural, in the evolution of the agent's species, or by the agent itself (in that case, we speak of the agent as choosing the purpose for which it acts). For we are disinclined to admit the artifacts designed to select for a type of outcome act purposefully (why we are thus disinclined is another question; for the present, it suffices that we are).

Being used for a purpose presupposes purposeful action; the thing so used does not necessarily exist for that, or for any, purpose. Serving a purpose presupposes either purposeful action or something that exists for a purpose, to the existence or operation of which it contributes. And, again, what serves a purpose, though it might exist or occur for that purpose, need not do so. The meteorite that struck down the aged, wealthy husband served the purpose for which his young, restless wife was preparing the arsenic; but it is doubtful that it fell for that purpose. The sensitivity of certain chemicals to light serves the purpose for which eyes exist and thus explains why those chemicals are found in the rod and cone cells of retinas, but those chemicals do not have such properties for that or, so far as we know, for any purpose.

Existing for a purpose is sometimes derivative from acting for a purpose: thus, artifacts. The derivation is not simple. The joiner constructs a chair in order to sell it and buy his dinner, but those are not the purposes for which chairs exist. The purpose of a chair is, rather,

that purpose of potential users for which it was purposefully made, by the joiner, to serve. If the account in *Genesis* is correct, then all the organic features that exist for a purpose are derivative from God's having acted purposefully and we are God's artifacts (His creatures). In that case, all existing for a purpose is artifactual. If Darwin's account is correct, and correctly glossed in the language of purpose, then some cases of existing for a purpose are primary, not artifactual. The capacity of animals to act purposefully is explained by features they have that exist for that very purpose, that is, to produce purposeful action.

7. The Mysterious Case of the Surplus Body

I refer, of course, to bodies of explanation. There are two such bodies named 'statistical' (sometimes 'probabilistic'), differing fundamentally in their structure, where the name suggests that there is only one. The extra body is therefore rarely noticed.

The leading accounts that philosophers, including Peirce, have given of statistical explanation all agree, amazingly, in ignoring the special character of explanation in statistical mechanics (Railton, *vide infra*, excepted).[17] Contemporary discussions of this topic begin with the 'covering law' model due to Hempel (1962), and it is convenient for us to begin there, too.

According to Hempel, statistical explanation is an inference, of a statement of the phenomenon to be explained, from premisses that include a statement of at least one probabilistic law. Such a law, in the simplest case, is of the form $P(E/C) = p$, where p is a real number, $0 \leq p \leq 1$, and $P(E/C)$ is the probability of E conditional on C obtaining. Thus, from a premiss that C and a premiss that $P(E/C) = p$, we can infer that E, the inference having a probability p of being correct, relative to the information supplied in the premisses.

Others deny that explanation is inference, and deal far differently from Hempel with the problem of the reference class (i.e., how C is to

[17] The literature on statistical explanation should not be confused with the much larger and more sophisticated literature on statistical inference. Although the former sometimes draws on the latter, the latter is not concerned with what constitutes explanation; it is concerned with the meaning of probability and the different kinds of inference that different theories of probability justify. Those issues are central to foundational debates in statistical mechanics, but relatively little of the literature on statistical inference deals with statistical mechanics, von Mises (1957 [1928]) and Jeffreys (1973 [1931]) being major exceptions.

be chosen) and the question of how, if at all, assessing the probability of an outcome either contributes to or is implicated in its explanation. Richard Jeffrey (1970) pointed out that assigning an explanandum a high probability does not always explain it and, more surprisingly, that some good explanations accord the explanandum a low probability. If things of a certain kind happen by chance (e.g., the decay of an atom of a radioactive element in a given period of time), then that is how a thing of that kind happens, even if its chance was low. But all agree that statistical explanations presuppose probabilistic laws of the sort described – laws that enable us to derive a probability for a given outcome from known conditions.

And that fails utterly to capture the reasoning in statistical mechanics, in which the laws assumed may be deterministic, not probabilistic, and in which the initial conditions are unknown. A typical example is the explanation of the evolution of an enclosed system of gas molecules from being less equally to being more equally distributed. There are trillions of molecules in even a cubic centimeter of gas, and any observable distribution of them (a macrostate) would be constituted by any of an enormous number of alternative arrangements (microstates) of the molecules. We must therefore remain ignorant of the actual arrangements, including the initial arrangement. Statistical reasoning shows nonetheless that the chances overwhelmingly favor changes from less to more equal distribution, until near-equality is reached. Rather than deriving a probability from known conditions, the movement of thought in statistical mechanics is almost the polar opposite: from ignorance of initial conditions (at the microlevel) to virtual certainty about the outcome (at the macrolevel).[18]

How is it possible that this obvious point has been overlooked or, at least, neglected? The explanation, in part, is that the formalism $P(E/C) = p$ does not distinguish between the two kinds of case. Thus, Hempel cited a range of examples, from radioactive decay to rolling dice, without distinguishing probabilistic dependence of an outcome on known conditions (radioactive decay) from ignorance of conditions possibly deterministic (rolling dice) (1962, pp. 121–2).

Perhaps the best-known alternatives to Hempel's model are two due to the late Wesley Salmon. In his 'statistical relevance' model, a fact explains an event if, putting it perhaps too simply, it accords it a probability that would be unaltered by any further fact other than that of the event itself (1970). Salmon is quite clear that such assessments of probability

[18] See chapter 5, section 1, for a more detailed discussion of this example.

presuppose probabilistic laws (1970). No distinction is made between different forms of statistical explanation.[19] Salmon's later, 'causal/ mechanical' model retains the idea of statistical relevance but adds to it the requirement of a causal/mechanical account, that is, some idea of the processes that lead probabilistically to the explained result (1984, p. 22). Here, at last, it becomes possible to make the needed distinction between mechanistic and other forms of statistical explanation, but Salmon did not make it. Instead, it is James Woodward who pointed out that the causal/mechanical model is not satisfied in statistical mechanics. For in that science, Woodward says, 'one abstracts radically from details of such individual causal processes and focuses on finding a way of representing the aggregate behavior of molecules' (Woodward 1989, p. 363).

Peter Railton's is the only model of statistical explanation – or probabilistic explanation, in his preferred term – that I know of, that explicitly excludes statistical mechanics and thereby makes the needed distinction (Railton 1977). He calls the explanations he models 'deductive-nomological-probabilistic' (D-N-P) explanations. We can skip over the reason Railton gives why they are deductive. The thesis germane to our interest is that D-N-P explanations must be causal (in Railton's sense of 'causal', i.e., they must cite mechanisms) as well as probabilistic: they are 'unsatisfactory unless we can back them up with an account of the mechanism(s) at work' (p. 208). Such a mechanism, not being deterministic, is a 'chance mechanism', and thus the model is restricted to genuinely indeterministic processes:

> It is widely believed that the probabilities associated with standard gambling devices, classical thermodynamics, actuarial tables, weather forecasting, etc., arise not from any underlying physical indeterminism, but from an unknown or uncontrolled scatter of initial conditions. If this is right, then D-N-P explanation would be inapplicable to these phenomena even though they are among the most familiar objects of probabilistic explanation. I do not, however, find this troublesome: if something does not happen by chance, it cannot be explained by chance. (p. 223)

[19] It is remarkable that Salmon, like Hempel and many other authors in this field, paid so little attention, at least in this context, to the explanations yielded by statistical mechanics. Salmon cited statistical mechanics several times, but always briefly and with respect to examples in which thermodynamic laws interpreted probabilistically are assumed, not explained (1970, pp. 209ff.; 1984, pp. 26, 180–1; 1998, p. 151). But the chief glory of statistical mechanics is its explanation of those laws: a fact of which these same authors were well aware and that has been much discussed in another context by philosophers of science, where the topic is theoretical 'reduction'.

It might be objected that explaining something as being due to chance is no explanation at all; but Railton's thought is that it is the chance mechanism that explains its effects, probabilistically, and not chance per se.[20]

Railton continues, 'What must be given up is the idea that *explanations* can be based on probabilities that have no role in bringing the world's explananda about' (p. 223, emphasis in original).[21] But that is an amazing claim. For, surely, statistical mechanics, even in its early, Newtonian phase, has provided some of the most impressive and successful explanations in modern science. They have been called 'explanations' and have been accepted as such and felt to be explanatory. To deny that they are explanations is, in effect, to impose one's narrower definition in lieu of a broader, established use of the term. Far better, I think, to admit that explanation takes different forms.

Explanations we may call 'statistical' fall into two classes. Those of the one class are the explanations that Hempel, Salmon, Railton, et al. evidently had in mind when they offered their models of statistical or probabilistic explanation. They are mechanistic in kind; for they all seek to explain particular outcomes by citing particular conditions. By our generous definition of 'mechanistic', they are so even if they lack the specification of mechanisms that Railton and Salmon say is essential to a complete mechanistic explanation. Let us follow Railton, but without insisting that mechanisms be specified, in naming this subclass of statistical explanation 'probabilistic explanation'.

The other class of statistical explanation is not mechanistic; at least, that is what I argue in some detail in the next chapter with respect, first, to statistical mechanics and, next, to natural selection. That of course contradicts the standard view that both sciences are mechanistic. Neglected by logicians, this latter form of statistical explanation needs a name. As the explananda of these explanations are anisotropic processes, that is, the (practically) irreversible evolution of systems toward final states or toward new states, let us call them 'anisotropic explanations'.

[20] This is in line with the growing tendency of philosophers to acknowledge a probabilistic causality (Suppes 1970; Fetzer and Nute 1979), in contrast to the past fashion, of claiming that quantum mechanics spelled causality's demise. I believe that Peirce would have welcomed this development, as witness his later theory of probability as a disposition (8.225, 2.664–5; cf. Burks 1964), which anticipated Popper's well-known propensity theory (Popper 1959) that Railton exploits. Deterministic causality may then be seen as a special case of probabilistic causality (where p=1).

[21] In a later paper, Railton acknowledges that classical, i.e., Newtonian, statistical mechanics is explanatory, but only as it bears on an ideal but unavailable explanation that, in this case, would be deterministic (1981, pp. 249–52).

The division of explanations between those that are deterministic ('deductive nomological', in Hempel's phrase) and those that are statistical is far less fundamental, I suggest, than is the distinction between mechanistic explanations, whether deterministic or probabilistic, on the one hand, and anisotropic explanations, on the other. The use of statistics is so different in probabilistic and anisotropic explanations, respectively, that the two being lumped together as 'statistical' is misleading though not incorrect. In the next chapter, I argue that teleological explanation is a subclass of anisotropic explanation. Teleological explanation is distinct from explanation in statistical mechanics, yet the two are far closer in nature to one another than either is to probabilistic explanation.

The failure to recognize that explanation in statistical mechanics is not probabilistic (in the sense we are now giving to the latter term), but is of another form altogether, has kept that form of explanation from being recognized. And, as teleological explanation is, at bottom, of that form, it has kept teleological explanation under its historic cloud of suspicion.

5

Final Causation

Does teleology have a future? On the one hand, final causation would seem to be excluded by modern science. On the other hand, there is Peirce's contrary view. Peirce's scattered remarks of various dates, taken together, suggest (a) that types of outcome play an explanatory role in some sciences and (b) that this is the key to making final causation intelligible. In the first four sections, we develop thesis (a), in the remaining three, thesis (b).

1. Explanation in Statistical Mechanics

Peirce appears to have been inspired by two innovations in nineteenth-century science: the molecular/kinetic theory of gases and heat and Darwin's theory of natural selection. As early as 1877, he wrote,

> Mr. Darwin proposed to apply the statistical method to biology. The same thing had been done in a widely different branch of science, the theory of gases. Though unable to say what the movements of any particular molecule of a gas would be . . . , Clausius and Maxwell were yet able, by the application of the doctrine of probabilities, to predict that in the long run such and such a proportion of the molecules would, under given circumstances, acquire such and such velocities. . . . In like manner, Darwin, while unable to say what the operation of variation and natural selection in any individual case will be, demonstrated that in the long run they will adapt animals to their circumstances. (W3:244)

Notice that two things are being claimed about each of these sciences. One is that the phenomena studied in each are processes that have a direction: each process tends – for all practical purposes irreversibly – toward results of a definite type. Later, in 1898, Peirce named processes

having this feature 'finious', as tending to bring about a final state of things (7.471). 'Final', however, is not *le mot juste*, since each organic adaptation makes further adaptations possible, by some of which it may be modified or made obsolete. And therefore 'finious' is less accurate than is 'anisotropic', which suggests direction but not finality. The other claim is that these phenomena are explained statistically.

We begin with statistical reasoning in statistical mechanics. Statistical reasoning in general and statistical mechanics in particular are highly technical subjects about which there continues to be a number of difficult issues and serious controversies.[1] What I will say is nonetheless very simple and consists in pointing out what is obvious but often ignored. I believe we can skirt the technical issues and controversies, while drawing a conclusion controversial in its own way.

If a gas is released into a container, it soon diffuses throughout that container; it does not subsequently collect together again in one or another corner. Liquids stirred together mix; stirring in reverse never separates them. If one end of an iron bar is heated and the bar is then left alone, the heat eventually spreads evenly over the whole bar and does not subsequently concentrate in any one part of it. The phenomenological Second Law of thermodynamics is a generalization of the familiar fact that heat flows spontaneously in but one direction, toward its equal distribution. In one formulation, it says that entropy, as a measure of the unavailability of energy to do work, grows irreversibly toward a maximum in any system closed to external influence. (Heat is available to do work only when it is unequally distributed, hence, when it can flow.) These phenomena and laws all have a direction: they are anisotropic.

By contrast, Newton's laws of motion, Maxwell's equations, Schrödinger's wave equation (or equivalent formalisms in quantum mechanics) are reversible: they are equally satisfied by a process run forward or backward. In that respect, all are like the theory of gear wheels. If gear A turned clockwise turns gear B counterclockwise, then B turned clockwise will turn A counterclockwise. The machine, reversed, obeys the

[1] Peirce wrote extensively and with great originality and power on statistical inference and statistical explanation, anticipating many of the more recent doctrines (see Niiniluoto 1993 for an accurate overview), but we have no need to discuss those contributions here. He appears not to have been deeply acquainted with the debates in statistical mechanics that began in his lifetime. As his scattered comments on that subject tend toward the technical but are framed in terms no longer generally understood, I have made less use of them than I otherwise might have. See Reynolds 2002, ch. 2, for a general account of Peirce's understanding of the kinetic theory; also, RLT, chs. 6 and 7 and Putnam's commentary thereon, RLT: 78–94.

same law. So, also, if any process governed by the aforementioned laws occurs in nature, there is no reason why it must always run in but one direction. And thus it would seem that anisotropic phenomena could not be explained by those laws, hence, neither by Newtonian nor by quantum mechanics. And, yet, they have been explained on the assumption of those laws, statistically.

When the Second Law was explained statistically, it was reinterpreted as a probability. For most systems of any interest, the probabilities it entails are very near either to 1 or to 0. We do not expect ever actually to observe heat flowing spontaneously from a colder to a warmer region. Nevertheless, the Second Law, thus reinterpreted, is consistent with a reverse flow of heat, however improbable that flow may be. Some have thought that the Second Law was thereby reconciled with reversibility in mechanics. This is a complex and much-discussed topic (see Hollinger and Zenzen 1985 for distinctions among several kinds of 'irreversibility'). Notice, however, that a low probability (i.e., an improbability) of reversion suffices to define direction: the direction is the one that is dominant. Hence, a probabilistic law may be anisotropic; if you wish, it is anisotropic to a degree. There is still a question about how such a law can be explained by laws that are not at all anisotropic.

Let us look in an informal way at the statistical explanation of the processes the Second Law describes. That explanation presupposes the molecular theory of matter and kinetic theory of heat. On the molecular theory of matter, there are trillions of molecules in a cubic centimeter of gas at normal pressure and temperature.[2] Suppose that S is a system of n molecules contained in a volume V spatially divided into a large number, m, of small but equal cells, where n is much larger than m. We can define 'a fairly even distribution' of S over V as consisting in there being $n/m \pm e$ molecules in each cell, where e is a whole number much smaller than n/m but much larger than zero. That type of distribution defines a class, C, that contains a great many alternative possible arrangements of the individual molecules of S. The complement of C, C', is the class of all other possible arrangements in V of those same molecules. Any arrangement belonging to C would constitute a fairly even distribution, and any arrangement belonging to C' would constitute an uneven distribution. Let us suppose that we can observe in most cases whether S is evenly or unevenly distributed (allowing for a broad range of borderline cases

[2] Billions in British and German numeration, trillions in American and French; in any case, now known to be on the order of 2.7×10^{19} molecules per cm^3 at one atmosphere of pressure and 0°C.

difficult to decide). Call the objects of those observations 'macrostates' of *S*. Call the particular molecular arrangements 'microstates'. Microstates cannot be observed and, thus, in observing a macrostate, we cannot determine which is the microstate that, of the many possible microstates that could constitute it, actually does constitute it. But any actual macrostate is constituted by exactly one microstate.

If *e* is not too small, *C* will contain a great many more possible microstates than does *C′*. So also, the class of all possible molecular motions that would lead from an initial uneven distribution of *S* over *V* to a more even distribution – we may call that class of motions *D* – will be enormously larger than its complementary class, *D′*, of motions that would sustain the original distribution or lead to other distributions no more or still less even. (We may speak here of macro- and microprocesses, on analogy with macro- and microstates.) At any given moment, then, the chance that *S* will remain unevenly distributed or become less evenly distributed is small. That improbability increases geometrically over finite periods of time; in cases of practical interest, where *n*, at a minimum, is in the trillions, the time period, at a minimum, is several seconds, and the temperature is not close to absolute zero, the probability is close to 1 that *S*, if unequally distributed, will become more equally distributed. Observed macroprocesses will thus tend to have a direction, which is toward even distribution. This theoretical conclusion, which is based wholly on statistical reasoning, given the basic assumptions of the molecular theory of matter, agrees with observed phenomena and explains them.

On the kinetic theory of heat, which identifies nonradiant and nonlatent heat with the random translational or vibrational kinetic energies of molecules, the same kind of reasoning explains the Second Law, amended as a probability. In it, entropy is interpreted as a measure of molecular disorder; order would consist, for example, in an uneven distribution of molecular kinetic energies, so that one part of the system can be observed to be warmer than another. In a thermodynamically closed system (a system in which neither matter nor energy either enters or leaves), the possible sequences of combinations of molecular events – possible microprocesses – that break down order, or that preserve disorder, vastly outnumber those possible microprocesses that do not. Therefore, in such a system, if it contains more than a few molecules, entropy can be expected to grow toward a maximum and then to stay near that maximum, with tiny fluctuations away from it, generally unobservable. Thus we never in any easily observed closed system notice a flow of heat from a cooler to a warmer region.

That is the basic idea. Actual calculations of the probabilities indicated for systems of given parameters is enormously complex, but that need not concern us. It is necessary, however, to acknowledge some difficulties in and historical developments of statistical mechanics. Initially, probability was introduced in this way: we do not know the exact microstate of a system at a given time, and, so, having no reason to suppose one possible microstate less likely than another, we count them all as equally probable. All that are consistent with an observed macrostate are equiprobable – even though we know that exactly one (but not which one) is actual! Probability, in that case, is a measure of our ignorance. But can a law of nature be based on ignorance?

In the effort to give physical meaning to probability, such devices were proposed as the idea of ensembles of systems (associated mainly with Gibbs' statistical mechanics, but the idea goes back to Boltzmann and Maxwell) and Boltzmann's ergodic hypothesis. The latter states (as currently formulated, though this is often distinguished as quasi-ergodicity) that over time an isolated system will pass arbitrarily close to each point on its energy surface in phase space. (A point on a system's energy surface in phase space is a microstate compatible with given macroscopic parameters, such as the system's total energy.) In the ensemble approach, probability is interpreted as a measure across many systems. On the ergodic hypothesis, probability can be interpreted as a measure across many times in the evolution of one system: that two microstates are equally probable means that they occur, to some degree of approximation, equally often.

It is not clear that either strategy succeeds in eliminating the role played by ignorance. The reasoning by which the Second Law and related phenomena are explained still seems to depend on our not knowing which possible system our system is or through which stages it passes during the given period. These questions, related paradoxes, and other foundational problems have been the subject of debate among physicists from the beginning of statistical mechanics.[3]

Quantum mechanics changed the scene dramatically. As it proposes fundamental laws that are probabilistic, it presents chance as being an objective feature of reality. Probability is therefore given a physical meaning directly, determined not by our ignorance but by physical law. Whether the quantum reformulation of statistical mechanics eliminates all reliance on ignorance is not as clear as one might wish. (David Albert (1994 and 2000, pp. 150–62) argues that it depends on which

[3] For general accounts, see Ehrenfest and Ehrenfest 1958 [1912]; Sklar 1993; von Plato 1994, ch. 3.

interpretation of quantum mechanics we adopt.) Suppose, however, that it does; for that is the hypothesis least favorable to the argument I shall make. Suppose that quantum indeterminacy in the development of a system makes it necessary to consider all possible microprocesses, even over a brief period, and that it would do so even if we had known what the initial microstate of the system was. Then our actual ignorance of that microstate is washed out: it makes no difference. If that is correct, then the equiprobablity of alternative courses of development is determined by the laws of quantum mechanics alone, and not by our ignorance.

The reasoning characteristic of statistical mechanics nevertheless remains the same, at least so far as that reasoning concerns us. To be sure, there are enormous differences. When the probability of a macrostate was the number of microstates compatible with it divided by the number of all microstates compatible with either it or its complement, then probability was a rational number and its calculation was combinatorial. With the ensemble approach and the ergodic hypothesis, finite sets of alternatives were replaced by integration over continuous differences. Thus, probabilities became real rather than rational numbers and the principles of their calculation migrated from combinatorics to measure theory. In short, the mathematics of statistical mechanics, already complex, was transformed and made more sophisticated. Then quantum statistical mechanics reintroduced an element of discreteness, in the energy levels of a system, and again changed the mathematics employed. But none of that affects this point: a macroprocess (or ensemble, ergodic set, etc.) is still explained by its being a far more inclusive type of alternative microprocess (system, path in phase space) than is its macroscopic complement. A change in the basis on which those alternatives are deemed equiprobable makes no difference to the logic of the ensuing explanation, no matter how much the detailed calculations are affected.

That is an obvious point and not disputed. And here is an equally obvious observation – also utterly unoriginal – that tends to be forgotten: a statistical explanation of this type refers to no forces or other particulars as accounting for the phenomenon explained. In Peirce's words,

> Take, for example, the phenomenon of the diffusion of gases. Force has very little to do with it, the molecules not being appreciably under the influence of forces. The result is due to the statistics of the equal masses, the positions, and the motions of the molecules. (NEM4:66)

Erwin Schrödinger made the same point in a popular lecture delivered in 1944. Referring to the diffusion of a small amount of potassium

permanganate in water, he said that the phenomenon 'is in no way due, as one might think, to any tendency or force driving the permanganate molecules away from the more crowded region to the less crowded one' (1967 [1944], p. 15). It is not as if something was shoving them along. Rather, it is a statistical fact that their random collisions and their otherwise mutually independent motions – in all directions – will add up, in short order, to uniform dispersion.

But if no particular conditions, forces, events, and so on account for the direction of such processes toward uniform dispersion (disorder), then that phenomenon is not explained mechanistically, as we have defined that term (chapter 4, section 4). Much less is the explanation mechanistic by narrower definitions that require the identification of a mechanism that does the job. The explanations that we have informally described are statistical but not in the sense of deriving a probability of an outcome by applying a probabilistic law to known conditions; that is, they are not mechanistic-statistical explanations or what we have named 'probabilistic explanation' (chapter 4, section 7). They are a quite different kind of statistical explanation, appropriately named 'anisotropic', as their explananda appear always to be anisotropic processes or anisotropic laws.

Complex as these anisotropic explanations are in quantitative detail, in their most abstract formulation they are simple: anisotropic processes are explained as being due to the type of process they are. We can expect an evolution of closed systems toward internal disorder because such an evolution is far more likely than is its alternative. The one alternative is far more likely than the other, because of what each is, and not because of any forces exerted.

If we want to employ causal language, by which to specify what is responsible for the direction of such processes (using 'cause' for a determinant, i.e., for that which is responsible; chapter 4, section 5), we could not say that its cause is anything mechanical. Instead, that which is responsible for the phenomenon is a type, namely, its type. The cause of the phenomenon is the type that it fulfills; it is the possibility that it actualizes. Nothing depends on choosing to employ this causal language; we can get along without it. But neither is there anything wrong with employing it, and it signals a surprising similarity of explanation in statistical mechanics to traditional talk of final causes, which also are types that explain their instances (chapter 4, section 3).

Final causes are traditionally cited to explain the emergence of order out of chaos, whereas the causes now under examination explain the

opposite tendency, toward disorder. Therefore, it is better not to call them 'final causes', nor should we identify anisotropic explanation in statistical mechanics as 'teleological'. We discuss the latter terms in section 5. However, explanation in statistical mechanics counts as a first departure from mechanistic explanation, a step in the direction of teleological explanation.

To that conclusion there is an obvious objection – one that may seem decisive if it is not examined. It is this: the only principles that statistical mechanics introduces, in addition to the mechanistic laws governing molecular motions, are theorems of pure mathematics, specifically, of probability theory. That is comparable to the use of the calculus in Newtonian mechanics. In either science, then, the explanatory principles are all mechanistic; the use of mathematics is simply to calculate the consequences of those principles when they are applied to cases of various description. From this point of view, the distinctness of the form of explanation in statistical mechanics, which we have been at pains to establish, seems to evaporate.

The objection ignores a crucial fact, that the application of probability theory requires that a subject first be described in statistical terms, for example, in terms of macrostates, ensembles, or ergodicity. This introduces a different way of looking at the world, a statistical way rather than a mechanistic way. The molecular/kinetic theory makes that possible but does not entail it. It is the statistical view, initially taken in an act of genius, that makes statistical calculation possible and thus is fundamental to statistical mechanics. Probability theory becomes relevant only thereby. The point is the same as that which must be made to defend Darwin's theory from the notorious charge that it reduces to a tautology. Yes, at the heart of the idea of natural selection there lies a tautology, also statistical (see section 3). But its applicability requires marshaling a great variety of facts (Darwin's achievement) under concepts we owe to Darwin's genius (or to his inspired borrowing from Malthus).

2. Reflections on the Preceding

An observed macrostate or process is constituted by an unobserved microstate or process. That may seem to be impossible. How can what is observed *be* something not observed? And, yet, that is the case with all observation. The strawberry at which I am looking is composed of cells, molecules, and so on, of which I know only by hearsay. As Aristotle pointed out, the object of observation is particular but the observation's

content is general. What is observed is observed as *a* particular (at a given location) *of* a given type. The full concreteness and exact characteristics of that particular escape observation. Hence, the same observation would have resulted from observing a different entity that varied from the given entity in unobserved respects.

As with observation, so with explanation: we never explain a particular outcome in its full concreteness; we explain it only as *a* particular *of* a given type. In mechanistic explanation, such particulars are related by a general law, that is, a law that specifies a spatio-temporal order among particulars of one type and particulars of other types. The same particular may therefore be subject to many different explanations, each pertaining to a different aspect of it.[4] A phenomenon, as something to be explained, is always an abstraction; it is real, yet it is selected by description from a more complex reality.

Now, it is sometimes assumed (e.g., Nagel 1961, p. 344; Railton 1981, p. 251; and even by James Clerk Maxwell: Sambursky 1974, p. 445) that statistical methods are resorted to in physics only in lieu of more exact methods that are unavailable to us. But that is false. Even were more exact methods available, they would not suffice. For they would not explain what is explained statistically, which is a different aspect of reality, a different phenomenon. Suppose that we were able to detect the exact microstate of a closed system S at time T and, again, its microstate at a later time T'. The latter state would be explained mechanistically from the former state, whether deterministically by Newtonian mechanics or probabilistically by quantum mechanics. But that would *not* be an explanation better than the statistical explanation that is available to us; for it explains a different fact. To see this, let us perform a thought experiment.

Imagine that we are Laplacean demons, somewhat modified.[5] Imagine our possessing fantastic powers of observation and calculation but only a very limited grasp of statistics. The world is assumed to be Newtonian. We can observe each of the trillions of molecules of known uniform mass in a thermally isolated, fixed volume of gas at time T, determine each of their locations and measure their velocities at T, and then

[4] Peirce said that this is generally understood, and then castigated Mill for not having understood it: EP2:315. It has more recently been emphasized by Carl Hempel (1965, ch. 1).

[5] Elliott Sober puts the demon to work in what seems to me essentially the same argument, though more circumspectly stated, with respect to evolutionary theory; the conclusion he comes to is weaker, I think unnecessarily (Sober 1984, pp. 127–134). See also chapter 11, section 1, regarding an argument Daniel Dennett makes with respect to psychology.

compute, in an instant, all of their subsequent motions, collisions, and exchanges of energy up to some later time, T'. That way, we have predicted and explained the microstate of the system at T', from knowledge of its microstate at T.

Suppose that, initially, the system was unequally heated. We still do not know such a fact as that in that period it became equally heated. For that is a statistical fact that cannot be apprehended without the use of statistical ideas and measures. But let us suppose that we can divide the space of the system into cells and calculate (from our detailed knowledge of molecular momenta) the sum of the translational kinetic energies in each cell, subtract each sum from every other and compute the average of those differences, both at T and at T', noting that the average difference is much lower at T' than at T. Thus we can note that heat flows, from T to T', toward more equal distribution. Thereby, we will also have explained that flow.

But we have done so only with respect to this system in the one period examined. Our calculations give us no reason to suppose that the same result will continue to hold for this system or that it holds for other systems. Our curiosity piqued, we continue our calculations up to some later time T'' (let it be five years later) and also we make similar calculations from millions more initial microstates, observed or supposed. (Our powers of calculation are stupendous.) Almost certainly, we discover that the flow of heat is not reversed from T' to T'' and that heat flows in the same direction in every other system considered. If there are exceptions, they will be few and brief. But that result, arrived at in that way, must seem to us an amazing coincidence. Why does this same direction of flow occur almost without exception in the systems we investigate? We demons gaze at one another with wild surmise.

Demonic powers of calculation are not a substitute for the statistical intelligence of a Maxwell, Gibbs, or Boltzmann. It is only by statistical reasoning that the universal tendency of heat to flow spontaneously toward equality can be explained. Only in that way does it cease to be surprising. Such reasoning therefore explains something – the Second Law – that cannot be explained mechanistically. Statistical mechanics is not the desperate resort of nondemons. Even demons could not do better than it; if statistically challenged, they could not do as well.

The mechanical steps by which one microstate of the system S evolves into another is one aspect of the reality of S. This same evolution constitutes an evolution from one macrostate to another. That the latter evolution conforms to the thermodynamic Second Law is a distinct aspect of S,

accounted for by statistical considerations. These two aspects are in reality inseparable but they are nonetheless distinct. And therefore they are explananda of different explanations. As it happens, those explanations are markedly different in kind, one being mechanistic, the other not.

The mechanistic aspect of *S*'s evolution is particular; its statistical aspect is general. The explanation of the former is of course in terms of the general laws of mechanics, but they have to be applied to the particular microstates of *S*, which only demons can know. The explanation of the latter is really an explanation of the general case of which the given case is an instance, and thus it requires no knowledge of what distinguishes this case from others (except for its location). Now we come to a point that may have been the source of some confusion.

Once the Second Law has been established, it can be applied to particular cases to explain or to predict the direction they take. The direction is an effect, via that law, of the given initial conditions; the law is a causal law, relating particular cause to particular effect, and, like all causal laws, it is mechanistic (by our definition) even if probabilistic. And therefore the application of this law to particular cases results in explanations that are mechanistic. All of this is so, even granted that the law's explanation is anisotropic, not mechanistic. Much of the work in statistical mechanics consists in calculations applying the principles of the science to cases defined by various parameters. This perhaps helps to account for the widespread conviction that explanation in statistical mechanics is wholly mechanistic. But that is to overlook the fact that statistical mechanics' fundamental achievement is to have explained those principles in the first place; and those explanations, we have argued, are not mechanistic.

That the Second Law is itself mechanistic may seem to be an unwelcome and awkward conclusion. It could be avoided by specifying that a mechanistic law must be reversible. Had we done so, however, then we would have been guilty or have appeared guilty of winning our argument on the cheap, by way of a definition. We might add such a clause to our definitions of the mechanical, the mechanistic, and mechanicalism now, but I see no need to do so. There is nothing problematic about allowing that a mechanistic law has a statistical, nonmechanistic explanation (Peirce suggested that all the laws of mechanics are to be explained that way: see below, note11). For the same reason, I am not moved to amend our definition of 'mechanistic' à la Railton (chapter 4, section 7) by adding that a mechanistic explanation or law must involve a mechanism. (If that were a general requirement, what would explain the operation of the mechanisms?)

What if, instead, we were to stipulate that a law explained nonmechanistically is nonmechanistic? That would have disastrous results. Take the nineteenth century's prime example of a machine, the steam engine. It operates by the flow of heat; it was in fact the example that led Carnot to formulate those principles that were the foundation of classical thermodynamics. If what is explained nonmechanistically is nonmechanistic, then the principles of the steam engine are nonmechanistic. And that would turn the received meaning of 'mechanical' on its head.

Which leads to a final reflection. It would be a serious error to suppose that a statistical phenomenon – that is, one explained statistically – is 'merely' statistical. To be sure, the statistically average family cannot sit down to dinner together; it could not, even if it consisted of four people rather than 3.75 people. Yet that statistic makes a difference, for example, to economic planners and business strategists. More dramatically, the general fact of heat's anisotropic flow is a statistical phenomenon that makes most things in this world work as they do. It is a reality – one far more consequential than most nonstatistical realities. But this sort of point is now commonplace, for example, that the rigid structures of familiar objects, hence, of all ordinary machines, are a consequence of the statistical properties of their microscopic parts. Clocks are made up of clouds, in Popper's phrase (1972, ch. 6).

3. Natural Selection

Darwin's *Origin of Species* was published in 1859 when Peirce was twenty years old. Its influence on him was profound. Here he stated Darwin's central idea:

> there are just three factors in the process of natural selection; to wit: 1st, the principle of individual variation or sporting; 2nd, the principle of hereditary transmission . . . ; and 3rd, the principle of the elimination of unfavorable characters. (6:202)

Eliminating the unfavorable is selecting the favorable. Implicit in the theory – indeed, in the idea of selection *simpliciter* – is a distinction between 'selection of' and 'selection for' (Sober 1984, pp. 97–102) – a distinction that we shall exploit mercilessly. If we are to defend what Peirce did with Darwin, not merely as reasonable in his day but with continuing pertinence in our own, then we need first to restate the Darwinian idea, not in Darwin's terms but in contemporary, neo-Darwinian terms, with some care, relying on the best recent treatments of it.

The point I will try to make is, I believe, independent of the controversy over units of selection (what is selected: the individual, the group, the gene, etc.?). Yet one must adopt some language, and, so, I speak of the genotype as that which is selected and of the type for which it is selected as being a phenotypical feature instanced by what the selected genotype determines. But it should be kept in mind that the selection of/for distinction is not the same as the genotype/phenotype distinction and that the difference between the 'of' and the 'for' that I stress does not depend on the one being genotypical and the other phenotypical. The difference I stress is of concreteness versus abstractness, and it would obtain no matter which units of selection are chosen.

Natural selection is *of* a particular genetic variant from among competing variants. Competing variants, or alleles, are within the same interbreeding population. The competition is in reproduction: which genes get transmitted to the next generation, or which in the highest proportion? Being selected is being retained in subsequent generations in higher proportion than competitors are. No variant can be selected that does not already exist and have existing competitors. Therefore, natural selection is not from among possibilities; it is from among actualities. Hence it is of something relatively concrete.[6]

Being concrete, what is selected has many features and it has each of those features in a quite specific way. But what is selected is selected *for* just one of its features – more precisely, for a *type* of feature that it exemplifies or results in exemplifications of. Features of the type selected-for give the organisms that possess them an advantage over competitors that lack them. The variant selected need not produce the best possible feature of the type. It may exemplify the type poorly, as long as competitors exemplify that same type much less well or not at all.

The type for which there is selection is thus distinct from the genotype selected. Though only a possibility, the type selected-for is cited in explaining why one variant was selected from among others: the variant

[6] Genotypes are concrete relatively to abstract possibilities or mere types. They determine specific structures in their individual members. The occurrence of monsters – members of the genotype in which that structure is not fully realized – is proof that a genotype is not an abstract type or universal. The instances of a universal must possess its defining attributes, whereas the members of a genotype are linked by descent, not by definition. Furthermore, it has been argued that a genotype is an individual, as it comes into existence at a particular time and place (fuzzy at the edges as these may be) and goes out of existence at a particular time (Ghiselin 1974). That is so, even though a genotype reproduces itself in many physical individuals.

was selected *because* it exemplifies that type (or results in exemplifications of it), while its competitors did not do so or did so less well. It is *because* lenses and focusing increase visual acuity that genetic mutations resulting in lenses and focusing were retained in subsequent generations; in fact, that happened in independent lines of animal evolution. The selection in those cases was *for* visual acuity and *of* concrete structures (or the genes that determine them) that improved visual acuity in specific ways.

At a closer level of analysis, of a narrower segment of the evolutionary process, we might say that there was selection for focusable lenses and of particular structures of that type. The of/for distinction is relative to the level of analysis, but the object of 'for' is always an abstract type and the object of 'of' is always something genetic or genetically determined, hence, concrete. It is important for a later purpose to remember that Darwinian selection, ultimately, is always for a type of effect that the organic feature selected has – an effect, namely, that contributes to the reproductive success of the organism that possesses that feature. There is, thus, always the possibility of distinguishing means from ends.

As the type selected-for is essential to explanation by natural selection, such explanation is like anisotropic explanation in statistical mechanics: both explain actual phenomena by the types they exemplify. Hence, it is not mechanistic. However, to determine the exact nature of explanation by natural selection, we need to distinguish three kinds of explanandum: (1) a particular adaptation, *S*, (2) the fact that there is selection for an adaptive type, *T*, and (3) the general trend in the biosphere toward adaptation.

(1) The existence of an adaptation *S* is explained in two parts, first, by its having occurred, which is more or less by chance, and, second, by its having been selectively retained in subsequent generations. Once having occurred, an allele's selection is explained by its being of type *T* (or having effects of type *T*), by its alternate not being of type *T* or exemplifying it less well, and by *T*'s conferring some competitive advantage. The type, *T*, is therefore an essential part of the explanation of *S* qua adaptation. It is an adaptation only if it was selected because of its being of type *T*, where being of that type enhances reproductive success.

It is qua adaptation – hence, in that aspect – that *S* is explained by natural selection. *S* could also be explained, had we knowledge enough, as a product of a complicated series of mechanical events. But, then, *S*'s enhancing reproductive success would seem a surprising coincidence, a bit of biological luck. *S*'s being an adaptation would not be explained. This parallels the point of our demon example in the preceding section.

The occurrence of genetic variation has several causes – mutation, hybridization, and so on – but chance predominates. Given the dependence of the selection of *S* on the occurrence of *S* and, hence, on a chance factor, the explanation of *S* by natural selection carries with it little power to have predicted *S*. Prediction would be possible only in unnatural situations (breeding, genetic engineering). The explanation, then, is essentially post hoc. Given an adaptation, we can postulate certain events such as chance mutations that first produced it and then selection 'pressures'[7] that explain why it had a good chance of being selectively retained. Darwin's discussion of the finches of the Galapagos Islands is a classic example of such post hoc explanation.

(2) But why, given the existence of alleles that differ in the degree to which they exemplify *T*, is there selection for *T*? It is because being of type *T* is adaptive. Thus, *T* is itself the answer, given the existing population and its environment. It is because of what visual acuity is that there is selection for it in certain populations (of animals) in given circumstances (predator/prey relationships). Because of what it is, visual acuity conduces to the reproductive success of those animals in that environment. The type *T* is thus an essential part of the explanation of why there is selection for *T*.

Again, the fact that there is selection for *T* depends on variants, that differ in respect to *T* actually occurring, which is usually a matter of chance. Therefore, selection for a given type cannot usually be predicted. But some types, broadly defined and not too remote from existing structures, are of such decisive significance for reproductive success that selection for them can be predicted for periods sufficiently long. For example, bacteria, under certain conditions, will eventually develop an immunity to an antibiotic. But how they will do so – the particular defensive mechanism (or gene determining it) selected – is not predictable, because it must first occur by chance.

(3) Finally, the trend toward adaptation (adaptations of any sort) is simply a corollary of there being variation in inheritable characteristics differing in such respects as visual acuity – types that explain why, in the particular circumstances obtaining, there is selection for them once genetic variants occur that differ in those respects. Such a trend can

[7] The mechanistic metaphor is useful but should be recognized to be a metaphor. Selection pressures can be explicated only in terms of an environment that makes some types of organic feature more favorable to reproductive success than others. As the type for which there is selection cannot be omitted from an explication of selection pressure, such a 'pressure' is not mechanistic.

actually be predicted, not as a certainty but as a high probability, for a world – the biosphere – in which the stated conditions, of variation, and so on, are met. This, however, with one caveat: absence of cataclysm.

That adaptations will occur is a corollary of what the eminent statistician R. A. Fisher called 'the fundamental theorem of natural selection': 'The rate of increase in fitness of any organism at any time is equal to its genetic variance of fitness at that time' (1930, p. 35). That is, a population of interbreeding organisms becomes better adapted at a rate equal to its genetic variance in properties that differ in adaptiveness. Like the Second Law, this is a law mechanistic in form, as it relates a dependent to an independent variable. Like the Second Law, it is nonetheless not deduced from mechanistic laws but is explained in another way. Fisher himself noted the parallel: both laws, he said, 'are properties of populations, or aggregates . . . ; both are statistical laws; each requires a constant increase of a measurable quantity' (p. 36).[8] Each law is demonstrated by reasoning about aggregates of individual events. Those events have no unitary physical cause, no mechanism that explains the law;[9] nor is the law deducible from mechanistic laws as a special case thereof.

In all three cases, the types for which there is selection play a crucial, ineliminable role in explanation. Such explanation, therefore, is not mechanistic. It is instead statistical in the anisotropic sense. It is irreducible to the mechanistic principles that govern the particular events to which it applies for the same reason that the corresponding explanations in statistical mechanics are irreducible.

To be sure, the process of selection consists of mechanical events. Genetic variants are produced mechanically (whether by chance or not), genes produce their phenotypic effects mechanically (or so for the sake of the argument we may assume; ontogenesis is of a still little-understood complexity), the resulting creatures die at a young age or survive long enough to reproduce, and those facts of life and death – being crunched in a predator's jaws, running faster than a predator, coitus and conception – are mechanical. However, there is no mechanism that does

[8] Fisher is usually (e.g., Schweber 1982) accorded the honor of having been the first to notice the statistical character of Darwin's reasoning and its relation to statistical mechanics, even though he was anticipated in this by Peirce in his best-known essay, published 53 years earlier.

[9] I am using the word 'mechanism' strictly, for a material system operating mechanically. It is common to refer to natural selection itself as a 'mechanism'; but clearly it is not a mechanism strictly speaking.

the selecting. Instead, there is a welter of diverse kinds of mechanical events. Thus, from a mechanistic point of view, there would be no reason for an ordered result – a new adaptation – to emerge out of that chaos. There is no mechanical cause of such an outcome nor any complex of causes related by mechanistic law to such an outcome.

As with molecular macroprocess and microprocess, we have here two aspects of one concrete reality, one statistical, the other mechanistic: they are inseparable but distinct and differently explained. And, as in statistical mechanics, it is the statistical or nonmechanistic aspect that ends up being the one most consequential. That is so even from a mechanistic perspective; for the statistical fact of adaptation results in purely mechanical changes that would not otherwise occur. At each stage in the evolution of life, living creatures have wrought changes in their physical environment. They have synthesized organic molecules from inorganic ones, deposited inorganic matter as a byproduct of organic processes (reefs, limestone), dug tunnels that aerate the soil, and so on. Many of these effects made further forms of life possible, which produced more and other physical effects. Creatures adapted for survival have transformed this planet and are now poised to transform other planets.

4. Evolution and Entropy

Why is talk of 'selection' justified in respect to organic evolution but not in respect to the growth of entropy? And why does natural selection lead to new forms of order, while the Second Law is the enemy of all order (except in the sense, as Peirce once noted, that chaos is the simplest order)? The two questions are connected.

A type of outcome may be said to be selected-for when (a) outcomes of that type exist because they are of that type and (b) the explanation of (a) is *not* that the type in question includes a preponderance of all equiprobable possibilities. When (b) holds, then what explains (a) is one or another form of selection.

It follows that types selected-for are inherently improbable, in the sense that they do not include a preponderance of equiprobable alternatives. Results of such types are systems that have low entropy or, in other words, that exhibit one or another form of order.

Selection does not make the improbable probable; for the specific entity selected has still to first occur by chance, most improbably. However, selection does make it probable, or less improbable, that *one or another*

improbable entity will eventually come to endure, perhaps proliferate. Even a Darwin could not have predicted what kinds of finch would evolve, from stray finch visitors, to exploit the new ecological niches in a recently vegetated volcanic island; but he would not be surprised if one or another improbable species of finch did evolve.

In the thermodynamic processes we have discussed, certain macroscopic types of outcome encompass nearly all of the equiprobable microscopic possibilities and therefore are highly probable. But in organic life, both the genotype selected and the type of feature selected-for are in that same sense highly, almost infinitely, improbable.

The Second Law is a universal law of physics. The organisms that are the product of natural selection are open systems that sustain themselves by exchanging matter and energy with the surrounding world. They degrade the energy and matter they consume, sustaining their own order by accelerating the increase of entropy overall. And that is possible only because the world as a whole is very far from thermodynamic equilibrium; for whatever reason, its matter and energy are not uniformly distributed.

Darwin's theory does not explain but, rather, presupposes organic life. It explains the evolution of new species, but only from old species. Any account of prebiotic evolution would have to be in terms of a nonequilibrium thermodynamics, a thermodynamics of open systems. Such a science now exists, the creation largely of the late physical chemist Iyla Prigogine and his school (Prigogine 1980). It is worth spending a moment sketching his approach, because it completes the picture begun by Peirce. Between the two statistical sciences Peirce compared – one that studies processes producing disorder, and the other, processes producing order – there lies an enormous chasm, which Prigogine showed how to bridge, as follows.

The nonlinear equations of chemical kinetics that govern autocatalytic and cross-catalytic reactions in thermodynamically open systems admit of different solutions, some corresponding to high-entropy disorder (equilibrium systems) and others to low-entropy order (named by Prigogine 'dissipative structures', because they depend on degrading matter and/or energy from the environment). The latter are molar organizations, for example, of spatial separation or temporal periodicity, that are to some degree self-sustaining (in that environment) and in some cases grow or are self-replicating. Which solution obtains (in Prigogine's figurative language, is 'chosen') depends on small fluctuations of molecular origin in, for example, chemical concentrations. Those fluctuations are, by

quantum mechanics, random. But once a system of molecules forms a dissipative structure by chance, it tends not to revert to its former, lower level of organization. Thus a random fluctuation is stabilized or in some cases amplified. Improbable as such fluctuations are, time is on the side of one or another such improbability occurring eventually. There is, then, over time, a predictably anisotropic emergence of unpredictably novel forms of organization – systems that are inherently improbable but that tend to endure and to proliferate once having occurred.

That was a very crude sketch of Prigogine's mathematically sophisticated, though incomplete, theory. For our purposes, it is important to notice a similarity and a dissimilarity to organic evolution. The similarity: like mutations, randomly occurring fluctuations tend to be retained if they are of one of the 'right' types (and, like mutations, most are not). A fluctuation's being of such a type explains why it survives. We may therefore speak of there having been selection for results of that type. The existence of particulars of such types can be explained mechanistically, but that such improbabilities are to be expected can only be explained statistically by an anisotropic explanation. The dissimilarity: the result of this prebiotic selection is a new form of order, but not an improvement in the chances for survival of an already existing entity (whether individual, society, or species). In other words, what is selected is selected because it is of the type selected-for but not because it has effects of a type selected-for. Prigogine's explanation of prebiotic evolution is thus a kind of anisotropic explanation that occupies a middle position between the explanation of the Second Law in equilibrium statistical mechanics and explanations of organic features by natural selection.

We are disinclined to name 'purposes' the types for which there is selection in prebiotic evolution. Why? I ventured an explanation earlier: a four-part explication of the meaning of 'purpose' as ordinarily used (chapter 3, section 6). A purpose has to be some agent's, that selects for that type of outcome; or it has to be some mean's, that is, something that is selected for its having results of that type. A purpose, in other words, must be possessed by something either as agent or means. But there is no agent in prebiotic evolution and there is no means/ends distinction, either. A dissipative structure is selectively retained because it is a means to its own survival, growth, or self-replication, and not because it is a means to an end distinct from itself. It is in that respect unlike an organic feature that contributes to the reproductive success of the individual or species that possesses it.

5. Peirce's Concept of Final Causation

Peirce's major statements on final causation[10] were made in 1902, among which is this:

> The signification of the phrase 'final cause' must be determined by its use in the statement of Aristotle that all causation divides into two grand branches, the efficient, or forceful; and the ideal, or final. (1.211)

Since Aristotle's list of causes is famously four, Peirce must have meant the two that account for becoming, omitting the two that account for being. The passage continues:

> If we are to conserve the truth of that statement, we must understand by final causation that mode of bringing facts about according to which a general description of result is made to come about, quite irrespective of any compulsion for it to come about in this or that particular way; although the means may be adapted to the end. The general result may be brought about at one time in one way, at another time in another way. Final causation does not determine in what particular way it is to be brought about, but only that the result shall have a certain general character.
>
> Efficient causation, on the other hand, is a compulsion determined by the particular condition of things, and is a compulsion acting to make that situation begin to change in a perfectly determinate way; and what the general character of the result may be in no way concerns the efficient causation. (1.211–12)

In this passage, two sorts of process are distinguished by their form, one involving variable steps with constant type of result, the other, constant rule by which one step follows another but with variable result. Such a rule is a mechanistic law, as we have defined that term. Processes with variable steps and constant type of result are those Peirce called 'finious' and we called 'anisotropic'. In either kind of process, the constant element is something general, a rule governing the steps or a type of outcome. It appears that Peirce's inspiration, in distinguishing the two causes in this way, is the kinetic theory of gases and heat and Darwin's theory of evolution, contrasted to Newton's mechanics and perhaps to Maxwell's electrodynamics.

There is a suggestion here of an empirical criterion by which to distinguish the two kinds of process. It would require grouping processes into

[10] Little has been written on Peirce's idea of final causation. The only extensive discussion of it prior to my 1981b was by Vincent J. Potter (1967, pp. 110–47). Arthur Burks presents a different, more mechanicalist view (1988). Other discussions that I know of are derivative from my paper and typically distort the idea, exaggerating its scope. What I say here is a refinement of my earlier statement.

classes (on some basis other than that of the form at issue, e.g., processes involving changes in distribution of heat and processes where one gear wheel turns another) and noting the contrasting constancies that characterize the processes so grouped. But the distinction depends on processes of the two kinds also having contrasting explanations; otherwise, it would not be a distinction between two causalities. Notice that a process can be grouped with different others, one grouping revealing its mechanical aspect, the other its anisotropic aspect, if it has one.

The characterization of final causation in terms of variability of steps is at odds with Aristotle's conception. For it was by the regularity of a process that that philosopher distinguished something's having a *telos* from its being coincidental:

> For teeth and all other natural things [i.e., things having a *telos*] either invariably or normally come about in a given way; but of not one of the results of chance or spontaneity is this true. (*Phys.* II, 8:198b35, Hardie and Gaye, trans.)

It is explanation by a type of result that is the 'truth' of Aristotle's doctrine that Peirce meant to conserve. He reversed the empirical criterion of finality in order to reveal the way in which types of result are (as we must now think) really responsible for outcomes, namely, the way that is implicit in some branches of modern science.

A fundamental difference between Peirce's and Aristotle's conceptions of final causation is that Aristotle opposed such causation to chance[11] whereas Peirce made chance to be an essential ingredient of final causation. It is chance that provides the materials from which selection for a type of outcome may be made. One advantage Peirce's conception has over Aristotle's is that the coöperation it assigns to chance and selection accounts for the emergence of novel forms of order. There is no room in Aristotle's philosophy for true creativity, either in organic life or in the fine arts or in science.

I think, nevertheless, that Peirce erred in 1902 by defining final causation too broadly. His caution of 1898, when he introduced the term 'finious' – 'if', he said, 'teleological is too strong a word' (7.471) – was better. For the final state of maximum entropy is not one that the Greeks

[11] Or to mechanical processes of which the initial conditions are so various and unknown that they can be taken account of only by means of probability theory. That is the contemporary definition of 'chaos', the subject of chaos theory. It makes no difference to Peirce's doctrine of final causation whether selection is made from among items that occur by chance in that sense or in the sense of being uncaused (introduced into modern physics by quantum theory). I shall therefore use 'chance' ambiguously for either case.

would have recognized as a form of order. It is, instead, a modern representation of that chaos from which they saw order as emerging. Nor would we, today, be tempted to describe it in teleological language. It is quite otherwise with organic features. Let us therefore call 'teleological' those anisotropic processes only that result in forms of order. These are the ones that, in Peirce's account but not Aristotle's, are due to selection (from among alternatives due largely to chance). A final cause, then, in Peirce's but not in Aristotle's sense of that term, is a type for which selection is made. The selection can be made consciously and deliberately, as by a human agent, or, in Darwin's phrase, naturally, by no agent at all. As Peirce suggested, though for the wrong reason, this conception of final cause includes but is broader than our ordinary idea of purpose.[12]

[12] The preceding has been an exposition of what I take to be the crux of Peirce's concept of final causation. He made related remarks that, for our purposes, we may ignore. As some readers may nonetheless feel that to ignore them is to distort Peirce's thought, I shall here briefly indicate what has been left out, and why.

In the 1902 writings on which we have drawn, Peirce wrote, 'Final causality cannot be imagined without efficient causality; but no whit the less on that account are their modes of action polar contraries' (1.213); again, 'Final causation without efficient causation is helpless' (1.220). Such has been our theme, about anisotropic processes generally: their particulars are mechanical; they would have no existence otherwise. Peirce added, 'Efficient causation without final causation . . . is worse than helpless . . . ; it is mere chaos' (1.220). That is in line with the Greek conception of chaos or Boltzmann's of entropy, as conforming to the laws of mechanics but bereft of a patterned result. However, Peirce continued, 'and chaos is not even as much as chaos, without final causation; it is blank nothing'. But, surely, there can be efficient without final causation. As Aristotle noted, rain does not fall for a purpose, e.g., to make crops grow. It just does fall, of necessity. There are two lines along which Peirce might have been thinking.

First, in 1891–3, he suggested that mechanistic laws are themselves explicable only as having evolved: 'Law is *par excellence* the thing that wants a reason' (6.12); 'Now the only possible way of accounting for the laws of nature and for uniformity in general is to suppose them results of evolution' (6.13). Somehow, a self-sustaining form of order can occur, first, by chance; then, being what it is, it takes hold. Or perhaps we need to postulate a Lamarckian power of habit taking, not biologically but cosmologically (6.16, 298ff.): this is the famous doctrine of evolutionary love, or agapasticism. I do not understand either alternative. How can any order be self-sustaining unless some mechanistic laws are presupposed that explain why this arrangement of the parts coheres while that arrangement does not? And, as for agapasticism, is not a tendency toward habit taking itself something that would require explanation as much as does a mechanistic law? The question Peirce raised, about how laws themselves are to be explained, is compelling, even if his answers to it are not. His answers ca. 1891–3 are part of the cosmology that W. B. Gallie dismissed as 'the black sheep or white elephant of [Peirce's] philosophical progeny' (1966[1952], p. 216; see pp. 222–30 for an extensive critique, and Reynolds 2002, ch. 5, for the best account to date, also critical). As that cosmology appears only in a brief period, and about ten years before Peirce developed his idea of final causation, we may question its relevance to the latter and, also, whether he retained it in later years.

And we may then define teleological explanation as explanation by final causes; it is a subtype of anisotropic statistical explanation.

6. Comparison to Recent Views

Peirce's conception of final causation can be measured against more recent attempts to make sense of teleological explanation. These latter may be divided into three groups, the historical/psychological, the physical, and the biological. We discuss them in that order.

A methodological dualism has sometimes been advanced by those who reject Descartes' metaphysical dualism (Dilthey's concept of *Verstehen*, Wittgenstein's discussion of rule following, Gadamer's hermeneutics); a teleological variant thereof may be traced back to Hegel's concept of teleology as presupposing *Verstand* and self-determination or freedom (1969 [1812/31], pp. 734ff.). In the Anglophone tradition, the groundbreaking work was by William Dray (1957), Peter Winch (1958), and Charles Taylor (1964); Georg von Wright's 1971 book completes the development. All these analyses associate teleology with human action exclusively; hence, they are limited to historical and/or psychological explanation. But can methodological dualism be maintained without metaphysical implications? Why is it that humans have to be understood differently from everything else?

For these reasons, I do not take Peirce's cosmology to be a proper part of his teleology ca. 1902.

Second, he held (and this in 1902) that the relation of a mechanistic law to the events it governs is an instance of final causation: 'The relation of law, as cause, to the action of force, as its effect, is final, or ideal, causation, not efficient causation' (1.212). Now, to be sure, a mechanistic law is not itself a mechanical cause; it exerts no force; its rule is somehow 'ideal'. That observation can be used to fend off one objection to the idea of final causation: final causation is no more mysterious than is mechanical causation, perhaps less so. But a law's governance of events does not conform to Peirce's own description of final causality: it does not consist in a process exhibiting variability of means. And therefore we cannot identify it with selection for a type of result. On Peirce's own grounds, then, the relation of mechanistic law to events, while indeed 'ideal', is not teleological.

Peirce several times represented the relation of law to the events it governs by this analogy: 'a law, by itself without the addition of a living reaction to carry it out on each separate occasion, is as impotent as a judge without a sheriff' (7.532, cf. 1.212, 213, 216). That has often been misunderstood to mean that the law needs an enforcer (Boler 1963, pp. 141–44; Turley 1975), as in theistic teleologies. But, like the latter, that only compounds the problem; for the law must still have some power over the agency that is to impose it on events (chapter 4, section 3). A careful reading of the passages cited shows that that is not what Peirce meant (Short 1981b, pp. 378–9): the sheriff is *the fact of imposition*, not an agent that imposes. Peirce's teleology is not theistic.

If Peirce's alternative makes sense, it is superior, as it divides methods differently, with more benign metaphysical implications. It makes rational explanation of human beliefs and actions to occupy one extreme of a continuum of anisotropic explanations, the other extreme of which is occupied by the statistical explanation of the Second Law. In between, we find the theory of dissipative structures, then the theory of natural selection, functional explanations in biology, and teleological explanations of animal behavior (about which, more in the next section). There is still a methodological dichotomy, between anisotropic explanations and mechanistic explanations, but its metaphysical correlate is not a division of kinds of being. Rather, it is a division of aspects of beings. Mind and machine are two aspects of the living human body (chapter 11, section 1), and human mentality is continuous with the nonmechanical, or anisotropic, aspects of any body or process.

Of those who see teleological explanation as having a place in the natural sciences, some reduce it to a form of mechanistic explanation and some do not. The mechanistic view (e.g., Beckner 1959, ch. VII; Woodfield 1976) has a long history (Hobbes, Lucretius), but recent versions can be traced back to the cybernetic theory of Norbert Wiener, modeled on thermostats, heat-seeking missiles, and the like (Rosenblueth et al. 1943), and to a quite different analysis by R. B. Braithwaite (1953). These bear deceptive similarities to the Peircean alternative.

Cybernetic devices act in ways subject to correction by 'feedback' from their effects, so that they maintain a steady state or draw progressively closer to a target. But few of us are tempted to use the language of purpose in describing these homeostatic devices and servomechanisms. As Richard Taylor (1950) pointed out, the analysis based on them therefore does not account for why we do use that language in other cases. Peirce's idea fits our pretheoretical sense better. For he would deny final causation to cybernetic devices on the ground that the variation of means they display (a thermostat clicking on, then off; a heat-seeking missile altering course) is not due to selection for a type of result from among random variations but is perfectly mechanical (cf. 5.473).[13]

Braithwaite (1953, ch. 10) proposed a 'plasticity' criterion of goal-directed behavior reminiscent of Peirce's empirical criterion of final

[13] What of a machine programmed to engage in random trials and to learn therefrom, as, for example, a chess-playing computer that, as it plays, improves its play? That is, I think, a good example of Peirce's idea of final causation and also a good example of how that idea is broader than our ordinary notion of purpose. See the explication of acting purposefully in chapter 4, section 6.

causation, namely, variability of means to a constant end. But, unlike Peirce, Braithwaite refrained from identifying the goal as a cause, and thus he assumed that the behavior's explanation is mechanistic. That was why Israel Scheffler (1959) was able to pose 'the difficulty of multiple goals': any instance of plasticity can be defined by a variety of quite different goals. An infant's crying results in cuddling from Mother but also in distracting Mother from her other occupations. If a goal is no more than a tendency's direction, any more or less constant effect will do to define that direction; death would be a goal of life. Only if the goal must also be a cause will cuddling count uniquely as the crying's goal.[14] The failure of Braithwaite's analysis is evidence in favor of Peirce's ontological boldness. The problem of multiplicity of goals has resurfaced in recent philosophy of mind, under another name but for the same reason (chapter 11, section 2).

In the 1970s, some prominent biologists – Francisco Ayala (1968, 1970), Jacques Monod (1970), Ernst Mayr (1974) – proposed variants of an idea sometimes named 'teleonomy'. Roughly, the idea is that biological processes (from ontogeny to animal behavior) that exhibit Braithwaitean plasticity and appear goal-directed are due to genetically determined programs that were selectively retained, in the course of evolution, precisely because they do determine processes that tend toward results of given types (their 'goals'). In one way or another, the suggestion is made that, while teleonomic explanation has unique features, it is mechanistic at bottom. However, while a teleonomic program depends on mechanisms, these select for a type of result from among chance happenings or from variations, more or less random, produced by the organism itself. The anisotropic tendency of the process can only be explained anisotropically, not mechanistically. More on this in the next section.

By contrast to the foregoing, William Wimsatt (1972) and Larry Wright (1973, 1976) developed an account of functional explanation in biology as being irreducibly teleological in logical form.[15] Like Peirce, they

[14] There is a benign multiplicity of goals: more abstract versus more specific formulations of the same goal and, also, a hierarchy of goals, one being means to another. One or another of such goals might be cited, at different levels of analysis, to explain the same phenomenon. That kind of multiplicity of goals would be no objection to Braithwaite's theory and is no objection to Peirce's. The multiplicity Scheffler cited, and that I have illustrated by the infant/mother example, is not benign and does defeat Braithwaite's theory.

[15] My own understanding of Peirce's idea was much influenced by the work of Wimsatt and Wright (Short 1981b, 1983), and it was only gradually that I came to understand how fundamentally they differ from him. My debt to their analyses remains undiminished.

model their ideas of teleology on natural selection; unlike Peirce, they insist nonetheless that the underlying reality is thoroughly mechanical. Teleological explanation, in their view, is irreducible to mechanistic explanation because it cites outcomes or types of outcome as causes; but what is thus explained, they say or imply, can also be explained mechanistically (Wright 1976, p. 59). Robert Brandon (1996) presents a highly regarded variation on the Wimsatt-Wright theme. Before turning to Brandon, as the best foil against which to argue that forms of explanation inescapably have metaphysical import, let us examine some aspects of Wright's very influential 1976 book. It has been influential despite being subjected to criticism never fully rebutted – which testifies, I think, to the robustness of Wright's underlying approach. The difficulties that that book encounters help to justify the Peircean alternative.

Wright argues that functional explanations (whether of organic features or artifacts) and explanations of goal-directed behavior are teleological because they invoke a 'consequence etiology'. In the case of behavior, Wright explains that etiology this way:

> So in general we may say that teleological behavior is behavior with a consequence-etiology; and behavior with a consequence etiology is behavior that occurs because it brings about, is the type of thing that brings about, tends to bring about, is required to bring about, or is in some other way appropriate for bringing about some specific goal. (pp. 38–9)

In the analysis of functions, 'tends to', and so on, is strengthened to 'does' or 'results in': something has a function, Z, if it 'is there because it does (results in) Z' (p. 81). Presumably, the thought is that organic features and artifacts are 'there' (in a population or a society) because they normally have or once normally had such effects. But that is not what Wright's formula states, and therefore it has been subject to a variety of criticism, such as that it fails to account for defective machines and vestigial organs; but we see how amendment can get around such objections. An objection harder to circumvent is that any effect that explains something's continued existence (e.g., the lethal effect of a gas leak that prevents its being repaired) is by this formula a goal (Boorse 1976). Similar objections can be made to Wright's analysis of explanation of goal-directed behavior. An action may neither tend to bring nor be of a type that normally brings nor be required for bringing nor be in any manner appropriate to bringing about its goal. People sometimes do the worst things possible to achieve their goals. Even dumb brutes make mistakes.

The root problem is that Wright's analysis omits selection, even though he says that 'natural selection . . . underwrites consequence-etiological explanations' (p. 96). Had his analysis exploited the idea of selection, it would have looked something like this: X (organic feature or artifact or action) has Z as its function or goal if and only if X was selected in a process that *selected for* that which makes effects of type Z more likely than they otherwise would be (i.e., selection was for that which tends to bring about Z or is required for Z or can be used for Z, etc.). There is no implication in this that what is thus selected is invariably the best thing for the purpose or even that it is any good for that purpose at all. Though the tendency of selection is to select what does work, not all that is selected must work. Mistakes can be made, instances of a type can be defective, the type itself may no longer be appropriate as circumstances have changed. Furthermore, on this analysis, the actual effects that something tends to have are irrelevant to assessing its purpose – even if, as in Boorse's gas-leak example, those effects do explain why the item continues to exist. So, this revision appears to evade all the objections that have been made to Wright's theory. But the revision is, of course, the Peircean analysis we have developed, that makes a purpose (function or goal, if you wish) to be a species of the genus, type selected-for.

Brandon distinguishes kinds of explanation by the kind of question – how or why or what for – each addresses. A teleological explanation answers a 'what for' question. He avoids some of Wright's problems by limiting his analysis to explanations of biological adaptations only:

> Put abstractly, a *what-for*-question asked of adaptation A is answered by citing the effects of past instances of A (or precursors of A) and showing how these effects increased the adaptedness of A's possessors (or the possessors of A's precursors) and so led to the evolution of A.
>
> The sense in which *what-for*-questions and their answers are teleological can now be clarified. Put cryptically, trait A's existence is explained in terms of what A does. More fully, A's existence is explained in terms of effects of past instances of A; but not just any effects; we cite only those effects relevant to the adaptedness of possessors of A. (1996, p. 41)

As with Peirce's account of teleological explanation in general, this analysis makes teleological explanations of adaptations to refer to the process by which they came to be. It differs from Peirce's account by specifying particular past effects, rather than selection for a type of effect. Thus it realizes Wright's idea of a consequence etiology less ambiguously than did Wright himself. But it does not succeed in making teleological

explanation distinct from mechanistic explanation. For, what Brandon calls the 'full', in contradistinction to the 'cryptic', version is clearly mechanistic. The past effects of instances of A were produced mechanically and the contribution of those effects to the successful reproduction of the possessors of those instances of A was also mechanical.

There is another surprisingly obvious respect in which Brandon's analysis fails. He intends to show how a 'what for' question is answered, but what X is *for* cannot possibly be its *past* effects. What X is for can only be a *type* of effect. Therefore, what Brandon says answers a 'what for' question does not answer it at all. Only his 'cryptic' version comes close. For the cryptic version omits reference to particulars, almost as if it asserted types as causes. The plausibility of Brandon's claim to have explicated teleological explanation requires us to focus on the cryptic version and forget about the full version. The full version is important only for Brandon's steering his analysis back into the safe harbor of mechanicalism.[16]

There is a last defect that must be mentioned. As we have argued ad nauseam, particulars do not explain tendencies. Thus, particular past effects do not explain why there has been a tendency toward the development of organisms well fitted for survival in their respective environments. But nothing is an adaptation that merely happens to conduce to survival; to be an adaptation, it has to have been retained *because* it conduces to survival. To explain it as an adaptation is thus to explain it as due to a process of selection for things having that type of effect. And therefore Brandon fails to explicate explanations (1) of particular adaptations qua adaptations, (2) of selection for adaptations of given types, and (3) of the general trend toward adaptation. All of these require citing real causes, not cryptic omission of details, and those causes must include the types of possible outcome for which there is selection.

7. Purpose's Realm

Equilibrium thermodynamics presupposes but is irreducible to mechanics: it introduces a new principle of explanation, making certain types of outcome explanatory. These explain the increase of entropy, an aspect of processes otherwise mechanical that cannot be explained mechanistically. Similarly, disequilibrium thermodynamics presupposes but is

[16] Perhaps that is exactly what Brandon intends, i.e., to make teleological explanation a kind of 'as if' that, in the end, is not to be accepted as objectively true. If so, he stands in a tradition that goes back to Kant, in his Third Critique.

irreducible to equilibrium thermodynamics. Dissipative structures are possible only in a world governed by the Second Law, yet their explanation requires the introduction of additional principles not deducible from the Second Law. Both forms of explanation are statistical in the sense we have dubbed 'anisotropic' and each explains anisotropic processes, though in the one case those processes are toward predictable disorder and in the other they are toward order of unpredictable forms. Explanations in disequilibrium thermodynamics introduce natural, that is, agentless selection for types of outcome, and thus they conform to Peirce's idea of final causation. We may therefore call them 'teleological', noting that they explain what classical teleologies were intended to explain, namely, the emergence of order from chaos. However, the word 'purpose', as ordinarily used, does not apply to such final causes.

The same pattern of presupposition without reduction persists as we pass from physics and chemistry to biology. Organic evolution presupposes the prebiotic evolution of dissipative structures and in particular the emergence of self-replicating structures, of which organisms are an especially complex subset. But once there are mechanisms for the inheritance of characteristics – mechanisms that admit of occasional variation in the characteristics inheritable – then a new principle of explanation becomes applicable. That, of course, is Darwin's principle of natural selection, the selection of characteristics for types of effect that conduce to reproductive success. The characteristics thus selected may be said to exist for a purpose, even if in some cases they do not serve or are incapable of serving that purpose. The achievement of Darwin is not to have banished purposes from nature but is to have shown how the purposes – that obviously are there – are natural.

We have entered an arena where overstatement is common, and disastrous. There is no basis for claiming that natural selection has a purpose, much less that it proceeds purposefully. It cannot have a purpose, as it was not itself selected. It just occurs. Nor has it a direction, as is often claimed, except in the most general sense, of producing species adapted for survival. There is no reason to suppose that new species must be increasingly intelligent or complex, and so on; they do not even have to be adapted to their environments better than earlier species were to earlier environments. Nor have we any reason to claim that organic species or their individual members have a purpose. For they, too, were not selected (speciation is a by-product of selection of organic features: when types evolve between which there is no interbreeding, there is speciation). Possible exceptions are species bred or individuals raised by men for their own

purposes; or is it hyperbolic to claim that cows exist to give milk, that the kitten spared drowning exists to catch mice, that the man who refrained from suicide in order to wreak vengeance exists for revenge? If the species *Homo sapiens* evolved naturally, then we exist for no purpose at all.

In natural history, the first purposes to appear are those types of effect for which organic features exist. Other kinds of purpose presuppose but are irreducible to these. We shall now trace the evolution of purpose. The story, as it progresses, becomes increasingly speculative, but it has the advantage of not being original: one or another speculation of the sort has been found persuasive by many, and it seems necessary that some such story be true. The point of a likely tale is to show how something is possible, in this case, the various forms of purposefulness.

Some organic features that exist for a purpose serve that purpose only by acting in a certain way, for example, the heart by beating. That action also has a purpose, then. But as the heart beats mechanically, we do not say it acts purposefully or that it beats purposefully (chapter 4, section 6). How did purposeful action emerge? Among the mechanisms that exist for a purpose are those, in animals, not plants, that vary the behavior of the organism as a whole, more or less randomly, in the absence of an outcome of type T and that reduce variation when outcomes of type T are attained. Reduction of variation takes various forms: the successful variant may be sustained or repeated, or acting in that way under like conditions may become a new habit of action (this is learning from experience), or the activity may simply cease for a time (we say that it achieved its purpose) but is repeated later. Any such reduction of variation is a selection for T. As T is a type of outcome for which there is selection, and as it explains the forms of behavior selected, we say that it is a purpose, specifically, a purpose for which the animal acts.

This sort of selection occurs within an individual's lifetime – that is why we say that *it* acts purposefully – and thus it is unlike natural selection, which takes place within populations of many individuals over many generations. Also unlike natural selection, it is driven by specific mechanisms (consciousness of the type of outcome selected-for is not a feature of all such mechanisms). However, the upshot cannot be predicted or explained by knowledge of the mechanisms alone; it depends on chance. And the tendency of the behavior toward one or another outcome of a given type can only be explained anisotropically, by the fact that there is selection for outcomes of that type. The type selected for must be cited to explain the tendency. Notice that purposeful behavior can appear mechanical when it is successful; its distinguishing feature is that it will be varied when it is unsuccessful.

In the next chapter, we argue that this form of purposefulness is essential to interpretation, hence signhood, hence significance.

Mechanisms determining group behavior, for example, of ant hills and bee hives, and instincts that create herd life, hunting packs, and the like, make one individual respond to the actions of another. Sometimes, as in the dance of the bees, actions are performed for the sake of eliciting such response: thus signaling – signs formed by individuals for the purpose of signifying. In most forms of group behavior there is variation and selection, hence, purposeful action by the group. Think of ants going off individually in all directions, until one finds a source of food, whence all soon begin to follow that same track. Agency must here be attributed to a society, which is something more than a population. Groups may also learn from their experience, forming new group habits as, for example, the white-tailed deer have in suburban areas of the United States learned not to fear humans. By mimetic instinct, a group can learn from the experience of just one of its members. One generation thus learns from preceding generations, independently of inheritance via reproduction: thus culture.

The phenomenon of signaling opens up several possibilities, closely connected: first, group deliberation and transmission of culture by means of signs. For signals, if not automatically acted on, become suggested alternatives; when elaborated, they become diagrams of possible action. Diagrams can be manipulated by the diagrammers; by means of instinct or acquired knowledge, they can trace the probable consequences of alternative actions. The survival value of such practices, by way of evading disasters and promoting flexibility of response, is obvious. However, group deliberation becomes individual deliberation when an individual responds to signs it has itself produced (Mead 1934). We noted (chapter 1, section 2) that Peirce, like Plato, held thought to be internalized discourse. The capacity to think for oneself and to act in despite of society is thus social in origin. Individual autonomy and varied personality are further examples of the irreducibility of new realities to their preconditions.

Always, new purposes are emerging. So far in our story, they are formed as means to fulfilling already established purposes. Focusing of light is a purpose of lenses, as a means to visual acuity, which is a means to individual survival, which is a means to reproduction of the genotype. But focusing is not reproducing; it is a distinct purpose. Nor is it predictable from the end to which it is a means. There could have been no purpose to focus light before organisms of certain kinds, with light-sensitive cells connected neurologically to centers of motor control, had evolved,

a fortiori by a series of chance events. Similarly, political power serves social cohesion, which is necessary to the survival of social organisms. But there could have been no political purpose before there was a society complex enough to have produced individual autonomy; for the latter increases a society's flexibility but also threatens its cohesion. With every new purpose, there is a new form of organization, adding to the variety of phenomena that we are pleased to observe.

We turn, now, to the emancipation of purpose from biology. Flexibility of response serves biological ends and was increased by the evolution of the pleasure/pain principle. For pleasure and pain are keyed to a wide variety of conditions beneficial or harmful either to the animal, to its ability to reproduce, or to others of its family, tribe, or species. Thus it can learn from experience to seek or to avoid things of types that it was not genetically determined to seek or to avoid. These are new purposes, but, more importanly, they are a new kind of purpose, formed by individuals. Those purposes are usually means to the biological ends that the pleasure/pain principle exists to serve. However, it is possible for an animal to learn to seek forms of pleasure that do not conduce to those ends, or to seek them beyond the point where that is useful, and similarly to avoid pains that, for biological purposes, might better be endured. Ergo, gluttony and lust and cowardice and sloth. In the case of economic and political animals: avarice and tyranny and cruelty, which are the pursuit of wealth and power beyond their social utility.

So also the development of human affection for its own sake, beyond its biological purpose, and beauty, strength, courage, generosity, loyalty, noble self-sacrifice, and so on for their own sakes, sometimes at the risk of social cohesion and group or individual survival (so Homer depicted the Trojan War). How is it that we have come to prize these deviations from biological norms and to disvalue others? That marks a new level in the formation of purpose, which we must now try to understand.

Most remarkable of all, we have learned to take pleasure in signs and sign interpretation for their own sakes, independently of any practical purpose. Practical discourse has been made poetry and story and history; diagrams have been made mathematics and scientific theory and pictorial art; auditory signals, music. Truth and beauty have become human purposes. Immense wealth and power have been devoted to science and art, sometimes knowingly at the expense of life. Our capacity to depict unactualized possibilities, fearful or attractive, and to define nonbiological purposes and rules of behavior subordinate thereto, has transformed structures that came into being initially as means to biological survival.

Thus the influence of religion and morality on politics, for example. In Aristotle's words, the *polis* originated to secure life, but continued for the sake of a good life. Once again, whence that judgment 'good'?

It is often supposed that pleasure and avoidance of pain are the ultimate purposes for which these other purposes are adopted as means. But that is to overlook a fact we have been at pains to establish: that a purpose formed is independent of the conditions that explain its formation. Aging Don Juans pursue pleasant ends beyond the point at which they cease to be pleasurable. Dogs chase rabbits even when they are fed well at home. We want to know the truth even when it is painful – quite beyond any practical purpose knowledge of the truth might serve. Besides, pleasure and pain have turned out to be highly malleable. We learn to take pleasure in things – caviar, alcohol, hard work – initially unpleasant. Art appreciation is taught. We have a moral duty to learn to take pleasure in doing our duty and in exercising self-restraint. But most importantly, our capacity to diagram and symbolize means that we can formulate possible purposes independently of any motive to adopt them. Sometimes, we then adopt them, arbitrarily or for reasons not well considered.

Human culture and irrational purpose would each seem to be an ultimate emancipation of purpose from biology. However, there is one further step to take. Purposes adopted irrationally tend not to endure. By contrast, some possibilities seem fated to become purposes and, once adopted, to persist and to gradually strengthen their hold. Among the latter are those that dominate human cultures: not only the arts but certain genres of art especially, not only religion but certain forms of religion especially, not only politics but certain political ideals especially seem to have staying power and a power to spread from one nation to another. In Peirce's words (in which 'idea' stands indifferently for a content of consciousness and a general type or possibility):

> it is the idea which will create its defenders, and render them powerful. (1.217)
>
> every general idea has more or less power of working itself out into fact; some more so, some less so. (2.149)

Again,

> ideas are not all mere creations of this or that mind, but on the contrary have a power of finding or creating their vehicles, and having found them, of conferring upon them the ability to transform the face of the earth. (1.217)

This last passage is particularly telling. Just as life, the first form that purpose took, transformed the physical surface of the Earth, so also human culture, the last embodiment of purpose with which we are acquainted,

has transformed the biosphere, making grasslands into pastures, rivers into electrical power, forests into libraries, and stones into cathedrals.

Peirce's conception of final causation turns out to be more Platonic than Aristotelian: it attributes a power to the type itself, independently of that type's being the nature of any existing individual or being otherwise embodied.

> [T]he idea does not belong to the soul; it is the soul that belongs to the idea. The soul does for the idea just what the cellulose does for the beauty of the rose; that is to say, it affords it opportunity. (1.216)

Platonism of this sort is presupposed in Peirce's theory of the normative sciences (chapter 3, section 1). Aesthetics discovers ultimate ends by their appeal to our unqualified admiration. Ultimate ends, then, are those possibilities that create in us a desire for them, causing us to adopt them as our purposes. As ethics establishes the rules that must be followed in order to attain the ends that are ultimate, we are here given a Platonic path to ethics, in place of the Aristotelian foundation denied.

6

Significance

Some further aspects of final causation, now to be noted, enable us to define 'interpret', 'sign', and 'significance'. Thus begins our systematic reconstruction of Peirce's mature semeiotic. We shall proceed independently of his own words, at first. But in sections 7–9 we will find that our definitions conform to the main features of his later statements, especially those of 1907 and 1909. Those definitions are further verified in the next three chapters, where they provide the framework within which we explicate the divisions of object and interpretant that Peirce made in his last two decades, as well as the taxonomy of signs that he based on those divisions.

Recall that Peirce intended semeiotic to be a science (chapter 1, section 5) and, as such, not an explication of ordinary usage of the word 'sign'. 'Sign' becomes a technical term justified, if at all, by the power of the system of semeiotic to illuminate a wide variety of phenomena. Recall also that this science was intended to provide a naturalistic account of the human mind (chapter 1, sections 3 and 4). Thus its key terms must span the human and the nonhuman. Again, this can be justified only by its success in enabling us to see facts not heretofore noted, in this case, continuities. Synechism is to be established not by fiat but by discovery. I make a particular point of this, lest you cavil at what might seem at first to be gross anthropomorphism.

A last prefatory caution. Because the basic categories of semeiotic are to be stretched widely, to cover phenomena far removed from human cognitive processes, we shall at first be emphasizing the primitive and illustrating our definitions by primitive examples. It is not to be supposed that human cognition is therefore being reduced to animal behavior. Peirce's

semeiotic is opposed to behaviorism and other reductivist philosophies. It would be opposed to them even if they were properly teleological and not, as they are, mechanistic. Within the broad semeiotic categories, we can distinguish forms of semeiosis that are uniquely human and that have uniquely human purposes (section 6 and chapters 8 and 9; cf. chapter 5, section 7). Continuity does not preclude but rather entails difference.

1. Teleology as Conjectural and Empirical

In chapter 4, we distinguished anisotropic explanation, which is by type of outcome, from mechanistic explanation, which is by particular conditions related by law to outcomes, and in chapter 5 we argued that anisotropic explanation is in some cases teleological. In teleological explanation, the explanatory type is one that is selected for, whether in natural selection or in purposeful behavior. In either case, the type selected for explains the item's selection: it was selected because it is of or has effects of that type. Thus we called these types 'causes', though they are not mechanical causes; they are distinguished as 'final'. When a final cause is either that type for which an agent selects or that type for which something has been selected as a means, it is ordinarily referred to as a 'purpose'.

Attributions of final cause are always hypothetical and always empirical. The same is true of attributions of mechanical cause. In this respect, either kind of cause is known in the same way. Neither sort of cause is observed independently of logically contingent assumptions, including some that are general; that is what makes them hypothetical. And observations of either are testable against other observations. The assumptions made are supported by the agreement of the observations they make possible and are thrown into question when those observations disagree. That is what makes the knowledge of either kind of cause empirical. It is empirical not in the Baconian sense of being derived from observation but in a Peircean or Popperian sense of being testable against observation. Neither certainty nor freedom from error is attainable, but errors may progressively be eliminated, knowledge refined, extended, and deepened, and uncertainty diminished over time, as inquiry continues. That is the structure of objectivity that is explored in some detail in chapter 12, and therein explained and defended. For the present, a pair of examples must suffice to render the idea plausible, as applying to knowledge of causes of either kind.

Take the observation of a compass needle moving when an electric current passes through an adjacent wire and compare it to the observation

of a bear tearing up logs and eating the grubs exposed. We say that we see the needle being moved, or perhaps we go so far as to say that we see it being deflected by the magnetic field produced by an electric current; similarly, we say that we see a hungry bear searching for grubs. Those observations depend on assumptions, about there being causes of motion, possibly including magnetic fields, and about ursine behavior and its relation to organic needs. We do not have to have an exact idea of what a field is or of how the bear's behavior is governed (e.g., whether by thoughts) in order for those observations to have meaningful content. That content may be explicated in terms of implications for what else is observable; thus the observations may be borne out, or not, by further observations. Does the needle return to its original position when the current is switched off? Does the effect diminish as the needle is removed further from the wire? And so on. Does the bear sometimes tear up logs and not eat the grubs exposed? And so on. If observations do not agree, we have reason to doubt what at first we had thought we observed. Maybe the bear was not looking for grubs. If observations do not agree reliably over many cases, we have reason to doubt the assumptions being made in them, for example, that there are such things as magnetic fields or that animals do engage in end-directed behavior.

The evidence for either kind of cause is one or another pattern of the respective sorts distinguished by Peirce – a constancy in sequence or a constancy in consequence (chapter 5, section 5). The postulated cause is more than the pattern it explains; for it entails the persistence of the pattern and it entails other phenomena as well. More importantly, our idea of it may be made more detailed and exact. When thus improved, a causal hypothesis acquires additional empirical import, and, hence, it may be subjected to additional and more stringent tests. Talk of the purposes for which organic features exist and for which animals act is no more speculative than is talk of mechanical causes. Both are conjectural but testable.

2. Valuation as Teleological

Something that exists or is done for a purpose may fail its purpose. It may serve its purpose well or poorly. It may be of such a nature that it is unlikely to serve its purpose as well as a possible alternative that was not selected. In this respect, knowledge of final causation differs from knowledge of mechanical causation, as the latter provides no occasion for valuation. If we judge a machine to be operating poorly, that is not

with respect to the laws of mechanics – for those it may not violate – but it is with respect to a purpose that the machine was designed to serve.

As purposes can be known empirically, so also valuation can be empirical and therefore objective. One does not have to share the purpose. We can judge whether Hitler was prudent to have attacked the Soviet Union and whether the scorpion's sting is a good defense against all predators; no sympathy with the purpose of either is assumed. It is sometimes said that judgments of value differ from factual judgments by their bearing on practice. But they have that bearing only when one who understands the judgment shares the implied purpose.

The dimensions of value vary. The familiar dichotomies – success/failure, better/worse, good/bad, correct/mistaken, justified/unjustified, real/unreal, true/false – do not apply uniformly everywhere. We have need of some of these, and now explore them.

Wherever there is a purpose, such alternatives as success/failure, better/worse, suitable/unsuitable apply. We shall make particular use of the terms 'suitable to', 'fitting', and 'appropriate to' a purpose, treating them as synonyms. There are all sorts of ways in which something may fit its purpose – being of the right shape or color or forcefulness, and so on. Fitness varies along another dimension as well: something may fulfill its purpose or be a means merely; and if a means, it may be sufficient or necessary or something that can be used for that purpose; it may be appropriate simply because nothing else is any better. The determination of what is appropriate depends on estimates of probabilities and cost/benefit analyses relating to entire sets of purposes.

Being mistaken is distinct from being bad or a failure, as it implies not only that the item in question was selected for a purpose, but also that the process of selection itself had a purpose. As natural selection has no purpose (chapter 5, section 7), it can make no mistakes. Of course, we speak of nature's mistakes – I seem to remember an older sister describing her younger brother as one – but such talk is jocular, I hope. In addition, being mistaken implies the possibility of correction. Not that every mistake can be corrected; some are fatal. But they must be the sort of thing that can sometimes be corrected by those who make them. A mistake, then, can only be made by an agent acting purposefully; for purposeful action, alone, both (a) has a purpose and (b) entails a capacity to modify behavior that is unsuccessful (chapter 5, section 7). The dog that barks up the wrong tree makes a mistake: given more information, it will switch to the right tree, if it is not too late to do so.

Let us focus for a moment on human action. When we judge whether an act was mistaken, we judge it in relation to its purpose and the circumstances so far as they are known *to us*, regardless of what the agent knew. When we judge whether that same act was justified, we judge it in relation to what *the agent* knew or should have known. In light – only – of all that was known or should have been known to the agent, and all of the agent's weighted purposes, was this action the one most likely (or one of the ones likely enough[1]) to achieve its purpose without subverting the agent's capacity to achieve purposes that, *in toto*, outweigh it? Thus, mistakes may be either justified or not. Justification introduces another dimension of valuation.

Perhaps the dichotomy justified/unjustified applies only to humans; for, are the lower animals free to act otherwise than in ways justified by the information available to them? Humans are notoriously subject to conflicting motives and irrational impulses. And yet we do speak of justification even when there is no possibility of unjustified action; we use the term even in reference to natural selection. What justifies an instinct – that is, which factors in evolution accounted for its selection over alternatives as the most cost-effective means of securing its end? But that is only another way of stating a teleological explanation. Teleology bridges the explanation/justification dichotomy, which is a special case of the fact/value dichotomy.

Many, if not all instances, of purposeful behavior have dual bases: a general rule (instinct, habit, custom, thought) and an occasion for applying that rule. There is, in addition, a basis for the rule, in thought or in experience. The experience may be that of an individual or that which accumulates in the evolution of its kind.[2] The action is justified, so far as it is, by the first basis – its occasion – and by the basis of its second basis. We often refer to the latter in what follows.

In the most primitive cases, a rule, 'Given C, do A', is based on an experienced probability[3] that doing A, given C, will suit a purpose P. C may

[1] There is also a cost-effectiveness that must be considered when we expend energy and delay action in order to determine which of several good alternatives is the best one. As General Sudarov said, the better is the enemy of the good enough.

[2] Experience in this sense does not require a unified consciousness; it requires neither unity nor consciousness. We shall speak of 'experience' and also of 'information' wherever particular events affect ('inform') processes of selection.

[3] The probability is distinct from the experience of it. The latter consists of a finite set of particulars, occurring either in the agent's experience or in the evolution of its type. The ratio of successful A's to occurrent C's corresponds only within some degree of error to

be, for example, an odor of rotting wood, and A is tearing up the indicated log, for that is likely to expose succulent, protein-rich grubs, good for eating (P). If C is frustration at not finding food where it was looked for, then A may be random wandering, which increases the chances of finding an alternative source of nourishment (P).

In human behavior, the basis of action typically includes, in addition to trials actually made and consequences actually experienced, possibilities conceived and their likely consequences calculated. But a possibility is present to us only through its representation; so also, calculation requires representation. Such valuative dichotomies as those between the real and the unreal (or merely apparent) and between the true and the false are also semeiotic. None of that can be assumed at this point.

I submit to your judgment this proposition: that the terms employed in this section – 'appropriate', 'justification', 'basis', and the like – are perfectly intelligible in this context and do not presuppose that there are minds or thoughts in the entities in reference to which they are used.

3. 'Interpret' Defined

There are two senses in which something may be said to be interpreted as a sign. In one, it is identified as a sign, as when one says, 'That word means snow' or 'Dark clouds mean rain'. Such interpretation is translation. Probably, it is humans alone who can do this, as it requires forming signs of signs; it seems essentially verbal. It does not explain how any sign acquires an object, as it relates signs only to one another. Exclusive attention to this form of interpretation has led some to espouse a semiotic idealism.

Interpretation in the other sense relates objects to one another as sign and signified. Animal behavior may be taken as a paradigm (and also as its first and most primitive instance). That behavior consists, for the most part, of responses to stimuli. It is purposeful, but responding to stimuli is not its purpose. A deer does not flee the sudden noise that startled it, but a predator; for it is to evade a predator that the deer flees. The instinct to flee is based on an experienced correlation of sudden noises to predators; the correlation is weak, but, unless the deer is near starvation, it is better for it to risk losing a meal than to risk being one. If no predator is there, the deer's flight is a mistake, albeit justified. Mistaken or not, the flight interprets the noise as a sign of a predator.

the objective probability. The latter must be a real number; it need not be rational. Also, the objective probability may change with time, increasing the discrepancy between it and the experience of it.

Our present aim is to capture in a general formula the idea thus illustrated. It must apply not only to animal behavior interpreting stimuli but to all cases of interpretation: to swatches of cloth taken as color samples, spoken sounds taken as meaningful, Greek tragedies taken as representations of the human condition, Mercury's apparent perihelion taken as evidence for Einstein's theory of General Relativity. We shall suppose that purposefulness is essential to interpretation and that a mark of purposefulness is the possibility of error. The kinds of mistake that can be made is our chief diagnostic tool.

We shall define 'R interprets X as a sign of O'. The definition will not presuppose any conception of sign. It is the whole phrase that is being defined. Nor must X be a sign in fact (as 'sign' is later defined) in order to be interpreted as being a sign. Misinterpretation – mistaken interpretation – is a form of interpretation. Any definition of interpretation that did not allow for misinterpretation would be defective.

O is the object that R interprets X to signify. O must be taken broadly: it need not be a physical object. R is not a person or other sign-interpreter; nor is it an interpretation; rather, it is that, such as a feeling, a thought, or an action, by which or in which X is interpreted. It is what Peirce named an 'interpretant'. As 'I' would be confusing, I have chosen to represent the interpretant by 'R', for 'response'. For any interpretant is either a response to the item interpreted or a feature of such a response.

Responses are purposeful actions, though not always bodily movements. That does not mean that they always are deliberate. They may be automatic, but the capacity so to respond must exist for the purpose of producing such responses, and those responses must occur in and normally contribute to a purposeful course of behavior. They are thus subject to correction even if first occurring uncontrollably. For example, we are so organized that sensation can trigger a thought (an observation), willy-nilly. This contributes in diverse ways to our purposeful behavior. But such thoughts are subject to correction subsequently if that behavior fails of its purpose (chapter 12, section 6).

The definition we shall give of 'R interprets X as a sign of O' is intended to be comprehensive, that is, to cover all the instances of interpretation that are to be recognized in the science of semeiotic. In many of the instances – and, in fact, in the most familiar, interesting, and important ones – R, X, and O are semeiotic in nature, independently of their relation to one another. For signs and interpretants can be objects signified, and interpretants may themselves be signs. However, our initial examples will be limited to the most primitive instances, wherein R, X,

and O are constituted as semeiotic entities only in virtue of the fact that R interprets X as a sign of O. That is, the entities that are R, X, and O, respectively, can be specified in nonsemeiotic terms.

R's purpose, P, like R, X, and O, is broadly conceived. It includes instances that are semeiotic in nature, as when our purpose is to know the truth, to chastise the erring, to tell a joke. But, once again, we will at first stick to primitive examples, wherein P is not itself semeiotic in nature.

As Peirce's semeiotic is famously triadic, it may upset some readers to find this fourth element, P, imported into the analysis of semeiosis. I cannot help that. I see no other way to make sense of Peirce's mature semeiotic or, independently of Peirce, to account for the intentionality of signs and of thought. This, however, does not threaten the essential triadicity of Peirce's philosophy. Recall that he held that every relation more than triadic can be reduced to a complex of simpler relations. Our analysis no more than recognizes that semeiosis occurs in a context, that context being one of purposefulness.

The crux of the matter is that the response which R is or is a feature of is part of a purposeful course of behavior and, as such, must take place on some basis that justifies it or seems to justify it relative to its purpose. Thus our definition of interpretation:

> R interprets X as a sign of O if and only if (a) R is or is a feature of a response to X for a purpose, P, (b) R is based on a relation, actual or past or apparent or supposed, of X to O or of things of X's type to things of O's type, and (c) O's obtaining has some positive bearing on R's appropriateness to P.

(i) Recall that there are many ways in which something may be appropriate, short of guaranteeing success (section 2), and notice that there are many ways of bearing positively, short of guaranteeing appropriateness. For both reasons, there is no implication that R's purpose will be achieved if O obtains. The kind of appropriateness at issue and the kind of positive bearing at issue depend on R's explanation: how do X and the relation cited in (b) explain (justify) R relative to P? (ii) 'Obtaining' is intended to cover a wide variety of conditions, from existing to being a fact to being a possibility; here again, which kind of obtaining obtains depends on how R is explained, for that will determine O's ontological category. (iii) The parenthetical clause in (b) disjunctively includes appearances and suppositions, which are themselves semeiotical; in primitive examples, the relations are all actual or past. (iv) 'Actual' is used syncategorematically, so that, for example, some relations among possibilities are actual

(orange actually is between yellow and red on the spectrum, green actually is not).

There are three features of this definition that require special attention. First, the definition does not entail that the O that X is taken to signify must obtain. But if it does not, then what is it that X is being taken to signify? The answer is: O. If that is problematic, it is a problem we endeavor to solve in section 10.

Second, O is specific. In the case of the bear digging for grubs in response to an odor of rot, O could be described as protein or as food or as a means to survival, and so on, but it can be described as specifically as this: grubs. How does it acquire that specificity? If the bear does not have a thought of grubs or a mental image of grubs or some other 'internal representation' of grubs, then how is it that grubs are what its clawing the wood interprets the odor as signifying? On the plausible assumption that the clawing is based on a relation, or past relation, of odor to grubs, and not to grubs-or-beetles, then that is how O is specified; otherwise, its specification would be broader, for example, grubs-or-beetles. (It helps but is not decisive if the clawing is suitable to exposing grubs and no other edible items.) If the bear discriminates by odor among kinds of grub – those good eating, those not – then O is yet more specific.

Third, O is not perfectly specific. If, over time, grubs of another species, also succulent and nutritious, replace those with respect to which the bear's grub-digging disposition was formed, that makes no difference to the bear nor to our semeiotic description of its behavior. There is a degree of looseness in this notion that fits the facts. However, we can only describe O in words and our words will often mean more to us than R interprets X as signifying. That has to be kept in mind, lest we find ourselves attributing detailed knowledge of grub anatomy, phylogeny, and so on, to bears.

In short: O is the object *so far* as it is determined by the relation on which R is based, *given* R's purpose, P.

4. 'Sign' Defined

If O does not obtain, then R is mistaken (R may be appropriate to P nonetheless, but fortuitously). This can happen in either of two ways. The relation on which R is based may be an actual but fallible relation of things of X's type to things of O's type; thus, O may fail to obtain despite X (no grubs in the log despite its delicious odor). Or the relation on which R is based may fail to be actual: it is past or apparent or supposed, merely.

(An example of a past relation as basis: a kind of rot once produced by nutritious grubs, when the bear's habit or instinct was formed, now has a different cause, one not good to eat.) We are inclined to speak of these two kinds of mistake differently. The first is, we say, to have taken a sign that is false or misleading at face value. The other mistake is, we say, to have taken something to be a sign that it is not or is no longer.[4] This way of stating the distinction implies a conception of sign: roughly, a sign is that the basis of interpretation of which is actual. We will follow this suggestion, but it needs more careful statement.

A basis is something that justifies. Justification is relative to purpose and, indeed, to a complete set of purposes and of probable benefits weighed against probable costs. It follows that one and the same relation may be a basis for interpretation sometimes and not other times (in a scarce season, a hungry bear may respond to faint indications it would not deign to notice in palmier days) or for one creature and not another. But we want to frame a conception of sign that is not relative to interpreters and their internal economies. Thus:

> X is a sign, S, of O, if and only if X has such a relation to O, or things of X's type have such a relation to things of O's type, that, for a possible purpose, P, X could justifiably be interpreted on that basis as being a sign of O.

This makes signhood to depend on an actual relation, not one that is past or apparent or supposed, merely. That is one difference between being a sign and being interpreted as a sign. A second difference is that a sign does not have to be interpreted. Interpretability – a specific type of potential interpretant – suffices. This potential interpretant cannot be arbitrary: it must be one that could be justified given a purpose to which O is germane – that is, would be justified granted the right complex of additional purposes, abilities, external constraints.

By this definition, even the most fallible, tenuous connection makes X a sign of O, since one can always imagine a creature desperate enough to take X justifiably to be a sign of O, on no better basis than that. Alternatively, we can assume human cognitive purpose, which is intrinsically semeiotical, not practical, and in relation to which we are free to draw improbable conclusions qualified by their respective degrees of probability. Being humans, we tend to make this assumption, readily granting that all sorts of things are signs, even if unreliable. But then we forget

[4] The expression 'false sign' is ambiguous: it may mean either a sign whose object does not obtain or something that is not the sign it appears to be. The first is unambiguously expressed as 'a sign that is false'. To assimilate it to the second would be like denying that lies are genuine statements.

that a purpose is assumed, and fail to recognize that significance is always relative to purpose.

Reference to a particular interpretant, R, drops out of this definition, as there are many individual acts, thoughts, and so on, by which X may justifiably be interpreted as a sign of O, on one and the same basis. The conception of signhood is nonetheless triadic, as reference is made not only to the sign and its object but also to a potential interpretant. That is what Peirce named the 'immediate' interpretant (chapter 7, section 2). It is a type that may be realized more than once and often in very different ways (a foxy odor makes a foxhound run upwind, a rabbit run downwind, and a man, who knows the scent, think 'Fox!').

By this definition, one and the same thing – say, one and the same physical object – may be many different signs (indeed, I can think of no exceptions), perhaps even an infinite number of different signs. For each relation that X bears to a different O, or that things of one or another type to which X belongs bear to things of other types, will make X a sign of a different O and, hence, a different sign. It is for this reason that the letter 'S' is used to designate a sign that X is. Our way of counting signs is different from our ways of counting the things – physical objects or other things – that are signs. X may be a sign, S, of O and it may also be a sign, S′, of O′, a sign, S″, of O″, and so on. It is one thing but many signs.

Relations of a variety of kinds would justify interpretation relative to corresponding purposes. The disjunction 'relation of X to O, or relation of things of X's type to things of O's type' is designed to cover that variety, but it does not begin to suggest all that is comprised in it. Peirce's triadic division of relations (chapter 3, section 3) accounts for his best-known division of signs, into icons, indices, and symbols (chapter 8, section 2). The relations that make something a sign are semeiotical in themselves when they are rules of interpretation (conventional or natural: see Peirce's concept of a legisign, chapter 8, section 1, or Augustine's of *signa data*, chapter 1, section 6). But they may also be nonsemeiotical, for example, causal laws, probabilistic correlations, resemblances, exemplifications.[5] Both justify interpretation, but the former would not, while the latter would, exist were there no interpreters.

There is a deceptive locution, 'sign to (someone or something)' that may confuse us if it is not examined. We say that X signifies O *to* Y if Y interprets X as a sign of O. Thus, X can be a sign to Y without being a sign

[5] The error of supposing that all of these, because their recognition is a semeiotic act, are themselves semeiotical, is exposed in chapter 8, section 4.

in fact. The same locution can be used to specify what is signified, in an individual instance of interpretation, where that specification depends not only on the relation of X to O in its bearing on R, but also on Y's purpose in forming R. Thus, we might say that the same foxy odor is a sign of its prey to a foxhound but of a predator (or of danger) to a rabbit. The object, fox, is unexpectedly divided into many objects. We can analyze this in terms of degrees of specificity: the most specific description of the object is 'fox', assuming that the odor is specific to foxes. Surprisingly, it is with the less specific descriptions that differences emerge: a fox is one kind of predator to a rabbit but one kind of prey to a hound. Our definition of 'sign' abstracts from these complications, but we return to them in chapter 7.

5. 'Significance' Defined

Whether semeiotical or not, the relation that justifies an actual interpretant, R, is logically prior to R. We may therefore speak of it as a 'prior relation'. It is a relation that X has to the object, O, independently of R. This talk of X's relation to O will be understood to include cases where the relation is between X's type and O's type.

As Peirce pointed out, we frequently speak of meaning or significance in a variety of related ways: for example, what Jones meant by what he said, what his words mean to Smith, their ultimate significance ('This means war!'), and so on. But we shall give pride of place to the significance of the sign itself, for example, to what Jones' words mean even if that is not quite what he had meant to say when he uttered them or what the oversensitive Smith took them to mean. Significance in that sense is what makes a sign a sign; moreover, it makes the sign to be the sign that it is, a sign of O and not of something else. To say that X has that significance is to say that X is S, that is, that it does signify O.

Now, this significance consists in X's being justifiably interpretable as a sign of O. And what justifies such an interpretation, assuming some relevant purpose, is the prior relation of X to O. We may therefore refer to a prior relation as the 'ground' of a sign's significance, and we may define significance as grounded interpretability.

6. The Breadth of These Definitions

As words are signs, anything we can speak of is an object of one or another sign, including things and events, individuals and their properties, classes

and types, signs and semeiosis, the nonexistent, mythical creatures and the myths themselves, fictions, possibilities, impossibilities, and absurdities. (The distinction between an object's being represented *as* what it is – the existent as existing – and being misrepresented – the nonexistent as existing – may be deferred to the next chapter.)

Signs and things taken to be signs may also be of any ontological category. It is not obvious, at first, that they can be. How could something be taken to be a sign if that thing does not exist or occur – that is, if it is not a paid-up member of the club of actuality? Well, it is only through actualities that something less than actual or more than actual can be 'taken' as anything at all. But so it can be. Take four-sided triangles as an example of absurdity. There you have it. An example is a kind of sign: it is something from which we can learn. So, even an absurdity can be a sign.

The example was presented to you through other signs, the words that represent it, and those words, themselves, are general types, not physical objects. You encountered those words through some of their physical instances, the tiny piles of carbon on this page. An absurdity is less than actual; a word, even a word for an absurdity, is more than actual, as it influences the disposition of the actual sounds and piles of carbon that instantiate it and as it determines how those instantiations are to be interpreted (see chapter 8, section 1, on the type/token distinction).

In speaking of absurdities as signs, we are assuming the theory now being developed, a theory that presumably will make sense of words that represent absurdities. For an absurdity can signify only if it is first signified. Another type of sign less than actual is a quality or 1st. As Peirce maintained in his phaneroscopy, 1sts occur or are actualized in 2nds. A particular shade of puce may signify puce in general, but only by being found in a concrete sample (which we help along by saying, 'Look at that awful color!').

Interpretants are responses or are exemplified by or embodied in responses, actually or potentially. As responses are actual, it follows that interpretants cannot be impossibilities. Impossibilities can be signified, but they can neither be embodied in nor exemplified by anything actual. However, anything that can be embodied in or exemplified by something actual may be an interpretant. A thought, for example, as it can be repeated, as it can be shared by different persons, and as it can be further developed, is more than actual. Yet it, rather than a particular occurrence of it, may be what interprets a given sign, albeit only as occurring in a particular response to that sign.

A feeling can also be an interpretant, though only when embodied in a response. A feeling in itself cannot be purposeful; a feeling in itself is not even an occurrence (chapter 3, section 4); nevertheless, it can occur by being embodied in something actual, and it can thus occur in a purposeful course of behavior and serve the purpose sought. For example, one who attends a tragic performance with the purpose, in Aristotle's phrase, of 'gathering in the meaning of things', will allow his emotions to be affected by the events portrayed, and the feelings of pity and fear evoked serve his purpose. That feelings are interpretants is shown by the fact that, although formed passively, they can be mistaken. Pity can be excessive or misplaced (thus sentimentalism). When the vulgar laugh at a noble action, they mistake the unfamiliar for the ludicrous. And so on. These are ways of missing the meaning of things (cf. chapter 7, section 5).

Purposes can only be types of outcome, but the types of outcome that can be purposes are unrestricted. Even an absurdity can be selected for, as in the case of circle squaring, that is, trying by compass and straightedge to construct a square equal in area to a given circle. In such cases, selection consists in rejecting failed attempts. The process will never be completed, because every attempt must fail. It may seem that this sort of thing should not be given a place in our theory, but that would be a grave error. For the pursuit of impossible purposes is a leading feature of human existence.

7. Peirce's Definitions of 'Sign'

As we have developed the preceding definitions systematically and thus independently of interpreting Peircean texts, the question naturally arises whether our work has become unmoored from the latter. In this section, we compare our definition of 'sign' to Peirce's many definitions of 'sign'. In the next two sections, we shall see that our account of significance accords with his most penetrating treatment of the same, in his 1907 manuscript, MS318.

In his later years, Peirce repeatedly defined 'sign', of which the following is a fair sample, not a complete list. In 1886, he wrote, 'A sign stands *for* something *to* the idea which it produces, or modifies' (1.339). In 1897: 'A sign, or *representamen*, is something which stands to somebody for something in some respect or capacity' (2.228). Published in 1903 but written earlier: a sign is 'Anything which determines something else (its *interpretant*) to refer to an object to which itself refers (its *object*) in the same way, the interpretant becoming in turn a sign, and so on

ad infinitum' (2.303). Written or, at least, spoken in 1903: 'a sign is something, A, which denotes some fact or object, B, to some interpretant thought, C' (1.346). Written in 1903: 'A *Sign*, or *Representamen*, is a First which stands in such a genuine triadic relation to a Second, called its *Object*, as to be capable of determining a Third, called its *Interpretant*, to assume the same triadic relation to its Object in which it stands itself to the same Object' (2.274, cf. 2.242, also of 1903, where the same idea is more elaborately stated but 'sign' is made a subclass of representamina, those having as interpretant 'a cognition of a mind'). Published in 1906: 'a sign, (stretching that word to its widest limits, as anything which, being determined by an object, determines an interpretation [*sic*] to determination, through it, by the same object)' (4.531). Written in 1908: 'I define a Sign as anything which is so determined by something else, called its Object, and so determines an effect upon a person, which effect I call its interpretant, that the latter is thereby mediately determined by the former. My insertion of "upon a person" is a sop to Cerberus, because I despair of making my own broader conception understood' (LW:80–1). Circa 1909: 'a sign endeavors to represent, in part at least, an Object, which is therefore in a sense the cause, or determinant, of the sign even if the sign represents its object falsely' (6.348); the sign, Peirce goes on to say, is similarly the cause of an interpretant in a mind, of which the object is then mediately a cause. Written in 1909: 'A Sign is a Cognizable that, on the one hand, is so determined (i.e., specialized, *bestimmt*) by something *other than itself*, called its Object . . . , while, on the other hand, it so determines some actual or potential Mind, the determination whereof I term the Interpretant created by the Sign, that that Interpreting Mind is therein determined mediately by the Object' (EP2:492); Peirce went on to explain that 'determine' must not be taken in narrow sense, if erroneous signs are to be accommodated (EP2:493, cf. 497 and 497n, quoted below).

The basic scheme, of object determining sign determining interpretant, clearly goes back to his 1868–9 papers in The *Journal for Speculative Philosophy*. We witness in these passages several struggles, some of which we have already traced in chapter 2. The interpretant that earlier had to be actual was later allowed to be possible; interpretants that often seem always to be mental are later sometimes affirmed to be not exclusive to human minds; and interpretants that earlier had to be signs, thus spawning an infinite *progressus*, later do not have to be. But there is considerable trouble over how the relation of object to sign and of sign to interpretant is to be conceived.

On the one hand, Peirce wants to say that signs are produced by their objects and produce their interpretants: varieties of causal language are employed, apparently efficient not final in meaning, and the frequently employed alternative term 'determines' is glossed in quasi-causal language. But that view, if understood in any simple way, is implausible, since a human being has a great deal of choice in how to interpret signs and in how to signify objects; and even lower animals are not acted upon by objects and signs simply, since it is their internal organization and the purposes for which they act that determine to which objects and signs they will attend. Peirce himself cited the examples of false signs and signs, for example, commands, of the future. How, he asked (EP2:492–3), can these be acted upon by their objects?

'Determines' lost any causal meaning when he answered, regarding the command, 'Ground arms!', that the object (the act commanded) 'determines the Sign although it be *created by* the Sign' (EP2:493, emphasis added). In a variant draft of the same letter to William James, he wrote, 'A *Sign* is anything . . . which, being *bestimmt* by something other than itself, called its *Object*, in its turn *bestimmt* the mind of an interpreter to a notion, which I call the *Interpretant*' (EP2:497). Recourse to the German word is of little help, as *bestimmen* and *bestimmt* have a range of meanings like that of the English 'determine' and 'determined', in which defining and deciding or being defined or decided is more prominent than is causing or being caused. Yet in a long footnote, Peirce wrote, 'You will probably object that *bestimmt*, to your mind, means "causes" or "*caused*." Very well, so it does to mine' (EP2:497n).[6] Here, at least, we have an indication that Peirce realized that the relation he was trying to formulate is *not* causal. He then proceeded to argue, regarding the object (tomorrow's weather) signified by a hygrometer, that, though it does not yet exist, it may in an 'indirect manner' cause but 'not *really* cause' the sign of it. That explanation, he concluded, is 'as lucid as a bottle of ink', which indeed it is.

If the determination at issue is not causal in the mechanistic sense, might it be teleological? In that case, one thing is determined by another as a means to an end. But that, too, is implausible as an explication of the determination of signs by objects. In semeiosis, it is the sign-interpreter, and not the object signified nor the sign nor the sign-producer, that must

[6] Oddly, he had published a more accurate account of the meaning of the German word four decades earlier (W2:155–7), though that was not in the context of sign theory.

always act purposefully. Even in the case of the command to ground arms, in which the object is the end commanded, it is not that end that produces the command, either as means to its fulfillment or in any other way.[7]

What, then, of noncausal meanings of 'determine', for example, deciding and defining? Deciding does not fit, but objects do in a sense define their signs and signs their interpretants. A sign is identified as a sign of such-and-so, its object: 'That's a stop-sign', 'Those scratches are a sign of bear', 'That's French for man'. An interpretant is identified by the significance it attributes to the item interpreted: 'I interpret the poem as the expression of a death-wish'. If that is what Peirce meant, then the determinant makes the determined to be what it is, not as a cause, final or efficient, but as a formal condition.[8] Of course, more than frogs are required in order to make the word 'frog' mean frogs – for example, a rule relating 'frog' to frogs – but the kind frog (at least as a possibility, if not as an actuality) is essential.

Within that idea there lurks another meaning of 'determine' (and also of *bestimmen*) that applies to the relation of object to sign and of sign to interpretant and that perhaps brings us still closer to Peirce's intention. This meaning is that of to limit, as in, 'The water's edge determines where your property ends'. Thus, if one is to command that arms be grounded, he is limited in the means he can employ. Qualities and general terms alone cannot indicate particulars, including an action (of a represented general type) that is to be performed *now* by *you*. Nor can pointing represent a general law. And so on. Each object limits, or determines, what may be a sign of it, and each sign similarly determines what may be an interpretant of it. The wool-gathering poet-soldier who contemplates the music of 'Ground arms!', and does not act (does not even resist acting, as in deciding to disobey) is not interpreting those words as a command. A command, being the sort of sign that it is, determines what may be an interpretant of it as a command, though there is great latitude in the

[7] In chapter 5, section 6, we considered that some possibilities may create and arm their defenders. But that is not a general account of all ends commanded, much less of all objects signified.

[8] To be sure, that is one of Aristotle's four causes, the so-called formal cause. This but testifies to the deep connection (not in English only: the German and Greek and English words are not etymologically connected) between different senses of determining or – to drag in another, more common term – of making. Making (as efficient cause) a sculpture, makes (formally) one a sculptor. What makes (produces) a sign and what makes (defines) that sign to be the sign that it is, are not always the same. My uttering the word 'frog' is not what makes that word or my utterance of it to mean frogs.

interpretations thus permitted. Notice that false signs and mistaken interpretations are also thus limited: one cannot falsely signify nor mistakenly interpret except by signs or interpretants, respectively, that fit those supposed objects or meanings. Thus, in an earlier draft of the 1909 letter to James quoted above, Peirce glossed 'determined' as 'specialized, *bestimmt*' (EP2:492); 'specialized' suggests this idea of being limited. And in section 1 of chapter 9, we will see that 'determine' must be understood in this sense, if we are to explicate a fundamental principle Peirce employed in classifying signs.[9]

Our definition of 'sign' agrees with Peirce's definitions in this wise. It is, first of all, triadic, making interpretation (actual or potential) essential to the relation between sign and object. Second, it agrees with the trend of Peirce's definitions in making interpretability, not interpretation, to be requisite to signhood. Third, it is as broad as Peirce's definitions came to be, in not limiting interpretation either to conscious thought or to further signs. Finally, it is consistent with the idea that objects determine their signs and signs determine their interpretants – in the only senses of 'determine' that make any sense of that doctrine.

8. Peirce's 1907 View

The 1907 manuscript, MS318, in which Peirce asserted the necessity for there being ultimate logical interpretants (chapter 2, section 10), is limited for the most part to consideration of those signs therein named 'intellectual'. But there is one passage (5.472–3) that addresses the nature of signs – or of their 'action' – generally. Immediately after announcing that topic, Peirce formulated the difference between final (triadic) and mechanical (dyadic) action. The ensuing discussion suggests that the 'action' of a sign is final, not mechanical.

> The action of a sign calls for a little closer attention. Let me remind you of the distinction referred to above between dynamical, or dyadic, action; and intelligent, or triadic action. An event, A, may, by brute force, produce an event, B; and then the event, B, may in its turn produce a third event, C. The fact that the event, C, is about to be produced by B has no influence at all upon the production of B by A. . . . Such is dyadic action, which is so called because each step of it concerns a pair of objects.

[9] What Peirce meant by 'determine' has been the subject of discussion: Ransdell 1976, 1977; Short 1981b, n. 3; Meyers 1992, n. 4; Short 1996a, n. 8. Liszka 1996, pp. 22–4, interprets the word as I do.

> But now when a microscopist is in doubt whether a motion of an animalcule is guided by intelligence, of however low an order, the test . . . is to ascertain whether event, A, produces a second event, B, *as a means* to the production of a third event, C, or not. That is, he asks whether B will be produced if it will produce or is likely to produce C in its turn, but will not be produced if it will not produce C in its turn nor is likely to do so. (5.472–3)

Did Peirce really mean A to be an *event* that produces B as a means to C? Does a single event ever produce another as a means to a third? It would be more natural to suppose that A is the animalcule that performs B as a means to C. That must have been what Peirce meant. Recall that this is in an unfinished manuscript consisting of many layers of revision; it may well contain poorly formulated thoughts and slips of the pen.

Determining 'whether B will be produced if it will produce or is likely to produce C in its turn, but will not be produced if it will not produce C in its turn nor is likely to do so', requires observation of more than a single sequence of events. As the idea is extended to 'animalcules', the needed observation cannot always consist in introspection or in asking an agent what his beliefs and intentions are. Even when introspection or first-person testimony is available, its reliability is testable against other forms of observation. Such observations must be of a persistent pattern of behavior, in which we see whether there is variation and selection for a given type of outcome, namely, whether B tends not to be selected except when it results in C. We are familiar with this thought from chapter 5.

The paragraph, 5.473, is long and convoluted; much later in it, Peirce concluded, 'it seems to me convenient to make the triadic production of the interpretant essential to a "sign"'. That production is, then, the 'action' that Peirce attributed to a sign; the sign acts in eliciting an interpretant and this action is triadic, not dyadic, that is, final, not mechanical. In 5.473, Peirce named this 'action of a sign' 'semiosy'. And in a variant passage, he wrote:

> by 'semiosis' I mean . . . an action, or influence, which is, or involves, a coöperation of *three* subjects, such as a sign, its object, and its interpretant, this tri-relative influence not being in any way resolvable into actions between pairs. (5.484)

That leaves it unclear exactly what the action is; the very word is ill-chosen for final causation. Let us seek more definiteness by asking two questions Peirce did not in this passage address. What is the final cause of semeiosis? And what does the selecting, that is, where is the locus of purposefulness?

We have said that it is the animalcule, A, that does B as a means to C. How does that ABC schema apply to semeiosis? It would make little sense to suppose that A is the object, B the sign, C the interpretant. How many objects produce signs in order to elicit an interpretant? (Someone who shouts 'Hey!' in order to draw attention to himself is an example.) And yet, Peirce's first example is of that type:

> an officer of a . . . company of infantry gives the word of command, 'Ground arms!'. This order is, of course, a sign. That thing which causes a sign as such is called the *object* (according to the usage of speech, the 'real', but more accurately, the *existent* object) represented by the sign. . . . In the present case, the object the command represents is the will of the commander. . . . Nevertheless, the action of his will upon the sign is not simply dyadic.

The action is not dyadic, for, as Peirce goes on to explain, the officer would not have used those words if he thought they would not be understood and obeyed.

The next sentence corrects the impression created by the preceding, that objects always produce signs as means to eliciting interpretants:

> However, although this condition is most usually fulfilled, it is not essential to the action of a sign. For the acceleration of the pulse is a probable *symptom* of fever and the rise of mercury in an ordinary thermometer . . . [is an index of temperature], which, nevertheless acts upon it in a purely brute and dyadic way.

Unfortunately, Peirce did not leave it at that but proceeded to try to assimilate this case to the former one, by asserting that 'a mental representation of the index is produced, which mental representation is called the *immediate object* of the sign; and this object does triadically produce the intended, or proper, effect of the sign strictly by means of another mental sign'. But that does not succeed in assimilating the one case to other; for the temperature still acts dyadically on the mercury column. And if it did succeed, then the admission that the condition of triadic production of the sign is 'not essential' would be undercut.

The invention of a mental 'immediate object' intervening between an index and its 'existent' object, producing an interpretant of the index triadically, is neither repeated nor reformulated elsewhere in the variant drafts that compose MS318. In that respect, it is quite unlike the doctrine of the ultimate logical interpretant that is repeated and reformulated several times. Nor is this mentalistic epicycle in Peirce's system repeated elsewhere in his writings. He many times distinguished a sign's immediate object from its real, existent, dynamic, or dynamical object (Chapter 7, section 3), but never elsewhere in this way – that is, on the basis that the

immediate object is needed to produce an interpretant. I think, therefore, that we may be permitted to leave the epicycle aside, as something Peirce tried but found wanting.

What, then, are we to glean from the whole, long passage? This: Peirce definitely did see semeiosis as being triadic or teleological. If, with that in mind, we look at sign-phenomena for ourselves, I think we can see that while signs are often produced as means to ends, they are not always so produced, and that signs themselves rarely or never (I can think of no cases) produce their interpretants as means to ends. And that leaves us with but one alternative (assuming that the temporal direction of semeiosis, so far as there is temporal direction, is from object to sign to interpretant), and that one fairly obvious from a common-sense point of view: interpreters produce interpretants as means to ends.

Consider an example. Jones pokes Smith in the back so as to call attention to himself, for the further purpose of asking for a loan. The attention – an interpretant of the poke – is therefore elicited as a means to an end. But it is not the sign – the poke – that elicits its interpretant as a means. It was Jones, the poker, who poked for that purpose. The poke, itself, compels attention mechanically. Our possessing the mechanism by which attention can thus be directed exists and operates for a purpose. Furthermore, its operation is not utterly mechanical: for example, we can ignore the insistent poking by a child when more important business is on hand. Regardless of whether Jones poked on purpose, Smith has a purpose in taking notice. And it is only thus that we can account for Smith's reaction as being not merely an effect of the poke but an interpretant of the poke, attention being directed not to the poke especially but to its cause.

Thus, A is the interpreter, B the interpretant, and C the purpose or final cause of the interpretant. That would satisfy Peirce's formula, 'triadic production of an interpretant'.

The purposefulness of semeiosis is rooted in the interpreter, not in signs or their objects. Antecedently existing things and events and conditions, and, also, types of things, and so on, become caught up in an end-directed process in which potential interpreters are already engaged, because those things are or were or appear to be or are supposed to be relevant thereto, and *thereby* they are taken to be signs and objects signified. Our concern with illness or with the weather leads us to exploit, for our own purposes, the observed correlation between pulse and fever or expansion of materials and temperature, taking one to be a more or less reliable sign of the other. So also, other animals exploit, for their

purposes, the correlations they have experienced. A sign is something suitable to being so exploited; thus, it is more likely to be so exploited than are nonsigns.

The sign's 'action' therefore depends on its relevance to the purposes of an agent; only so does it have an effect. The sign makes or can make a difference: in that sense it 'acts', when it acts at all. But it acts only through influencing an agent that, independently of that sign, is pursuing some purpose. Talk of a sign's action is only another way of talking about how a sign determines its interpretant. Nothing is a sign except for its objective relevance to the purposes of possible agents.

9. Significance and Purpose

So far, so good. But there is more to 5.472–3 than that. Consider the context into which this discussion of the triadicity of semeiosis was introduced. MS318 as a whole attempts to explain how signs, intellectual signs in particular, signify. What makes 'horse' to designate horses, or a thought of horses to be about horses? It is significance in general that must be explained, without deriving it from thought. But, in opposition to his earlier views, that explanation is no longer supposed to be that one sign interprets another ad infinitum. Instead, Peirce claimed that there must be an ultimate logical interpretant – an interpretant of intellectual signs that is not itself a sign (or is not an interpretant in virtue of being a sign). Why? How does that help? How does that contribute to an explanation of significance? Then we come upon 5.472–3 and its introduction of the idea that semeiosis is teleological.

If we read 5.472–3 in the context of the rest of MS318, it is unmistakable that Peirce's view in 1907 was that the end-directedness of semeiosis accounts for the significance of signs. That is why 5.472–3, which is about the end-directedness of semeiosis, is part of MS318, which is about how signs signify. The necessity for there being ultimate interpretants then becomes clear. For nothing is end-directed or purposeful that does not face the prospect of success or failure, and, hence, semeiosis must terminate, or be capable of terminating, in something subject to success or failure – something, that is, in which the sequence of sign-interpretants can be brought to the test. And in the case of intellectual signs, that is the ultimate logical interpretant.

In primitive forms of semeiosis, where the interpretant is an action, the matter is simple and obvious: the bear uncovers grubs or it does not.

That is its success or failure. It is therefore looking for grubs, and, as it is responding to some stimulus, that stimulus is therefore a sign to it of grubs. But what of human speech and thought, especially when speech and thought are engaged in for their own sake and not for a practical purpose? If we seek knowledge as an end in itself, what distinguishes success from failure? What determines whether we have succeeded in knowing the truth? The chain of signs interpreting signs has to terminate in an interpretant that makes the whole chain testable, giving us not an infallible mark of success or failure but some indication, however fallible, of whether our conclusions are true.

The experimental method associated with modern science reverses the idea of theory pursued for the sake of practice: it is practice engaged in for the sake of theory. A true opinion is one that *would* lead to satisfactory results *were* we to rely on it for some practical purpose. An experiment tests a theory by seeking a result *as if* it were practically important (sometimes it is, but that is irrelevant to pure science). That is Peirce's later, corrected, subjunctive version of pragmatism, that the meaning of a proposition is how it *would* influence conduct – that is, our habits of action – *were* it believed and *had* we some practical purpose to which it was germane. (The concept of truth is rooted in the practical need for reliable beliefs; but in science that concept is subtilized, so as to comprise *possible* practice and so as to be defined in terms of an impractically long – indefinitely prolonged – process of inquiry. Cf. chapter 12, section 4.) Therefore, when knowledge is sought for its own sake, it must nevertheless be translatable into habits of action. Words and formulae not so translatable are not testable and therefore can neither succeed nor fail and therefore are without purpose and therefore have no significance. They are at best part of a verbal game that has no meaning. That is why habits of action, or modifications thereof, are what Peirce identified as the ultimate logical interpretants.

'Horses have double stomachs', which is false, can be translated into other words and explicated in words without limit; but the statement refers to horses and says something about horses because it (and its translations and explications) can be tested by actions taken with respect to horses, for example, a dissection of an equine cadaver. And it is by its use in statements having that characteristic – that they can be tested by doing things with horses – that the word 'horse' means horse. The objective determinant of success or failure is where we look for the object signified.

In conclusion: the point of 5.472–3, in the context of MS318, is that the end-directedness of semeiosis explains the significance of signs. The requirement that there be ultimate logical interpretants is no more than a corollary of this. The account we have given of significance fits Peirce's view of 1907.

10. Intentionality Explained

In chapter 1, section 4, we devised this formal mode criterion for something's material mode intentionality: X possesses intentionality if it cannot fully be described without implying a grammatically simple affirmation about it which cannot be expressed without employing one or another intentional verb (or gerund, etc., derived from that verb). And we defined a verb as intentional if it is transitive and its use in a simple singular affirmation does not license an inference that its object, if indefinite, exists or, if propositional, obtains.

Now, 'signifies' is an intentional verb so defined, for to say that S signifies O is not to imply that O exists or obtains. But it is impossible fully to describe a sign without saying what it signifies. Therefore, signs possess intentionality and significance is a form of intentionality. That is very far from being the end of the argument. For, in the first place, we have identified the significance of signs with their potentiality for being nonarbitrarily interpreted, and one might suppose that their intentionality is thereby reduced to something else. In the second place, those who agree that there is irreducible intentionality may think it depends on something like human thought, inexplicable naturalistically.

Now, to be sure, on the account developed here, significance is not an intrinsic property of anything. It is a function of possible interpretation; thus it is irreducibly triadic. But acts of interpretation are themselves irreducibly triadic, as they consist in assigning objects to signs. And, as those objects need not exist to be so assigned, interpretation entails intentional inexistence. Put it another way: any interpretative act is purposeful and purposeful behavior cannot fully be described except in language employing intentional idioms (as defined here) or implying propositions only expressible by intentional idioms. Take even the simplest, most primitive examples of purposeful behavior: seeking and fleeing. 'Seek' and 'flee' are intentional verbs: what is sought or fled might not exist. Furthermore, any purpose may fail to be attained, indeed, is not yet attained while still pursued. The challenge posed to reductionists

is this: can such behavior be described fully without introducing the inexistent?

The same primitive examples are evidence against the view that all intentionality depends on something like human thought. If they can neither be reduced to the nonintentional nor plausibly accounted for by mental states, then we have persuasive albeit not conclusive evidence that intentionality is independent of mental functions (unless very broadly construed!).

There are many who will still maintain that all talk of purpose, justification, mistakes, interpretation, signs, and significance is 'mentalistic' and can be extended as we have done only by an inexcusable act of gross anthropomorphism. If such is your view, then please do restrict these words as you wish. That will not matter, as long as you admit the differences that have been pointed out, (a) between explanations that cite laws linking particular conditions to particular effects and explanations that cite types of outcome, and (b) between phenomena that can be described fully without use of intentional idioms and phenomena that cannot. You may similarly restrict the word 'cause' to mechanical causes, as long as you admit that types of outcome are explanatory, no matter how they are denoted. It is not the words but the distinctions and continuities they label that matter.

Reductionism is motivated by the conviction that any unreduced instance of intentionality must be something inexplicable, an intolerable mystery, perhaps violating fundamental principles of modern science. And that is why mere failure to reduce the apparently intentional does not persuade the reductionist of his error but only spurs him on to greater efforts. At most, it convinces him that, while intentional language may be irreducible and yet of some ineliminable practical importance, what it seems to refer to is unreal. Nothing inexplicable may be allowed to be real. It is therefore of some interest that Peirce's semeiotic entails a naturalistic explanation of unreduced intentionality. We conclude this chapter by restating that explanation, which has been developed piecemeal over many chapters.

An analogy may be drawn to a logical conundrum that has been solved repeatedly, first by Plato in his response to Parmenides, then by Russell in reference to Meinong, by Strawson, and by Quine.[10] How can we speak

[10] Plato states the problem in his *Theatetus*, solves it in his *Sophist*. Quine's criterion of ontological commitment similarly turns on a distinction between reference to particulars

meaningfully of what is not? If there is no bird in the bush, then to what can we possibly be referring when, mistakenly, we say, 'There is a bird in that bush' or (a somewhat more complex case) 'The bird in that bush is a robin'? If nothing is referred to, then are we saying anything? (And if we are not saying anything, then there is nothing we are saying mistakenly.) The same problem arises if we say, truly, 'There is no bird in that bush' ('What isn't there?' 'A bird.' 'What bird?'). The solution is that we are applying general terms to individuals that do exist (Plato supposed that what I am calling a general term is the name of a Form). The bush exists and birds exist, but none of the latter is related to the former by being in it. 'Being in' is, here, the general term that is being applied to existing individuals, whether in a false affirmation or a false presupposition or a true denial. In denial, we are saying, in effect, 'Take any existing individual in the set of birds: that individual is not in this bush'. In affirmation, that denial is denied. The details will be familiar to most readers. (There are other parsings to the same effect; nor does it matter, at least for this one purpose, whether we sum up the matter by saying, 'See? No reference is made to anything nonexistent', or by saying, 'See? That's how one refers to the nonexistent'.)

Now, just as general terms enable us to speak intelligibly of the nonexistent, so also selecting for a type of outcome relates individuals to the nonexistent. The grubless log pawed is like the birdless bush misdescribed. The general applied to the particular gives to airy nothing a local habitation, whether the general is a word or a disposition. The explanation of intentionality, then, is final causation: that is the reality denoted by intentional idioms. More specifically, the explanation is that there is selection for a type of outcome, where that selection is itself due to mechanisms that were selected for their selecting for that type of outcome. (For only in such case is a mistake possible.) Thus we can say that the bear is digging for grubs and may be doing so mistakenly, guided by fallible signs. There is nothing mysterious about this, because there is nothing mysterious about there being selection for a type of outcome. Only, one must overcome the mechanicalist prejudice that mere types are not explanatory.

A final note. This explanation of intentionality is in terms of purposeful action, and not in terms of final causation in general, much less in terms of

and description by general terms; only, Quine denies Plato's assumption that the latter also refer (1961, ch. 1). Russell's theory (1905), later named 'of definite descriptions', and Strawson's alternative (1950) are well known and much reprinted.

anisotropy (Peirce's 'finiosity') in general. Outside of purposeful action, which appears to be limited to animals, no mistakes are possible, and where no mistakes are possible, there can be no intentionality, hence, no interpretation; but all significance is relative to potential interpretation. That is why I disagree with those (e.g., Sebeok 1985; Pape 1989; Emmeche and Hoffmeyer 1991; Hoffmeyer 1996; Santaella 1999a, b) who suppose that Peirce's semeiotic extends to all of life or even cosmologically. It may be so extended, I think, only at great cost to its own significance. By being extended so far, it explains much less.

7

Objects and Interpretants

In the 1900s, Peirce introduced several divisions of object and interpretant. That was in an article, notebooks, letters, drafts of letters, and uncompleted manuscripts. He was groping his way and never came to a definite, clearly articulated conclusion. Much of this effort was directed toward providing principles for a sign taxonomy, developed in those same years.

We can see in that taxonomy (chapters 8 and 9) that he needed two quite different trichotomies of interpretant. One, following from the teleological structure of semeiosis, pertains to each sign: the *immediate interpretant* is a potentiality in which consists the sign's interpretability; the *dynamic interpretant* is any interpretant actually formed (from zero to many); and the *final interpretant* is another potentiality, the ideal interpretant of that sign for the interpretative purpose. The other trichotomy is an application of Peirce's phaneroscopy and distinguishes among signs: an *emotional interpretant* is a feeling or 1st; an *energetic interpretant* is an action or 2nd; and a *logical interpretant* is a 3rd, being a thought or other general sign or a habit formed or modified. An immediate interpretant may be either emotional, energetic, or logical, and so also dynamic and final interpretants may be of any category, actually or potentially. A sign's final interpretant, for example, is that potential feeling or potential action or potential thought, habit-change, and so on, that would best satisfy the purpose of interpreting that sign.

The distinction between interpretants that are *ultimate* and those that are themselves signs, mentioned earlier (chapter 2, section 10, chapter 6, sections 6–9), is made within the class of logical interpretants and is

different from the two trichotomous divisions of interpretant, despite Peirce's own occasional interchange of the similar terms 'final' and 'ultimate'. Thus, an interpretant may be final without being ultimate, and conversely.

Because of Peirce's many changes of conception (especially of the final interpretant) and terminology, there has been much uncertainty about his divisions of interpretant and, in particular, a tendency among Peirce's commentators to conflate distinct divisions (especially identifying the ultimate, the final, and the logical interpretants). To find our way out of these dark woods, we need to attend, first, to the different ways the respective divisions are defined. They do not compete; they are made on entirely different grounds. Second, we need to attend to the uses Peirce made of those divisions. Both trichotomies are required in his sign taxonomy, while ultimate interpretants are required to complete the account of significance (as in the preceding chapter).

Let us dispel, now, one frequently encountered but mistaken argument. 1sts and 3rds are may-be's and would-be's, respectively, and therefore are not actual occurrences; hence, emotional and logical interpretants are not actual occurrences. But neither are immediate and final interpretants. The conclusion often drawn is that immediate and emotional interpretants are the same, as are logical and final interpretants, every energetic interpretant being dynamic, and conversely. That inference is a non sequitur. While 1sts and 3rds, in themselves, are not actual, a 1st can qualify something actual and a 3rd can be instantiated in something actual. There is, then, no obstacle to speaking of a 1st actually occurring or of a 3rd actually being exemplified. (In the preceding chapter, we were careful to say that an interpretant may be an aspect of a response to a sign.) Thus, a dynamic interpretant may be either emotional, energetic, or logical. Nor is there anything wrong with speaking of a potential occurrence of a 2nd or, therein, of a 1st or 3rd. Thus, the potentiality that an immediate or a final interpretant is, is either emotional, energetic, or logical, depending on the sign and the purpose.

Peirce's distinction between a sign's *immediate object* and its *dynamic object* was more clear, more constant, and less original than were his divisions of interpretants. It goes back to the Stoic distinction (chapter 1, section 6) between the object as represented, or *lekton* – the immediate object – and the object as it exists independently of its representation – the dynamic object. If I say, 'my mother', and my father says, 'my wife', the object of these two expressions is the same in herself but is represented differently,

as mother of the one and as wife of the other. You could apprehend the object as represented in either expression without knowing that it is the same person. The problem with this dichotomy lies in explaining (a) how signs signify something beyond their immediate objects and (b) how, in practice, we distinguish the two. As we saw, the Stoic view led to the skeptical conclusion that the real object of a sign is unknowable. That is reflected in Kant's doctrine of the *Ding an sich*. Peirce's central philosophical aim may be said to have been to defeat such conclusions.

1. Much Groping, No Conclusion

Before attempting a systematic exposition, let us review Peirce's struggle with the divisions of the interpretant. There is only one place, that I have found, where he named the emotional, energetic, and logical interpretants. That was in the 1907 MS318 (primarily at 5.474–5), wherein he wished to focus on just the one type of interpretant named 'logical'. For that purpose, it was convenient to label the three alternatives. Nevertheless, the trichotomy is clearly invoked, without benefit of labels, in other places and, especially, where the immediate/dynamic/final trichotomy is also invoked.

In a letter of 1904 to Lady Welby, the two trichotomies are presented, unlabeled, in successive paragraphs:

> [W]e may take a sign in so broad a sense that the interpretant of it is not a thought, but an action or experience, or we may even so enlarge the meaning of sign that its interpretant is a mere quality of feeling. (8.332)

That is the emotional/energetic/logical trichotomy. Next,

> I am now prepared to give my division of signs, as soon as I have pointed out that a sign has two objects, its object as it is represented and its object in itself. It has also three interpretants, its interpretant as represented or meant to be understood, its interpretant as it is produced, and its interpretant in itself. (8.333)

As indicated by the first sentence of the second passage, it was Peirce's desire to distinguish various classes of sign that drove his divisions of objects and interpretants; many other passages testify to the same motive. The same sentence indicates that he did not think he had already, in the preceding paragraph, presented the trichotomy of interpretants he proceeds to state. The latter is clearly the immediate/dynamic/final trichotomy, though the final is obscurely described as the interpretant

'in itself'. It would seem, then, that the two trichotomies were thought to be distinct.[1]

None of the divisions is formally labeled here, but later in the same letter Peirce referred to immediate and dynamic objects and to immediate, dynamic, and 'signified' interpretants (8.335–9), leaving poor Lady Welby to guess which label goes with which definition. Notice that the language used to describe the two objects, '[as] represented' and 'in itself', is duplicated in describing two of the three interpretants, but with much less clarity of meaning. Is it really the interpretant that is represented or to be understood? And what determines how it is *meant* to be understood? Finally, what does 'in itself' mean when applied to interpretants? But 1904 is only the beginning of Peirce's work on these distinctions.

The same kind of evidence for Peirce's having intended two distinct trichotomies of interpretant may be found in his 'Prolegomena to an Apology for Pragmaticism', in the 1906 *Monist*. Here, he first stated the emotional/energetic/logical trichotomy without labeling it as such and then, in the same paragraph, presented (for the first and only time in a published article) the other trichotomy, formally labeled:

> I have already noted that a Sign has an Object and an Interpretant, the latter being that which the Sign produces in the Quasi-mind that is the Interpreter by determining the latter to a feeling, to an exertion, or to a Sign, which determination is the Interpretant. [Thus the emotional/energetic/logical trichotomy.] But it remains [N.B. *remains*] to point out that there are usually two Objects, and more than two Interpretants. Namely, we have to distinguish the Immediate Object, which is the Object as the Sign itself represents it, and whose being is thus dependent on the Representation of it in the Sign, from the Dynamical Object, which is the Reality which by some means contrives to determine the Sign to its Representation. In regard to the Interpretant we have equally to distinguish, in the first place, the Immediate Interpretant, which is the interpretant as it is revealed in the right understanding of the Sign itself, and is ordinarily called the *meaning* of the sign; while in the second place, we have to take note of the Dynamical Interpretant which is the actual effect which the Sign, as a Sign, really determines. Finally, there is what I provisionally term the Final Interpretant, which refers to the manner in which the Sign tends to represent itself to be related to its Object. I confess that my own conception of this third interpretant is not yet free from mist. (4.536)

The immediate interpretant, we see, has been changed from what is 'meant' to the 'right understanding' of the sign. This is an improvement,

[1] If I seem to be belaboring this point – not only here but in much that follows – it is because I am belaboring it. I do so out of considerable frustration that it continues to be denied (e.g., by Liszka 1990, 1996, ch. 2), despite Short 1981b and 1982; cf. Short 1996a.

as the 1904 formulation implies that signs are produced with an intention to signify, which would not fit natural signs, such as symptoms. 'Right understanding' implies a purpose in interpretation, as well as something objective that determines whether that purpose is in process of being satisfied.

Whether the conception of the final interpretant was improved is another question; Peirce's self-confessed 'mist' seems real. There is much evidence, in unpublished manuscripts between 1904 and 1908, of a prolonged struggle with this idea. In the October 8, 1905, entry in Peirce's 'Logic Notebook' (MS339), the final interpretant is first named 'significant', but later in the same day, and in what appears to be one continuous act of writing, it is named 'representative' and is defined as representing the sign in respect to being a 'Rheme [term], Proposition, [or] Argument'. The next day, ' the Representative Interpretant is that which correctly represents the sign to be a Sign of its Object'. The entry for October 12 is to like effect. In the 'Logic Notebook' we witness Peirce thinking aloud, as it were. He himself, in his entry of March 23, 1867, wrote, 'Here I write but never after read what I have written for what I write is done in the process of forming a conception'.

The entry for April 2, 1906, is more helpful. Now the final interpretant is named 'normal':

> The Normal Interpretant is the Genuine Interpretant, embracing all that the sign could reveal concerning the Object to a sufficiently penetrating mind, being more than any possible mind, however penetrating, could conclude from it, since there is no end to the distinct conclusions that could be drawn concerning the Object from any Sign. The Dynamic Interpretant is just what is drawn from the Sign by a given Individual Interpreter. The Immediate Interpretant is the interpretant represented, explicitly or implicitly, in the sign itself.

As before, the dynamic interpretant is one actually formed, and the immediate interpretant, 'represented . . . in the sign itself', is what would be 'revealed in the right understanding' of the sign, presumably given knowledge of signs of that kind (e.g., knowledge of the language employed by a speaker, or experience of a correlation of cause and effect that grounds a natural sign). The normal interpretant goes beyond what is required for the immediate interpretant: it is 'all' that would be revealed to a 'sufficiently penetrating' mind.[2]

[2] 'Normal' is from the Latin *norma*, a carpenter's square, more generally, any standard; as the passage quoted makes clear, Peirce used the term in that sense and not in its democratically degraded sense, wherein the common is made normative. The normal

But what does such penetration require beyond knowledge of the sign itself? Peirce continued,

> I have omitted the *intended* interpretant. So far as the intention is betrayed in the sign, it belongs to the immediate Interpretant. So far as it is not so betrayed, it may be the Interpretant of *another* sign, but it is in no sense the interpretant of *that* sign.

What we need other signs to know cannot be part of a given sign's immediate interpretant. The implication, contrary to what the last sentence quoted might seem to indicate, is that it can be part of its dynamic and final interpretants. For other signs bear on 'all that the sign could reveal concerning the Object'. One sign may be interpreted in light of what we know from other signs. The final interpretant may be distinguished, then, as one to which other signs are also relevant.

That option is seized in another manuscript of 1906 (MS499), in which the term 'Normal Interpretant' occurs and in which Peirce employed the example of courtroom testimony. The witness tells his story 'without the slightest idea that it can be doubted'. To accept it in that way is its immediate interpretant (here named, not very helpfully, 'the Objective, or Naïve, or Rogate Interpretant'). 'The dynamical interpretant is the judgment of the fact which listening to the witness's testimony actually produces on the minds of the jury'. And, 'The Normal Interpretant is the modification of the verdict of the jury in which this testimony ought logically to result'. As a modification of a verdict based on other evidence, the final interpretant draws on information gleaned from other testimony, as well as from this testimony. Note that the final interpretant is the conclusion that 'ought' to be drawn, while the dynamic interpretant is the one actually drawn. The two can be the same, but even when they are, they differ in definition.

The immediate and final interpretants correspond, negatively, to the two kinds of interpretative mistake we have distinguished (chapter 6, section 4), of taking X to be a sign, S, that it is not and of taking a sign, S, that is false or misleading, at face value. If an interpretant, R, actualizes an immediate interpretant that X actually has (because grounded in a real relation of which X is a relatum), then a mistake of the first kind is not made. And a mistake of the second kind is not made if R actualizes or would be part of an actualization of the final interpretant. The

interpretant is not the one that would 'normally' (commonly) be formed, but is the one that ought to be formed.

jury may properly understand what the witness is saying, thereby actualizing that testimony's immediate interpretant, yet, by taking those words as the unvarnished and whole truth, fall short of actualizing their final interpretant.

MS318 omits mention of immediate, dynamic, and final interpretants altogether (in two places 'final' is used in the sense otherwise given to 'ultimate'). But that is because in that essay Peirce was not concerned with sign classification and did not need this other trichotomy. Some passages read as if he were trying to merge the two trichotomies under the emotional/energetic/logical rubric. One of these has been so influential (by which I mean that it has misled so many) that we must give it some close attention:

> Now the problem of what the 'meaning' of an intellectual concept is can only be solved by the study of the interpretants, or proper significate effects, of signs. These we find to be of three general classes with some important subdivisions. The first proper significate effect of a sign is a feeling produced by it. There is almost always a feeling which we come to interpret as evidence that we comprehend the proper effect of the sign, although the foundation of truth in this is frequently very slight. This 'emotional interpretant', as I call it, may amount to much more than that feeling of recognition; and in some cases, it is the only proper significate effect that the sign produces. Thus, the performance of a piece of concerted music is a sign. It conveys, and is intended to convey, the composer's musical ideas; but these usually consist merely in a series of feelings. If a sign produces any further proper significate effect, it will do so through the mediation of the emotional interpretant, and such further effect will always involve an effort. I call it the energetic interpretant. The effort may be a muscular one, as it is in the case of the command to ground arms; but it is much more usually an exertion upon the Inner World, a mental effort. It can never be the meaning of an intellectual concept, since it is a single act, [while] such a concept is of a general nature. But what further kind of effect can there be?
>
> In advance of ascertaining the nature of this effect, it will be convenient to adopt a designation for it, and I will call it the *logical interpretant*. (5.475–6; the bracketed word was inserted by the editors of the *Collected Papers*)

The description of the three interpretants as forming a series of responses to one and the same sign, progressing from recognition of the sign as such to a comprehension of its meaning, suggests a progression from an immediate to a final interpretant, as if Peirce had been trying to meld the two trichotomies.

However, there is no progression from immediate to dynamic to final interpretant: any dynamic interpretant must actualize the immediate interpretant (see below, section 2), and the progression – where one occurs – is from one dynamic interpretant to another, until, finally, the

final interpretant is made actual. Furthermore, each of the emotional, energetic, and logical interpretants is described in this passage as an interpretant actually formed, hence, as dynamic. Finally, a comprehension of a sign's meaning is merely an actualization of its immediate interpretant, not its final interpretant, which, as we have seen, goes beyond mere comprehension of meaning in that sense. For these three reasons, we must conclude that, if Peirce was trying to make the one trichotomy do the work of both, he failed.

Perhaps the emphasis, in this passage, on feelings and actions occurring as steps in a process of interpretation is due only to the topic, announced in the first sentence: 'the "meaning" of an intellectual concept'. For only 'intellectual' signs bear a meaning that is 'logical' or a 3rd. Now, as we know from Peirce's phaneroscopy, 3rdness incorporates 2ndness and 1stness. Hence, to account for the understanding of an intellectual sign, Peirce had to describe a sequence, from 1st to 2nd to 3rd, as in MS318. And even in that context, Peirce pointed out that feelings and actions, respectively, *can* be the complete interpretant of a sign, if the sign is, say, a piece of music or a command (about which more in section 5).[3]

Sometimes, Peirce's accounts of immediate, dynamic, and final interpretants are complicated by a distinct issue, as to whether signs comprise other things as well as words spoken by one person to another. It appears either that he saw semeiotic as a generalization of that narrower case or that he felt he had to begin with the narrower case to make his broader conception intelligible to neophytes. Here, for example, is a statement, admirably clear about the divisions that now concern us, that limits interpreters to persons:

> I define a *Sign* as anything which on the one hand is so determined by an Object and on the other hand so determines an idea in a person's mind, that the latter determination, which I term the *Interpretant* of the sign, is thereby mediately determined by that Object. A sign, therefore, has a triadic relation to its Object and to its Interpretant. But it is necessary to distinguish the *Immediate Object*, or the Object as the sign represents it, from the *Dynamical Object*, or really efficient but not immediately present Object. It is likewise requisite to distinguish the *Immediate Interpretant*, i.e. the Interpretant represented or signified in the Sign,

[3] In 'Logic Notebook' entries of October 21 and 23, 1906, Peirce also wrote as if he were trying to merge the two trichotomies, this time under the immediate/dynamic/final rubric. But even then, each subdivides triadically, sometimes as being 1sts, 2nds, or 3rds and other times in relation to 1sts, 2nds, or 3rds, respectively. E.g., on October 23, he wrote, 'The *Dynamical Interpretant* is the actual effect produced upon a given interpreter on a given occasion in a given stage of his consideration of the sign. This again may be 1st feeling merely, 2nd an action, or 3rd a habit'.

from the *Dynamic Interpretant*, or effect actually produced on the mind by the Sign; and both of these from the *Normal Interpretant*, or effect that would be produced on the mind by the Sign after sufficient development of thought. On these considerations I base a recognition of ten respects in which signs may be divided. (8.343)

That was in a draft letter, dated December 24, 25, and 28, 1908, never sent to Lady Welby. In a letter of December 23, 1908, actually sent,[4] Peirce gave a similar definition of 'sign', but then added, 'My insertion of "upon a person" is a sop to Cerberus, because I despair of making my broader conception understood' (LW:81).

Be that as it may, in the 1907 MS318, Peirce had already tried out the idea that the concept of a sign's object is a generalization from that of a word's utterer, and similarly that the idea of an interpretant generalizes from persons who interpret:

> [S]igns mostly function each between two minds, or theatres of consciousness, of which one is the agent that *utters* the sign . . . while the other is the *patient* mind that *interprets* the sign. . . . [B]efore the sign was uttered, it already was virtually present to the consciousness of the utterer, in the form of a thought. But, as already remarked, a thought is itself a sign, and should itself have an utterer . . . and so back. . . . [I]n some cases, neither the series of utterers nor that of interpreters forms an infinite collection. When this is the case, there must be a sign without an utterer and a sign without an interpreter. . . . But why argue, when signs without utterers are often employed? I mean such signs as symptoms of diseases, signs of the weather. . . .
>
> Having found, then, that neither an utterer, nor even, perhaps, an interpreter is essential to a sign, characteristic of signs as they both are, I am led to inquire whether there be not some ingredient of the utterer and some ingredient of the interpreter which not only are so essential, but are even more characteristic of signs than the utterer and the interpreter themselves. (EP2:403–4)

There is nothing in this passage to suggest that Peirce held that purposefulness is one of the utterer's features retained in the generalized conception of a sign's object. The examples of medical symptoms and meteorological signs tell strongly against such an hypothesis.[5]

[4] It would appear that the unsent drafts of December 24–8 are drafts of the letter (mis- or post-) dated December 23; cf. chapter 9, section 4.

[5] Joseph Ransdell's reading (1976, 1977) is that the object of a sign somehow produces the sign purposefully, as if with an intention to have itself signified (the World as Narcissus?). That reading does not depend on this one passage alone. In the published statement of 1906 quoted above, Peirce wrote that the dynamic object 'is the Reality which by some means *contrives* to determine the Sign to its Representation' (4.536, emphasis added), and in other passages as well the same idea may be found. Ransdell's interpretation is grounded in much that is suggestive. But do all suggestions have to be taken?

The variations and struggles with subsidiary issues never end. In a draft letter of 1906 to Lady Welby, Peirce wrote,

> There is the *Intentional* Interpretant, which is a determination of the mind of the utterer; the *Effectual* Interpretant, which is a determination of the mind of the interpreter; and the *Communicational* Interpretant, or say the *Cominterpretant*, which is a determination of that mind into which the minds of utterer and interpreter have to be fused in order that any communication should take place. This mind may be called the commens. It consists of all that is, and must be, well understood between utterer and interpreter at the outset, in order that the sign in question should fulfill its function. (LW:196–7)

This is a helpful way of looking at the nature of communication (see Bergman 2000), but it hardly fits the full breadth of Peirce's semeiotic; and yet the context in which it is presented suggests that, at the moment, Peirce was entertaining that idea in lieu of the immediate/dynamic/final trichotomy.

On November 1, 1909, midway in his seventy-first year and at the end of his 'Logic Notebook', Peirce tried two times to define 'sign'. His commentators usually write as if his concept of sign remained unchanged from 1867. The first try is followed by the words, 'Well, on the whole, or rather not on the whole by any means, but as another phase of reflexion, I think this won't do'. The next definition – the Notebook's last – is crossed out.

2. Immediate, Dynamic, and Final Interpretants

Having reviewed many of Peirce's variant schemes and failed experiments, I think we may now be allowed to focus on the strands of his thought that appear essential and to ignore others that do not cohere with them. As indicated earlier, the strands we shall hold to be essential are those that (a) are implied by the 1907 account of significance and (b) are required for the 1903–8 taxonomy of signs.

In a letter to Lady Welby (of March 14, 1909, seven and a half months before his final, failed attempts to define 'sign' in his 'Logic Notebook'), Peirce declared that his 'three grades of Interpretant were worked out by reasoning from the definition of a Sign' (LW:111). That definition was not given in the letter; the three grades he went on to describe are the immediate, dynamic, and final interpretants.

'My Immediate Interpretant is implied in the fact that each Sign must have its peculiar Interpretability before it gets any Interpreter' (ibid.).

Ergo, a sign does not have to be interpreted in fact; being interpretable – not arbitrarily, in any way, but in some 'peculiar' way – suffices to make it a sign.[6] The immediate interpretant is that peculiar interpretability. 'The Immediate Interpretant is an abstraction, consisting in a possibility'. Like any possibility, an immediate interpretant can be actualized in diverse ways.

An interpretant actually formed is dynamic. 'The Dynamical Interpretant is a single actual event' (ibid.). Any dynamic interpretant of a given sign must actualize that sign's immediate interpretant, in one way or another; otherwise, it would not be an interpretant of *that* sign. And, yet, dynamic interpretants of the same sign will differ from one another: 'My Dynamical Interpretant is that which is experienced in each act of Interpretation and is different in each from that of any other'.

An implication of the preceding is that there is a sense in which a sign cannot be misinterpreted. But it is a peculiar sense and has been the source of some confusion. Let me explain. Reference to a sign, S, can be either opaque or transparent. Used transparently, 'S' refers to the X that is, or is thought to be, that S and that might be another sign as well or instead. Used opaquely, 'S' refers to X only as being S. If I shout to you, warningly, 'Duck!' and you take me to be making a joke, then I might explain your prostrate form to curious passersby this way: 'He took my warning to be a joke'. But, of course, it is the word, X, I shouted that you took to be a joke. Had you taken X to be a warning (a genuine warning), you could not have taken it to be a joke (a joking, pretend warning), as it could not be both. In saying that you took my warning to be a joke, I am using 'warning' transparently. In saying that you could not take a warning to be a joke, I am using 'warning' opaquely. Using 'S' opaquely, S cannot be interpreted at all without actualizing S's immediate interpretant; in that sense, it cannot be misinterpreted. But this does not entail infallibility. A mistake can be made, namely, in getting X wrong, by taking it to be a sign that it is not. Using 'S' transparently, we can say that someone took S the wrong way, as being a sign it is not.

'[E]ach Sign must have *its* peculiar Interpretability', Peirce said. He did not mean that the interpretability is peculiar to the sign (i.e., shared by no other sign), which would be false, but that the sign (using the word opaquely) has only that one interpretability. A different immediate interpretant would constitute a different sign. It follows that something, X, having diverse immediate interpretants, $Ri_1, Ri_2, \ldots$, is that many distinct

[6] See also passages of 1902 and 1903 quoted in chapter 2, section 9.

signs, S_1, S_2,. . . . That fits the way we ordinarily count signs. Suppose that in nineteenth-century Arizona a rancher observes some puffs of smoke on the horizon: they are a sign of fire; they are also a sign that the Apaches are on the warpath. We count these as two signs, albeit they are the same smoke.

As a sign's immediate interpretant is specific, there must be something that determines it. That is what we have called the sign's prior relation to its object, or the ground of its significance (chapter 6, section 5). Different grounds make the same thing different signs: a causal regularity makes the smoke a sign of fire, a conventional code makes it a sign of war. Now, interpretation presupposes purpose, and therefore an immediate interpretant presupposes a possible purpose – a purpose relative to which an interpretant of that type could be justified on the corresponding ground. But sometimes different purposes, even antagonistic ones, overlap, making something the same sign on the same ground. Thus, Apache and rancher, alike, have an interest in decoding the smoke signal, but the one in order to join an attack, the other to prepare a defense. The two, opposed actions actualize in different ways the same immediate interpretant.

Suppose that a dynamic interpretant, Rd, gets X right: X is indeed the sign, S. As we have noted before, there is a second way in which Rd may nonetheless be mistaken. It can fail to be the right interpretant for its purpose. For we do not interpret signs merely to understand them. We interpret signs in order to learn from them. We seek knowledge or guidance, and so on. When I shouted, 'Duck!', I did so with the purpose of saving your life, a purpose that I assumed you shared. If, with that purpose, you heeded my warning, but did not duck low enough or fast enough, then your response failed its purpose, even though it did properly interpret my exclamation as being a warning. The same immediate interpretant can be actualized in different ways – in this case, by your ducking faster or lower, as you might have done had you had additional information.

Or suppose that I did intend my exclamation as a warning, but as a false warning, meant to distract you so that you would not notice the automobile, driven by my accomplice, that is running you down. Then, in failing to detect the falseness of my warning, you failed your purpose, of preserving your life, since you were run over while ducking a projectile that was not there. To think, 'I am being warned – falsely!', would also have been to interpret my exclamation properly, as a warning, even while correcting its message.

Thus we have the idea of a final interpretant, the interpretant ideally adequate to the purpose for which the sign is being interpreted. But here we run into a difficulty. Is there one final interpretant per sign, corresponding to the overlap of purposes to which that sign is germane, or are there different final interpretants of the same sign, relative to the different purposes with which it may be interpreted? Peirce struggled with the concept of a final interpretant, but for other reasons; he missed this issue, probably because he usually had in mind one purpose only, which was to know reality.

In the same letter to Lady Welby, he continued, 'and the Final Interpretant is the one Interpretative result to which every Interpreter is destined to come if the Sign is sufficiently considered. . . . The Final Interpretant is that toward which the actual tends' (ibid.). The picture evoked is that of scientific inquiry, conceived by Peirce as an indefinitely prolonged 'fixation of belief' carried out by an indefinitely extended community of inquirers, all of whom have the same ultimate purpose.[7] But some signs, such as 'Duck!', have to be interpreted the right way at first, or not at all. And where ultimate purposes differ, as practical purposes usually do, agreement in interpretation may be limited to actualizing the sign's immediate interpretant. There is then no final confluence of interpretations.

Peirce often wrote as if the whole truth about a sign's object is its final interpretant (e.g., at 8.314), again reflecting his preoccupation with science. That does not comport with the breadth of his semeiotic, wherein interpretants, including final interpretants, may be actions, feelings, or habits, as well as representations: nonrepresentations cannot be truths. But science, too, is not interested in the whole truth. It seeks understanding or explanation or a general theory – any of which requires one to abstract from vast quantities of detail. Ignoring, forgetting, putting to one side, refusing to consider are almost as important to scientific understanding as is knowing.

We are compelled to amend Peirce on this topic. We shall have to allow distinct final interpretants of a given sign, S, relative to the different purposes for which that same sign may be interpreted. And in none of these would the whole truth about S's object be represented.

[7] The choice of the term 'final' for what he elsewhere named the 'normal' or 'significate', etc., interpretant is thereby explained. In 'The Fixation of Belief', Peirce identified truth as the final fixation of belief toward which scientific inquiry tends (chapter 12, section 4). We will retain the term 'final' even though it does not fit all cases of ideally adequate interpretant.

3. Immediate and Dynamic Objects

In a passage of 1904 quoted earlier, Peirce wrote that 'a sign has two objects, its object as it is represented and its object in itself'. Similarly, in a 1909 draft letter to William James:

> As to the Object, that may mean the Object as cognized in the Sign and therefore as Idea, or it may be the Object as it is regardless of any particular aspect of it, the Object in such relations as unlimited and final study would show it to be. The former I call the *Immediate* Object, the latter the *Dynamical* Object. For the latter is the Object that Dynamical Science (or what at this day would be called "Objective" science) can investigate. (8.183; identified at EP2:492)

The in-itself, no longer unknowable, is what science can discover the nature of. Again, in a 1908 draft letter to Lady Welby: 'It is necessary to distinguish the *Immediate Object*, or the Object as the sign represents it, from the *Dynamical Object*, or really efficient but not immediately present Object' (8.343). Here, the adjective 'dynamical' is explained as being due to the object's dynamism (its being 'really efficient') rather than to its being studied by 'dynamical science'. But it comes to the same thing.

Evidently, a sign's immediate object corresponds to its immediate interpretant; when interpretation is intellectual, the object as the sign represents it is the object as it would be understood in that sign's immediate interpretant. The words 'as unlimited and final study would show it to be' suggests that the dynamic object corresponds in similar manner to a sign's final interpretant. But we have seen that final interpretants may be many and are always less than what unlimited and final study would result in.

The dynamic object was perhaps introduced to explain the difference between success and failure (more narrowly, between truth and falsehood). The success of an interpretant, given its purpose, depends on reality, not on representation. The immediate object is the Stoics' *lekton*, Dion as represented, while the dynamic object is the real Dion, Dion as he exists independently of being represented. But we can only know Dion as represented, as all knowledge is representation. If that is true, then how can reality be distinguished from representation? To posit a reality that is unknowable is empty verbiage, metaphysical nonsense, Kant's *Ding an sich*. Thus all reality would seem to be representation. We are back to idealism again, semiotic idealism specifically.

On the same day, March 14, 1909, that he wrote the letter to Lady Welby quoted in the preceding section, Peirce also wrote to William James:

> We must distinguish between the Immediate Object, – i.e. the Object as represented in the sign, – and the Real (no, because perhaps the Object is altogether fictive, I must choose a different term, therefore), say rather the Dynamical Object, which, from the nature of things, the Sign *cannot* express, which it can only *indicate* and leave the interpreter to find out by *collateral experience*. For instance, I point my finger to what I mean, but I can't make my companion know what I mean, if he can't see it, or if seeing it, it does not, to his mind, separate itself from the surrounding objects in the field of vision. (8.314)

The essential move made in this passage is to distinguish between the immediate and dynamic objects in terms of the process of interpretation. In that way, the distinction is not attempted, or pretended, to be made from some vantage point, such as God's, that we do not in fact possess. It is made from within our experience, as interpreters of signs. (That is why, in this chapter, we first discussed interpretants, before discussing objects.) The consequence is that the dynamic object is not presented as something that might very well turn out to be unknowable; it is by conception something knowable, if never known completely.

The words 'indication' and 'collateral experience', emphasized by Peirce, are key. As we can see from the example provided, indication is indexical, but not purely so. It is by indices plus icons or symbols[8] that we pick out particular objects of general description. The index points us in a direction, the symbol or icon tells us what to look for in that direction. Such objects have a spatio-temporal location.[9] Having such a location, they can be picked out again, by other indices linked to other icons or symbols.[10] Thus there is collateral observation of the object of the original sign. Collateral experience consists in an interpretation of

[8] The icon/index/symbol trichotomy, introduced in chapter 2, section 8, is stated more formally in chapter 8, section 2. Roughly, an icon signifies by exemplification or resemblance, an index by a causal or other existential relation, and a symbol by an instituted rule, conventional, customary, or instinctual.

[9] Space and time are a frame within which all objects indexically signified may be located; but do not suppose that indices presuppose that frame, as in Kant's theory. On the contrary, our apprehension of space and time, as a stable framework embracing all that is actual, emerges from successful reidentifications of objects indexically signified and, hence, from coordinating indices successfully. As in experience, so in reality: time and space involve 3rdness as well as 2ndness. That there is such an order is a contingent fact. But for it, we would know no facts; it is contingent nonetheless.

[10] 'Linked' is often too weak. An index may be or embody an icon or it may replicate a symbol; chapter 8, section 2. Thus observation; chapter12, section 6.

diverse signs as being signs of the same object. It requires the identification of the object of one sign with that of another, and that, in turn, requires that each sign – that is, each sign having an object of which collateral observation is possible – be an index or have an indexical component.[11]

Since different signs of the same (dynamic) object represent it differently, correction and supplementation of a given sign becomes possible. It is easy to see how supplementation is possible, under the conditions described. But how can a sign be corrected by another of the same object? Correction is possible because a false description can contain enough truth to enable one to identify the object misdescribed, given that a place or direction is indicated. 'Look at the monkey!' you exclaim. 'That', I regretfully respond, 'is no monkey; it is your long-lost son'. I looked in

[11] From 1907 to 1911, Peirce made several references to collateral observation or experience. There appear to be some serious inconsistencies among them. For the most part, collateral experience is presumed to be prior to the interpretation of a proposition, providing acquaintance with the object that it is about or, generalizing from this, that any sign 'denotes' (e.g., 8.178–9). The ideas of object and of denotation assumed here are those of actual individuals and indexical reference to them (EP2:404–9, 6.138, 338), from which it would follow that denotation is not a property of all signs. At one place, Peirce seems unsure about this: 'no sign can be understood – or at least no proposition can be understood – unless the interpreter has "collateral acquaintance" with every Object of it' (8.183).

In the same draft letter of 1909 (8.177–85, EP2:492–7), collateral observation is definitely distinguished from 'acquaintance with the system of signs', the latter being 'the prerequisite for getting any idea signified by the sign' (8.179). Peirce's example (simplified), 'Napoleon was lethargic', is helpful up to a point: acquaintance with the language enables us to identify 'Napoleon' as being a proper name, but to know who it names, hence, what the proposition is about, we must have some other knowledge of that individual – knowledge not provided by the proposition itself (which, to make the point more obvious, is outrageously false). But then Peirce extended this idea, in still the same document, to other substantive terms, such as 'lethargic'. We do not know what that word means if all we know of it is that it is an adjective. That makes sense, too, but, then, how can acquaintance with the system of signs be distinguished from what collateral observation is necessary to?

I have emphasized the references to collateral observation that make it subsequent to, or, at least, not implicated in, understanding the sign, rather than prior to and necessary to that understanding. Thus, in a 1908 letter to Lady Welby, Peirce spoke of the dynamic ('Dynamoid') object as that which the sign 'must indicate by a hint; and this hint, or its substance, is the Immediate Object' (LW:83). I take the immediate object to be that which acquaintance with the system of signs enables us to apprehend. Following its hint, we can then by collateral observation acquire other experience of (or relate other experience already had to) what we take to be the same object otherwise represented, hence, as the dynamic object common to diverse immediate objects.

In the most interesting class of signs, the two views of collateral experience coalesce (see chapter 10), which perhaps explains Peirce's equivocation.

the direction you indicated and saw nothing monkey-like in that direction but your son, with whom I had intended to reunite you. Hence the correction.[12]

Icons and symbols are general, in the different senses associated, respectively, with 1stness and 3rdness (chapter 3, section 5). By themselves, signs of neither kind can signify anything that may be discovered to have a character further or different from that which they represent. By association with indices, however, they can. The portrait with a name attached, or by other means having a known history, can be compared with the original or with other representations of the same person. A symbol, which is a type that represents a type, can be replicated, its replicas signifying, indexically, instances of the type represented.[13] These, investigated, may change our idea of their type (chapter 10).

We see, then, that collateral experience is not restricted to individual objects and yet that, when the object is not individual, collateral experience of it is possible only through indices of individuals, namely, those that exemplify the type or embody the quality. The same applies to the ideal entities of mathematical knowledge, whose properties can be demonstrated only through experiments on individual diagrams, whether geometrical or algebraic; the diagram is physical, not ideal, the diagrammed is ideal, not physical. The same applies to fictional entities, so far as their nature is fixed in song, story, or image; for it is individual exemplars of the latter that we may hear, read, or view. Huck Finn is not real and yet it is true that he was a white boy living in Missouri in the early nineteenth century who befriended a runaway slave, and was befriended by him, and felt guilty about it; as he was not real, there is no truth about, say, the color of his eyes or anything else that is not recorded or implied in Twain's novels.[14]

[12] The correction depends, of course, on our being able to distinguish between signs, or between mutually inconsistent sets of mutually consistent signs, by their degree of reliability. Reliability is again determined not from an impossibly godlike perspective from outside our experience, but from within experience, in our efforts to establish maximal coherence of interpretation. Signs whose immediate interpretants do not fit, uncorrected, within that coherence are deemed false; those that are of a kind that are often false are unreliable. Judgments of truth and reliability are themselves fallible, i.e., subject to correction in our ongoing attempt to wrest order from chaos.

[13] Signs replicable, their replicas, and the differences between them and other signs are discussed in chapter 8, section 1.

[14] Two other marks of Huck's unreality, in addition to his being indeterminate where undescribed, is that he can have incompatible properties if described inconsistently and that we can choose which novels, etc., define him. If a film version of Huck has blue eyes

In all of these cases, the spatio-temporal world of real individuals, existing dynamically, plays a central role. Without that world, there would be no indexicality and no collateral experience and, hence, no objective inquiry or objective knowledge. But that is not to say that all that we have collateral experience of reduces to the individuals through which we have it. Huck Finn is not piles of carbon on a page. Nor is a law reducible to the individual events that attest to its reality.

The distinction between immediate and dynamic objects opens up the darksome pit of metaphysical disputation, deferred to the next section. For the present, let us determine where the dynamic object fits into the formal scheme of Peirce's semeiotic. We know that each sign has an immediate object, corresponding to its immediate interpretant, but does each sign also have a dynamic object? Is a sign's dynamic object always distinct from its immediate object? And do distinct final interpretants correspond to distinct dynamic objects or may different such interpretants share a dynamic object?

We begin with the last of these three questions. In the same long paragraph to James from which we have just quoted, Peirce later explored an example, wherein his wife,[15] on waking, asks him what sort of day it is. Her purpose in asking pertains to her plans for the day, and so he does not respond at length but merely says that it is stormy. The dynamic object of his response, he wrote, is the real meteorological conditions at the moment. Those could be the object of a study limited only by time, resources, and human competence, but the final interpretant of his response, for his wife's purpose, is her making the right choice as to whether to take her umbrella with her. Even were his wife a meteorologist, interested in refining a theory about storms, the final interpretant of her husband's report would not be what Peirce said it was, 'the sum of the *Lessons* of the reply, Moral, Scientific, etc.' (8.314).

In saying that, he evidently wanted to use the idea of a final interpretant (qua whole truth) to define a sign's dynamic object, namely, as that which corresponds to the final interpretant. We, to the contrary, shall define the dynamic object as that which (a) can be signified by diverse signs and (b) accounts for any difference there may be between final and immediate interpretants. It can account for the latter without being exhaustively

and we choose to make that film canonical, then his eyes do have a determinate color. A difference between physics and theology is that in physics there are no canonical texts.

[15] The actual words are 'suppose I awake in the morning before my bedfellow, and that afterwards she wakes up and inquires, "What sort of day is it?"' (EP2:498). In the *Collected Papers*, 'bedfellow' is silently emended to 'wife' (8.314). *Autres temps autres moeurs.*

represented by any final interpretant of any of its signs. Thus, to different final interpretants of the same sign, as well as to the final interpretants of many different signs, there will correspond the same dynamic object. This amendment preserves the idea that the dynamic object is 'the Object as it is regardless of any particular aspect of it, the Object in such relations as unlimited and final study would show it to be'.

For reasons already noted, collateral observation, although it involves indices, is not limited to the actualities that are indices' direct objects. There is, then, no reason to suppose that there is any sign lacking a dynamic object, of which the immediate object is only an aspect or presentation. The point is made by the least likely example: colors presented in samples really are arranged in one way on the spectrum and really do have a physical explanation, about either of which one might make a mistake.

However, while no sign represents its dynamic object completely, many signs, such as pure icons and pure indices, cannot misrepresent their objects. For the object of such a sign is exactly whatever is presented or indicated. As no further representation is made within that sign that might be false of that object, the sign cannot be mistaken or misleading. In that respect, while there will be differences, there can be no discrepancy between such a sign's immediate and dynamic objects.

A last word on this distinction: the immediate and dynamic objects are not different entities. The distinction pertains, rather, to how one and the same object is considered. The immediate object *is* the dynamic object *as* it is represented, however incompletely or inaccurately, in a given sign.

4. Peirce's Realism

It might be supposed that, as the distinction between immediate and dynamic objects is made within the process of interpretation, it entails idealism. The dynamic object, it might be said, so far from being 'external' to the sign process, is an artifact of it. That would make Peirce's theory a semeiotic gloss on Kant's conception of physical objects, according to which they are products, as it were, of an intellectual synthesis of the sensuous manifold.[16] The distinction between immediate and dynamic objects would be Kant's between the subjective (unstable and private)

[16] Or, if we take additional features of Peirce's semeiotic into account, that introduce further changes, we might say, rather than that it is a gloss on, that it is a semeiotic transformation of Kant (Apel 1980, ch. 3).

and the objective (stable and interpersonal) but equally phenomenal. The idealist argument is ever so. It is always that any meaningful distinction must be internal to thought or to experience or to the process of representation (right, so far), and hence (here is where the trouble begins) that any meaningful idea of the external makes it to be internal, after all.

One could cite Berkeley or what Kant's refutation of Berkeley (KdrV, B274–9) left unrefuted (1.35–9). But consider a more recent and more semiotical idealist, Nelson Goodman:

> If I ask about the world, you can offer to tell me how it is under one or more frames of reference; but if I insist that you tell me how it is apart from all frames, what can you say? We are confined to ways of describing whatever is described. Our universe, so to speak, consists of these ways rather than of a world or of worlds. (1978, pp. 2–3)

I suggest that this transition, from truism (all but the last sentence) to absurdity (the last sentence), is a non sequitur. Nonetheless, it poses a challenge to the realist: to *show how* a description of the world as being distinct from all descriptions of it can be consistent with itself.

The consistency at issue is not between propositions but is between a proposition and its assertion. Suppose Mugsy mutters, 'I ain't talkin''. Taking his statement literally, the proposition Mugsy asserts – that Mugsy is not talking – is perfectly consistent; it is something that could be true. But it is not consistent with Mugsy's assertion of it. Mugsy asserting it makes it false. In effect, Berkeley and Goodman accuse realists of that kind of inconsistency. As soon as the realist says that there is or might be a world independent of all conception of it, the idealist leaps to his feet, shouting 'Gotcha!' For, in speaking of a world unconceived, the realist is conceiving of it (Berkeley, *Treatise*, Pt. I, §23). The idealist point is that there is no way of affirming realism without being guilty of that kind of inconsistency.

There is a currently popular type of argument for what is sometimes named 'scientific realism'. Realism is presented as a hypothesis that the world is pretty much as science describes it; it describes it, at least in its physical aspects, as existing independently of our descriptions of it. The argument is that this hypothesis is the best explanation for science's success (Boyd et al., in Leplin, 1984). If there were no entities that exist independently of our theories and that produce sensations in us regardless of whether the laws of their action are known to us, then why do our experiences so consistently support some of those theories (and enable

them to be made ever more precise and detailed) and contradict the rest? This argument is, I believe, both sound and thoroughly Peircean (chapter 2, section 7; chapter 12, section 6). But it does not address the idealist counterargument that talk of independent existence is meaningless. The argument is sound *if* meaningful. In short, this argument, often said to be 'internal' to science, fails its purpose, if its purpose is to refute idealism.

We have no room here for a thorough defense of realism against the charge of meaninglessness,[17] but we can suggest the lines a Peircean defense might take. To begin with, phaneroscopic analysis exhibits the experiential roots of the concepts of independent existence (2ndness: chapter 3, section 4) and of what would be or would have been (3rdness: chapter 3, section 8). It shows that *haecceity* does not reduce to our experience of it, and that lawfulness does not reduce to experienced regularity. Thus, although conjectural, it is not meaningless to assert that there is something we have experienced that is other than our experience of it, and it is not meaningless to assert that that is something that would exist unexperienced, unperceived, unrepresented. That the object of an indexical sign exists independently of that sign, and that it has a character independent of any symbol we replicate in describing it, is a meaningful hypothesis. As to whether the hypothesis is true, that is another matter. The argument cited above, that it is the hypothesis that best explains the progress of science, is a compelling argument for its truth, once its meaningfulness is granted.

But what of the idealist objection that we contradict ourselves when we assert that there is (or even that there might be) something unconceived or unrepresented? Here again, we must turn to the category of 3rdness and, as well, to the related idea of symbolic signs. A symbol is a sign general in meaning; thus, when applied to a particular object, whether an individual or a natural kind, it is capable of being supplemented and made more specific (chapter 8, section 2; chapter 10). For example, when rationality is predicated of the species *Homo sapiens*, we can ask, 'In what way are humans rational and to what degree?' The idea of human rationality can be made more precise, detailed, specific, concrete. Hence, symbols can be used to represent that which is unrepresented. Put less paradoxically, a symbol may be used to represent that which is not yet

[17] Not surprisingly, there is more to what Goodman has to say on the topic than the one passage quoted above, and there are different ways of parsing his arguments; cf. Putnam 1992, pp. 109–23, and my comment thereon in chapter 11, note 6.

represented precisely or specifically. Thus we can know some things that we don't know. We know that the red ball lying in the corner of the room has an inside of unknown color and a diameter of unknown length. Peirce's reference to an 'external permanency' (W3:253, see chapter 2, section 7) requires no more than that; he is not claiming that there is something utterly unconceived or in no respect represented. We see, however, that the reality of what would be and of what would have been is implicated in the hypothesis of an external reality or, if you wish, of scientific realism. For, to say that the ball has a diameter unknown is to imply that it would prove to have a determinate measure (within fairly narrow limits of precision) were it measured.

The dynamic object is exactly that about which more can be learned. Therefore, it must be independent of our experience of it. Peirce's conception of the dynamic object, formed after 1903, embodies his slowly developed realism. That realism, as Max Fisch wrote (1986, p. 195) has three components: his long-standing 'Scholastic realism', affirming the reality of 'generals', both 1sts and 3rds; his adoption in the 1890s of Scotus' idea of *haecceity* or 2ndness; and his view, formed not much later, that the reality of law is not reducible to any set of actual events but is subjunctively and counterfactually conditional. This last divides the reality of generals into two kinds, that of mere possibility or what may be (1sts), and that of potentiality or what would be (3rds). The independent existence and yet cognizability of the dynamic object – its being signifiable by diverse signs – involves all three of the phaneroscopic categories. I shall refer to this, simply, as Peirce's realism.[18]

For two reasons, it would be misleading to name Peirce's realism 'scientific'. First, most of those who subscribe to scientific realism harbor a nominalistic chariness of realism with respect both to generals and to counterfactuals. Second, most of the same suppose that science is mechanistic, fundamentally. Hilary Putnam introduced the term 'internal realism' for scientific explanations of science's success, opposing it to 'metaphysical realism', a view he finds 'incoherent' (1978, p. 123). But neither of these

[18] Fisch's 'three-category realism' is needlessly prolix, as is the Peircean label 'extreme Scholastic realism' that Susan Haack prefers (Haack 1992). Haack makes the important observation that Peirce's 'position is that there are real generals, not that generals are real', contrary to what Scholastic realism is usually thought to assert (p. 23). In the distinction of a real general, such as a natural kind, say, horses, from a merely possible kind, say, pink horses, all three categories come into play. But Peirce also made some possibilities more real than others (2.149, 5.453), so an exact statement of his realism would be subtle and complicated.

terms fits Peirce's realism. Putnam attributes three distinguishing characteristics to metaphysical realism, at least one of which Peirce accepted and another of which Putnam rightly says he rejected (p. 125). That implies that Peirce occupied an intermediate position that Putnam presumably thinks is untenable. The three characteristics are (a) that reality 'transcends complete formalization in any one theory', (b) that 'the world is independent of any particular representation of it', and (c) that a theory held at 'the ideal limit' of inquiry might yet be false. Peirce affirmed (b); to deny it is idealism. Indeed, there is nothing in Putnam's internal realism that Nelson Goodman could not happily accept. And Peirce rejected (c). The possibility of affirming (b) while denying (c) is precisely what Peirce's realism is all about.[19]

5. Emotional, Energetic, and Logical Interpretants

The purposefulness of semeiosis is reflected in the immediate/dynamic/final trichotomy of interpretants, and its realism is reflected in the corresponding distinction between immediate and dynamic objects. Purpose requires realism. One cannot act purposefully in a world that varies with our whims. Where purposeful action occurs, we have semeiosis; where it is possible, we have signs. We have now to determine how broadly this theory applies. Can actions and emotions, as well as statements and thoughts, be divided into the immediate, the dynamic, and the final? If so, what are their corresponding objects?

We have anticipated this discussion, to a degree, in our general account of significance, and any further such discussion must anticipate, in turn, the next two chapters, on Peirce's taxonomy of signs. The theory of signs is an organic whole: it cannot be understood without knowing the parts and the parts cannot be understood except in relation to one another. One must therefore begin with the parts, partially understood, or with the whole, partially understood, and then revisit those topics with improved understanding, as one proceeds.

Taking thoughts and other logical interpretants to be unproblematically acceptable as interpretants, let us now attend exclusively to energetic and emotional interpretants. Peirce used only a few examples, repeating

[19] And what of characteristic (a)? That is more important than might at first appear. The three characteristics are joined inseparably in Kant's notion of the *Ding an sich* and in Thomas Kuhn's late Kantianism, but we shall see later (chapter 12, section 7) that Peirce rejected the false dichotomy that, at least in Kuhn's and Putnam's reasoning, (a) stands on one side of.

them many times over without developing them in any depth. His prime example of an energetic interpretant is the action performed in response to the command, 'Ground arms!'; that utterance is, then, the sign interpreted in the action, of setting rifle butts on the ground. His prime example of an emotional interpretant is the feeling evoked by a musical performance. Had we time, I would apply Peirce's doctrine about music to other forms of art and explore the implications thereof for various issues in the philosophy of art. And I would apply what he says both about music and about military commands to moral judgment and moral experience; for moral judgment seems designed both to elicit feelings and to command action. However, we shall have to limit the discussion almost entirely to music and to military commands, making the argument that in those cases, at least, there are distinctions between immediate and dynamic objects and among immediate, dynamic, and final interpretants. Whether the realism thus entailed applies elsewhere must be left as a question for another author.

When his officer barks, 'Ground arms!', a soldier might think, 'I'm being told to bring my rifle butt down', and then, after a perhaps brief deliberation, he does as he is told. In that case, the command is interpreted by a thought, a logical interpretant that is itself a sign, and the action interprets the thought. Deliberation, however, is not desired by the officer. And where deliberation is not wanted, thought is unnecessary. The well-trained soldier will simply obey. In that case, the command's interpretant is energetic. Of course the soldier must understand the words. He must be capable of thinking about what they mean. But he does not have to repeat them over to himself or translate them into other words in order to act on them. It would be a mistake to suppose that interpretation must always be intellectual. If it were, we could never get from thought to action. Words (spoken or thought) must sometimes translate into action immediately if any are to do so eventually (cf. Aristotle on the 'practical syllogism', *De motu an.* 7).

Although performed automatically, the action has a purpose and the purpose is that of the soldier, not his commanding officer. The officer has a purpose in commanding that action, which, normally, is the purpose he thinks the action will serve; but the purpose with which the action is performed is that of he who performs it. The soldier's ultimate purpose might be to be a good soldier, part of an effective fighting unit, or to beat hell out of the enemy – purposes presumably shared by the officer – or (an unwilling soldier) it might be to avoid punishment. But the proximate purpose is none of those; it is simply to do as ordered to do,

that is, to perform an action of that type. The type of action commanded therefore determines whether the action is successful, in its proximate, not its ultimate, aim. That type of action is therefore the dynamic object signified; for a dynamic object is that which determines whether an interpretant succeeds or fails. The relation of interpretant to dynamic object that determines success or failure varies from case to case. In the case of commands and a purpose to obey, the relation is one of fulfillment.

There is, however, an ambiguity here. The command expresses the will of the officer, but not always adequately. Does obedience require doing what one is literally told to do or doing what one's officer wants done? Peirce wrote, 'The object the command represents is the will of the officer that . . . ' (5.473). The words, 'will . . . that' indicate that by 'the will of the officer' Peirce meant a type of action desired, and not a psychological state of the officer or his act of volition. If that is correct, and if the officer misspeaks or if what he wants done is more specific than he can say, then there is a distinction between the immediate energetic interpretant of his command and its final energetic interpretant, granted a purpose to obey. What fulfills the immediate object does not necessarily fulfill the dynamic object. But this distinction, between immediate and dynamic objects, can be made, either intellectually or energetically, only on the basis of collateral observation.

The aim of military training is precisely to reduce the need for collateral observation and deliberation to a minimum, that is, to reduce the distinction between immediate and final interpretants to a minimum. On the parade ground, that ideal is very nearly achieved. For example, there is no difference between what the words 'Ground arms!' mean and the way in which the officer desires that action to be carried out: the soldiers' training makes those words to mean a precise motion, a precise degree of snap. And, on the parade ground, to second-guess the command would be wrong, even were the commander in some improbably distracted state and not saying what he meant to say. The commanded's purpose in that case is to do exactly as told; hence, the dynamic object is the type of act specified and not necessarily the type willed by the officer, pace Peirce. In a backhanded way, this testifies to the reality of the semeiotic distinctions we have been drawing, as it is their absence that is so marked a feature of a well-trained platoon on parade.

In battle, those same distinctions come back into play, though in ways problematic for the commanded. Suppose that an officer shouts, 'Charge up the left side!', there being, at the top of the hill, an enemy machine-gun nest that must be taken. In this context, the officer means more than

he can say. The soldiers should understand, by collateral observation, that they are to take advantage of the cover the left slope provides and thus run crouching up the ravine. And in this case their immediate purpose is to do as they are desired to do by their officer; it is not simply to do what he expressly says to do. Thus, not every actualization of the command's immediate interpretant would realize a final interpretant, to fulfill its dynamic object. The immediate object of the command – what the words expressly say to do – embraces but is less specific than the dynamic object – the exact will of the commander.

If, in the heat of battle, the officer, rattled, says 'left' when he means 'right', then the immediate object of the command is at variance with its dynamic object. That distinction can be made from collateral evidence later, in the officer's court-martial. If collateral observation at the time enables his troops to make the distinction, but without leisure to question their officer, they will, with some trepidation, do as they were meant to do, thus correcting the sign – the words of the command – even in interpreting it. In this case, obedience ceases to be automatic and thought is necessary: there are logical interpretants, taking up the brief period of the soldiers' hesitation, intervening between the sign and its energetic interpretant. Of course, the soldiers cannot but fear that they will be charged with not doing as ordered: obedience has become ambiguous. This, notice, is quite different from the kind of case that also arises, where the officer meant what he said but appears nonetheless to have commanded the wrong thing (relative to a purpose shared by officer and men); but that is not a question about how the command should be interpreted; it is a question of whether that sign should be ignored altogether.

Not all energetic interpretants interpret commands. Recall our zoölogical examples, of bears digging for grubs, rabbits evading foxes, and the like. In those cases, the conditions for success are not a type, either willed or expressed. The immediate object is that grubs or fox are present, and the dynamic object is the actual situation. The dynamic object determines causally, rather than by definition, whether the energetic interpretant succeeds in its aim; more precisely, it determines the interpretant's appropriateness in those circumstances to its aim.

We may thus draw a tentative conclusion, that energetic interpretants fall into the same teleological and realistic structure as do logical interpretants. There is a distinction between a sign's object (its dynamic object) and what the sign represents that object to be (its immediate object), and, thus, there are distinctions between a sign's immediate and

final interpretants and between the success or failure of any dynamic interpretant of it. To extend this analysis to statements of moral duty, one would have to address the question, among others, of where lies the locus of that authority that in the military case resides in the commanding officer. Is it God, the community, reason, or something else? If the extension can be made, then our preceding discussion, of the nuances of military obedience, will translate into a corresponding set of moral distinctions, doubtless with additional complications.

What of emotional interpretants? 'The performance of a piece of concerted music . . . conveys, and is intended to convey, the composer's musical ideas; but these usually consist merely in a series of feelings' (5.475). We may presume that Peirce was not expressing a naïve view. He must have meant specifically musical feelings. Musical ideas or feelings are not mere sound, and yet are nothing apart from sound; they may be described in emotive language ('sad' and the like) and yet are never adequately so described. Description – a logical interpretant, dependent on collateral observation by which the feelings heard are compared with human emotions conventionally designated – lacks the specificity, complexity, and nuance of the musical ideas described. As Mendelssohn said, 'It is not that music is too vague for words; it is too precise for words'. That is why the proper or complete interpretant of music is emotional, not logical. But that emotional interpretant is not one's ordinary feeling, for example, of sadness; it is the same complex of feeling as that embodied in the piece of music heard.

It follows that the music is an icon; it represents the qualities it embodies. That such feelings are first possessed by the composer does not mean that he is somehow 'expressing' his own emotions, referring to them, or telling us about them; they are feelings contemplated. The composer might feel delight and triumph in getting so much of sadness and despair into sound. Those feelings are reproduced in the audience, so far as the music, and not the sound merely, is heard. And they are reproduced in the same contemplative mode, that is, in heightened form and unattached to the audience's own actions and sufferings. We delight in contemplation, even of the sad. Aristotle said that man by nature desires to know, and there is no reason to suppose that such knowledge has always to be conceptual; witness his own account of art, in the *Poetics*.

Normally, if the music is at all interesting, one does not 'get it' all, at least on first hearing; that is, one's feelings fall short of those in the music itself. One can get so little of it that a logical interpretant of the music mistakes its ideas entirely. The same applies to other art forms: one

can mistake Van Gogh's violent colors as 'cheerful' or miss the darkness in Frost's poems. Misjudgment suggests that the judge's emotional interpretant is deficient, but a misjudgment can also be at variance with that interpretant. Is not one who thinks a Van Gogh cheerful in fact made to feel uneasy by it? He is judging conventionally, rather than aesthetically, on the basis of his feelings. Contrariwise, accurate judgment can focus feeling, enabling it to become heightened; hence the role of criticism. But a judgment in the first place interprets feeling and, in the last place, is no substitute for feeling.

The upshot of all of this is that, as with any pure icon (chapter 8, section 2), there can be no mistaking the dynamic object of a piece of music, once one has grasped its immediate object. For the two are the same, being distinguished only by the respect in which they are apprehended: they are feelings merely as heard or those same feelings in all their relationships (say, as studied in musicology). Any defect in one's interpretation of the music is, therefore, a failure to hear all that there is in it. The aesthetic purpose of listening is, so far, not fulfilled; when it is fulfilled, it is fulfilled in the listening itself.

Music is a limiting case of sign,[20] as is the pure icon in general. A pure icon does not signify anything that it does not contain. The feeling as contained in the sounds is the sign, in itself is the object, in the experience of the listener is the interpretant; the distinctions among these three are relational, not substantive. We should therefore expect the application of Peirce's semeiotic to music to be somewhat truncated, as it is. But the basic teleological structure is the same, as is the objectivity of the question of whether the (emotional) interpretant was adequate to its purpose. This testifies to the reality of what music represents, that there really are such feelings, as objective possibilities, independently of whether anyone has them.

And might not these same distinctions, and the same realism, apply to our moral emotions, so that shame, remorse, indignation, contempt, admiration, pride, and so on, may be said to be emotional interpretants in which objective moral qualities (their dynamic objects) are either apprehended or, as the case may be, mistaken? Of course, to apply Peirce's semeiotic to moral emotions, one must show how these are subject to correction in light of collateral experience. But are they not? A window is broken; an aunt appears; there am I, slingshot visibly in hand; her

[20] Yet there have been some interesting applications of Peirce's semeiotic to the analysis of music, preeminently, Cummings 2000, and more recently, Kruse 2005.

harangue is replete with factual assertions pertaining to a long chain of effects a broken window has, for example, on the family exchequer; these I am to interpret logically; but the tenor of her discourse, whether or not it includes such overtly emotive language as 'bad' and 'ashamed' (as in, 'Aren't you?'), is one that requires an emotional interpretant (remorse) in which this rupturing of the moral order – an objective fact – is apprehended; and perhaps an ensuing energetic interpretant of that emotion, as itself a sign of moral order ruptured, is expected, namely, a resolve to reform. And if I am innocent, her harangue will still have an emotional interpretant, but one in which the harangue is corrected: I will feel indignation at being falsely accused. This indignation registers a discrepancy between the dynamic object (my relation to the moral order) and her presentation of it, which is the immediate object of her vivid word portrait.

What Peirce's semeiotic adds to the familiar logic of moral discourse is the implication that it has a dynamic object, that it is not a merely conventional or subjective structure. By taking feelings to be at least sometimes interpretants of signs, and therefore subject to correction, it implies that feelings have objects (just as the signs they interpret do) and that those objects are of a nature (say, good or evil, right or wrong) that is independent of their being signified. In this respect, Peirce's semeiotic runs counter not only to other versions of semiotic theory, which tend to make even the natural sciences relative to social convention, but to the modern notion that any value – aesthetic, moral, political, or cognitive – is subjective. The usual argument, that these values must be subjective, because they are objects of emotion, not sensation, is confuted.

As we noted earlier (chapter 2, section 3), Peirce in his first published writings set out to attack modern philosophy at its root, by attacking Cartesianism. He proceeded to defend a medieval conception of realism (chapter 3) and to reverse modernity's rejection of teleological explanation (chapters 4 and 5). His semeiotic, I suggest, is a continuation of that same impulse, and has the import, among other imports, that emotion is cognitive (cf. Savan 1981). It is not cognitive in the sense that has sometimes been urged, that it is at bottom conceptual or a judgment; rather, it is cognitive merely as feeling. Mere feeling is cognitive so far as feelings occur within the teleological structure of semeiosis, as interpretants.

8

A Taxonomy of Signs

Peirce referred to certain types of sign – icon, index, symbol and term, proposition, argument – from early on, but he did not initiate a comprehensive semeiotic taxonomy until 1903. And he never completed it. Every formulation was tentative. Nevertheless, its main lines were clear from the start. What a sign is in itself is divided triadically, as monadic, dyadic, or triadic; so also, what a sign is in its relation to its object is divided triadically; and so also, what it is in its relation to its interpretant is thus divided. Any sign will belong to one each of each of those three triads. That yields twenty-seven combinations, though seventeen of them are ruled out on formal grounds. Thus, in 1903, ten classes of sign were recognized. That number increased when distinctions were made between immediate and dynamic objects and between immediate, dynamic, and final interpretants. In 1908, there were ten trichotomies and sixty-six classes of sign, and a potentiality for further subdivisions.

We begin with the three triadic divisions of 1903. The principles presupposed, especially that by which certain combinations are proscribed, shall be left unexamined until the next chapter. For it is the examples that motivate the principles. In the end, examples and principles are both justified, if at all, by their being shown to fit together, each with the other. In 1903, in the midst of his first efforts to form a semeiotic taxonomy, Peirce wrote:

> The principles and analogies of Phenomenology enable us to describe, in a distant way, what the divisions of triadic relations must be. But until we have met with their different kinds *a posteriori*, and have in that way been led to recognize their importance, the *a priori* descriptions mean little; – not nothing at all, but little. Even after we seem to identify the varieties called for *a priori* with varieties which

the experience of reflection leads us to think important, no slight labor is required to make sure that the divisions we have found *a posteriori* are precisely those that have been predicted *a priori*. In most cases, we find that they are not precisely identical, owing to the narrowness of our reflectional experience. It is only after much further arduous analysis that we are finally able to place in the system the conceptions to which experience has led us. (EP2:289)[1]

As we noted in chapter 3, Peirce used formal categories as a guide to phaneroscopic investigation, and the latter to give those categories empirical content. This is nowhere more evident than in his exploration of semeiotic phenomena, in which key terms were often used in a search for their meaning.

In the next chapter, we explore the more elaborate and less certain taxonomy of 1908. Despite the length of this discussion, it is not a complete review of all that Peirce said about classes of signs. It treats in some depth, and as systematically as the subject permits, his major statements. But there are variations, many not yet published, that I have not taken into account.

1. Qualisign, Sinsign, Legisign

In 1903, Peirce, writing of 'signs' narrowly as those interpreted by 'a cognition of the mind' (EP2:291), announced the three basic trichotomies:

> Signs are divisible by three trichotomies: first, according as the sign in itself is a mere quality, is an actual existent, or is a general law; secondly, according as the relation of the sign to its Object consists in the sign's having some character in itself, or in some existential relation to that Object, or in its relation to an Interpretant; thirdly, according as its Interpretant represents it as a sign of possibility, or as a sign of fact, or a sign of reason. (EP2:291)

As each sign must belong to one each of each of the three trichotomies, we confront an expository difficulty: to illustrate any division, we must cite signs that can be understood only as exemplifying other divisions as well, with which, therefore, some acquaintance must be presupposed. However, the second division is already familiar to us: it is of icons, indices, and symbols. And the third division is also familiar, at least as limited to language: linguistic signs that represent possibilities, facts, and

[1] For many years, the main published source for Peirce's classifications of signs has been 2.227–327. But, as that contains some inaccuracies corrected in EP2, and as EP2 contains other germane materials as well, all citations in this chapter are to the latter volume whenever possible.

reason are (or so Peirce intended) terms, sentences, and arguments, respectively.

The first trichotomy is based on what signs are 'in themselves'. What does that mean? To be interpreted or to be interpretable, a sign must first of all be something, hence, a quality, or an existent, or a law (a 1st, a 2nd, or a 3rd, in Peirce's phaneroscopy). Thus:

> According to the first division, a Sign may be termed a *Qualisign*, a *Sinsign*, or a *Legisign*.
>
> A *Qualisign* is a quality which is a sign. It cannot actually act as a sign until it is embodied; but the embodiment has nothing to do with its character as a sign.
>
> A *Sinsign* (where the syllable *sin* is taken as meaning 'being only once', as in *single*, *simple*, Latin *semel*, etc.) is an actual existent thing or event which is a sign. It can be so only through its qualities; so that it involves a qualisign, or rather, several qualisigns. But these qualisigns are of a peculiar kind and only form a sign through being actually embodied.
>
> A *Legisign* is a law that is a sign. This law is usually established by men. Every conventional sign is a legisign. It is not a single object, but a general type which, it has been agreed, shall be significant. Every legisign signifies through an instance of its application, which may be termed a *Replica* of it. Thus, the word 'the' will usually occur from fifteen to twenty-five times on a page. It is in all these occurrences one and the same word, the same legisign. Each single instance is a replica. The replica is a sinsign. Thus, every legisign requires sinsigns. But these are not ordinary sinsigns, such as are peculiar occurrences that are regarded as significant. Nor would the replica be significant if it were not for the law which renders it so. (EP2:291)

The distinction between legisign and replica is better known as the type/token distinction, that terminology also being due to Peirce (4.537).

The inclusion of qualisigns and sinsigns proves wrong the multitude of commentators who assert that on Peirce's view a sign is a 3rd. To the contrary, it is clear that a sign – that is, that which is a sign – can be of any category. The triadicity of signhood consists in the fact that something can be a sign only in triadic relation to two others, its object and its interpretant. *Being* a sign involves relations to others, but *that which* is in such relations may itself be of a lower adicity.

A qualisign is, for example, the color embodied in a cloth sample; in itself, that color is a mere possibility, its actually occurring in the sample being an addition to it; and what it represents is nothing other than itself, not as embodied but as a possibility that might be realized elsewhere or nowhere.

There are some difficulties in the preceding passage. Notice the flat-out contradiction between 'the embodiment [of a qualisign] has nothing

to do with its character as a sign' and 'these qualisigns [in a sinsign] are of a peculiar kind and only form a sign through being actually embodied'. The latter are so peculiar that they are not qualisigns at all, as that term was defined. Peirce ought to have said, rather, that a sinsign must have qualities but that its significance is not that of those qualities qua signs.

Legisigns present other problems. One is this: is a legisign any 'law that is a sign' or must it be 'established' as a type 'which, it has been agreed, shall be significant'? For, laws that are not established in order to signify are nonetheless significant. For example, a law of nature exemplifies lawfulness in general and a criminal code is evidence of a certain level of culture and social organization; but examples and evidence are two kinds of sign that Peirce distinguished as iconic and indexical, respectively. Now, as laws can neither be qualisigns nor sinsigns, there is no place in Peirce's taxonomy for such signs as these, unless we suppose that they are legisigns. But, then, not all legisigns are established in order to signify, since, presumably, laws of nature are not established in order to signify and criminal codes, though they are established to signify, are not established to signify a general level of culture. But we cannot blithely conclude that a legisign may be any law; for the distinction between laws that are and laws that are not established in order to signify is so important as to warrant explicit recognition in the taxonomy.

Usually, but not always, Peirce referred to legisigns in the narrower and more interesting sense of being laws established in order to signify. That unacknowledged inconsistency is proof that his taxonomy was in the process of being developed and that even in basic respects he had not thought it through. We in fact need to use the term in both senses, one to capture those signs that are 3rds and the other to distinguish those 3rds that were instituted in order to signify. But, for the most part, we use it in the latter, narrower sense. Therefore, let us introduce the term 'legisign(G)' to represent the general sense and use 'legisign' in the narrower sense only. Legisigns(G) may have instances, but these are not spoken of as replicas.

The essential feature of legisigns and their replicas, distinguishing them from all other signs, is that the purpose they have to signify constitutes their significance. Peirce did not put it quite that way, but spoke of legisigns as being laws, or as having laws, for the interpretation of their replicas; however, as those laws exist for that purpose, it comes to the same thing, and, by emphasizing purpose, we make it easier to avoid some mistakes commonly made by interpreters of Peirce's semeiotic. Signs of other categories may be used or produced purposefully in order to signify, but

their significance does not depend on that purpose. A color embodied in a patch of cloth is an icon of such color whether or not anyone uses the patch with the purpose of exhibiting that color. A portrait may be drawn with the purpose of representing the sitter, but it is a likeness (or not) independently of that purpose. Someone can build a fire to produce smoke in order to draw attention to himself (a hiker lost in the woods, for example), but the smoke rising at that location signifies fire in that location, regardless of any purpose to signify. If a code is used (say, a line of three fires, signaling distress), then that is a different matter. Then the smoke signifies more than fire, and that additional significance is apprehended only by one who knows the code and who supposes that that code is being employed in this instance with the purpose of signifying by its means. In short, the difference is between purpose exploiting forms of significance that exist independently of that or any purpose and purpose constituting significance.

Legisigns and their replicas are so important, and Peirce's commentators have so often confused other classes of sign with them, that extended commentary is necessary. One prevalent confusion is to suppose that causal laws are legisigns and that instances of causation are their replicas, as for example in the case of medical symptoms. The argument is that, to interpret the instance, one must know the law. But the difference is that the law does not exist and the instance does not occur in order to signify; they do not have that purpose. At the other extreme, it is often thought that all legisigns are conventional. But not all that exists for a purpose or is done for a purpose is created or is done with conscious intent (chapter 5), much less by express agreement with others. Peirce said only that 'every conventional sign is a legisign', not that every legisign is conventional. The two extremes meet in the absurd doctrine that laws of nature are social conventions (this is otherwise known as 'the social construction of reality').

Consider the mating display of the ruffed grouse: it consists of a puffed-out breast, booming sounds, and frenetic antics, not unlike the behavior of young males of another species. The instinct for such display exists in order that male grouse may elicit sexual response from female grouse; there would be no such instinct were there not a correlative instinct on the part of female grouse to respond accordingly; the two instincts evolved in tandem. Female response interprets the display as a sign of the male's readiness to mate; readiness to mate is the object signified. But there is no connection between that sign and its object except in the instincts that exist for the purpose of signifying and interpreting. Thus the

display, as a type of behavior, is a legisign. Its replicas have a correspondingly particular purpose, namely, to win the heart of a particular female. Such purpose requires no capacity to think; thinking would only complicate matters, as we know from our own experience. Replicas of legisigns are what St. Augustine named *signa data*, in contrast to *signa naturalia* (chapter 1, section 6). *Signa naturalia* are natural in the sense of occurring without a purpose to signify, but *signa data*, and the legisigns they replicate, are in some cases natural in the sense of not being man-made.

As Peirce noted, replicas are sinsigns, and legisigns cannot signify except through their being replicated. The replica must possess qualities that make it a replica of a given legisign, and in that respect it is an icon of that legisign. However, accidental conformity to a pattern is not replication; it is an icon of the type but not a replica of it. It is not a replica even if the type is a legisign. A pattern of bird droppings similar to the word 'repent' is not an instance of that word; much less does it mean that we should repent. To be a replica, an icon of a legisign must be produced in order to replicate that sign: it must occur for that purpose. It follows that a replica is related to the legisign replicated not only iconically but also indexically. For a purpose to replicate presupposes the existence of the legisign replicated and is therefore indicative of that legisign. Normally, the purpose is furthermore to signify by means of replication, but in some cases, such as exercises in elocution or penmanship, a legisign is replicated for other purposes.

It is relevant that the bird-droppings example is not only improbable but extremely improbable. Legisigns would be sources of confusion if they were likely to be iconized accidentally. If we had a language of sneezes, think of the misunderstandings that would arise, especially during hayfever season. To establish a legisign is thus to create a potentiality for exemplars that would otherwise be extremely improbable.

Exactness of replication is not important. Evidently, a legisign is a union of two laws, one defining replicas and the other determining how they are to be interpreted; but Peirce did not expressly distinguish the two (except rarely, e.g., at 2.292), perhaps because the law of interpretation predominates. No matter how clearly the parrot enunciates a string of swear words, it is not swearing, whereas no matter how much a grouchy and profane old man mumbles and slurs and mispronounces his words, he is swearing. Think, too, of the enormous range of shapes and sounds that we count as instances of the same word; it is the law of interpretation that groups these (cf. Saussure: chapter 1, section 5). The purpose to signify by means of a rule of interpretation is of the essence; formation

rules are a means merely and close conformity to them is only a practical help. And yet, we identify a legisign and its replicas with visual or auditory patterns and their instances, and not with the meanings that rules of interpretation give to them.

It follows that there can be no legisign before there is a law of its interpretation, whether conventional, customary, or instinctual. (As the grouse case shows, the laws of formation and interpretation do not have to be operative in the same individuals; but, if not, their separate purposes must subserve a joint purpose.) Once a legisign is established, then the rule of its interpretation can change. For that reason only, a legisign may have existed before its (current) rule of interpretation evolved.

As Peirce conceived of the matter, legisigns and their replicas are distinct signs. The immediate interpretant of the legisign is how, in general, its replicas are to be interpreted: it consists in a rule of interpretation, not of itself, but of its replicas. The immediate interpretant of a replica is different from that, as it depends on that rule's application to the particular case. This will become clearer in section 3, when we discuss iconic, indexical, and symbolic legisigns. For the present, note merely that while mating display in general means readiness to mate, this particular grouse's mating display means that *it* is ready to mate. The distinction is one that the grouse in question would particularly insist on.

The legisigns that most interest us are linguistic, and one test of Peirce's semeiotic is whether it can accommodate, and perhaps improve on, the insights of recent philosophy of language. Paul Grice famously distinguished hearer's meaning, word or sentence meaning, and utterer's meaning (Grice 1989, chs. 5–6, 14). In Peircean terms, the first is a dynamic interpretant of an utterance (or perhaps only of the sounds, in cases were the words are misheard). The second is the immediate interpretant of the legisigns replicated. The third is the effect in the hearer that the utterer intended to achieve by means of the hearer's recognizing that intention (that is the original formulation of the doctrine, p. 219). To recognize the speaker's intention, the hearer must recognize the word or sentence meaning of the utterance, that is, he must identify the legisigns replicated and know the rules for interpreting their replicas; in addition, the hearer must apply those rules to the given circumstances, thus actualizing the immediate interpretant of the utterance. But why is this a recognition of the speaker's intention? It is so for two reasons, inseparable from one another. First, the intention is recognizable because, and only because, such legisigns are not normally replicated except with an intention to signify by their means. Second, the replicas do not signify by means

of the legisigns replicated unless they were formed with that intention; hence, the immediate interpretant of the replicas includes recognition of their being formed with an intention to signify. Thus Grice's central insight was anticipated by Peirce; it is a special case of the general theory of legisigns – signs that signify through a purpose to signify.[2]

The institution of language presupposes a commonality of purpose at some level. That follows directly from its consisting of legisigns. Our purposes in speaking are related to the purpose for which the language exists, no less than the grouse's behavior is related to the purpose of its instincts so to behave; but, with us, the relationship is complicated. Why, for example, do we have in our language such a word as 'justice'? It must serve some social purpose that persons can demand justice, claim that injustice has been done, seek to restore justice, debate whether a law or policy is just, and so on. But the purposes for which individuals speak of justice very often are inimical to any social good the concept serves. In this connection, recall Kant's comment on lying. To lie is to use linguistic capacities that exist so that we may convey truth. But the effectiveness of those capacities either to convey truth or to deceive is eroded in proportion as they are used to deceive; for lies are often found out. Lying is self-destructive, honesty is not. There are philosophers who have suggested, with more or less reference to Peirce, a Kantian sort of ethics grounded, not, as with Kant, in pure reason, but in our being language-using, social animals, having a basic need to communicate (e.g., Jürgen Habermas in many writings but with special reference to Peirce in 1995 and Karl-Otto Apel 1980, ch. 7).

2. Icon, Index, Symbol

'According to the second trichotomy, a Sign may be termed an *Icon*, an *Index*, or a *Symbol*' (EP2:291). This division is based on the relation of a sign to its object – in later formulations, to its dynamic object (1904, 8.335; 1906, 4.536; 1908, 8.344, 368). That relation is the one we have called the sign's 'prior relation' to its object or the 'ground' of its significance.

It is essential not to confuse the ground of significance with significance itself. They are separate relations. Nelson Goodman pointed out that resemblance, unlike representation, is reflexive and symmetric

[2] It might seem appropriate to discuss J. L. Austin's speech act theory at this point, but that is deferred to the next chapter (sections 2 and 5), as comparison of it with Peirce's doctrine depends on his later, more extensive semeiotic taxonomy.

(1968, pp. 4–5). Some have taken this as a decisive refutation of Peirce's idea of an icon, the ground of which can be a resemblance. What it refutes, instead, is the identification of ground with significance.[3]

a. Icons

'An *Icon* is a sign which refers to the Object it denotes merely by virtue of characters of its own which it possesses, just the same, whether any such Object exists or not' (EP2:291)

Again, 'An *icon* is a sign which would possess the character which renders it significant, even though its object had no existence; such as a lead-pencil streak as representing a geometrical line' (2.304). That is to say, an icon's significance is grounded in its own qualities, and not in any relationship to another that requires the latter to exist. Nor does its significance depend on a rule of interpretation, although rules aid in the use of icons by directing attention to relevant features (see section 3). Anything that signifies on the ground of its own qualities alone is an icon.

Peirce most often referred to icons as 'likenesses', for example, in this from 1895: 'Most icons, if not all, are *likenesses* of their objects. A photograph is an icon' (EP2:13). Much later, in 1911, he wrote similarly: 'firstly, Likenesses, or, as I prefer to say, *Icons*, which serve to represent their objects only in so far as they resemble them in themselves . . . ' (EP2:460–1). However, a pure icon cannot signify an object by being similar to it. For how is that particular object picked out? Many things might be similar to the given sign, even in the same respect. As Peirce elsewhere said, particulars can only be signified indexically. A photograph is an effect of a physical process and thus it is an index of the subject photographed; only so is the image it contains an icon of that subject instead of a possibility merely. A painted portrait is an icon of the features

[3] There is more to Goodman's arguments: he also said that in important respects a picture resembles other pictures more than it resembles what it pictures. But all that proves is that pictures and resemblances are different sorts of sign; being a picture perhaps requires an index of the pictured and/or an index of an intention to have created a resemblance and/or satisfaction of some convention determining what sorts of object or kinds of resemblance can be a picture or a picturing. Finally, he said that one item can resemble another without representing it. But that is to suppose that representation involves a use or intention to represent. That is arguable, but it is not part of Peirce's stipulated definition of 'sign' (the justification of which, recall, is not its conformity to ordinary usage but is the fruitfulness of the science based on it; chapter 1, section 5). Whether X is interpreted, or presented in order to be interpreted, in light of its resemblance to Y, or supposed resemblance to Y, is additional to the fact of resemblance, if any; but it is the fact that justifies an interpretation based on resemblance, and that makes it a ground of significance in Peirce's sense (chapter 6, section 5).

depicted, whether or not anyone ever had them; only as accompanied by an indexical sign, such as an attached label, or in respect to the history of its production, is the painting a portrait of an actual person (whom it might fail to resemble).

Thus, in several places, Peirce distinguished a pure icon as having for its object only that 1stness that it itself is. In 1904: 'A pure icon is independent of any purpose. It serves as sign solely and simply by exhibiting the quality it serves to signify' (EP2:306); that is, the quality is exhibited independently of the purpose served thereby. And a year earlier: 'A pure icon can convey no positive or factual information; for it affords no assurance that there is any such thing in nature' (4.447).

When a pure icon is itself a qualisign or 1st, then its embodiment or occurrence involves 2ndness in addition. The color of a red fire engine is a qualisign, an icon of the qualitative possibility that it is. By embodying that qualisign, the fire engine is itself an icon or, as Peirce put it, it is 'iconic' (2.276): it is an iconic sinsign. In some passages of 1903 (EP2:273, 282), he used 'hypoicon' for an iconic sinsign. The prefix suggests that it is a substratum supporting an icon.

Some of Peirce's favorite examples (paintings, photographs) of hypoicons are arguably causes, not embodiments, of icons, as when a picture painted on a two-dimensional surface creates in the viewer an image of depth. Or perhaps there are icons, here, at two levels. First, some lines on the two-dimensional surface are like the visual experience of the three-dimensional world; second, the illusion of depth thus created is part of an image, contained not on the two-dimensional surface but in the viewer's experience (cf. Aldrich 1958), that is like, say, cows in a field, foreground, with a cottage in the distance, background. The first is an icon of depth, the second is an icon of a pastoral scene.

Verbal art also forces us to admit some complexity to our idea of iconicity; or, contrariwise, it is the concept of iconicity that helps us to recognize some of the complexity of art. Verbal art is the use of words, many of which are in Peirce's sense symbols, to create icons. A novel or dramatic production, and much poetry and song, is replete with assertions, but these are not to be taken as assertions, normally; rather, they serve to create a story or an image or other icon. Sometimes, of course, the mode is mixed: in didactic works, some statements are intended to be taken as such; in historical novels and confessional poetry there is indexical reference to past events.

But even when the mode is not mixed, our experience of the real world informs our experience of the fictional world, presented iconically

in art. In abstract painting, for example, our sense of weight, volume, up and down, informs our experience of the image; we feel Delaunay's disks rotating or rhythm in Mondrian. And, sometimes, even in nondidactic works, indexical reference creeps in and the icon is impure. The pleasure of viewing some of Escher's prints consists in their making the impossible look possible: they are lessons in illusion, proving the power of the eye. But that pleasure depends on an implicit comparison of the image to what is really possible; the real world is not forgotten. Then there is quotation of art in art, as in Charles Ives' music. In itself, this is an iconic relation between icons, but, as mediated through the audience's memory of the music quoted, the relation is indexical, and the surprise of recognition, or knowledge of the historical circumstances of the music quoted, is part of the pleasure proper to hearing the quotation.

Return, for a moment, from the complexities of art to the simplest icons, namely, iconic qualisigns. The idea that a qualisign signifies nothing but itself is not so silly as it might at first seem. As so often in semeiotic analysis, we have to make distinctions that are relational rather than substantive. The red color embodied can serve to focus attention on that color in itself, independently of its occurring anywhere. The color in itself is the object; as embodied, it is the sign.

Sinsigns can be iconic in other ways than by embodying or creating a qualisign. A red fire engine is a hypoicon of its red color, but it is itself an example, hence, an icon, of fire engines. Since fire engines are individual things, they cannot be exemplified (though they may be depicted) by mere images; if one points to a picture and says, 'That's an example of a fire engine', it is not the picture but the vehicle depicted that is the example. It is not as embodying a qualisign, but qua individual thing, that a fire engine is an icon of fire engines. In cases such as these, a sinsign is iconic in its own right and, hence, is not, in that respect, a hypoicon. So also a 3rd, say, a law, can be an icon qua 3rd, for example, as exemplifying 3rdness or some more specific form of 3rdness, such as a law of nature, an inverse square law, or a law of only approximate validity. Obviously, a 2nd or 3rd is any number of distinct icons; which of these signs, if any, we notice depends on our interests and susceptibilities.

But iconicity is a matter of 1stness; how can 2nds and 3rds be icons except in respect to embodied 1sts? Recall that in his phaneroscopy, Peirce spoke of the 1stness of 2ndness and the 1stness of 3rdness, the reference being not to contained qualities but to the possibilities realized

in the 2nd qua 2nd or 3rd qua 3rd (chapter 3, section 8). Thus, in 1903, he wrote,

> An *Icon* is a Representamen whose Representative Quality is a Firstness of it as a First. That is, a quality that it has *qua* thing renders it fit to be a Representamen. Thus, *anything* is fit to be a substitute for anything that it is like. (EP2:273; I have emphasized 'anything' and deleted Peirce's emphasis on 'substitute')

Thus, icons are found not only among qualisigns but also among sinsigns and legisigns(G).

We should note, however, that, like likenesses, samples and examples are not pure icons. The sample is of a more inclusive quantity, existing or potential, that will vary from the sample in various respects, and an example is of a type that comprises a continuum of possible variations. Even a type of quality or 1stness is a 3rd, not a 1st. But in purely iconic signification nothing is signified but an exact 1stness that the sign itself is or realizes. Even a type is a pure icon of that very type alone and not of types in general. Samples and examples are icons, but only with the addition of some commentary, expressed or understood ('Tigers look *like* that', 'Red is any color *close* in shade to this'); or there must be a habit or instinct to generalize from the instance, by which a general concept is formed. Such commentary and such conception is symbolic, not iconic; such a generalizing tendency is a power to symbolize, even to form new symbols.

But surely the main use of icons is as samples, examples, or likenesses, to each of which uses the iconic component is essential. We may therefore speak of likenesses, examples, and samples as icons, if we remember that such icons are impure, having either an indexical or a symbolic component. We should note, too, that the fact that the most salient icons are impure does not entail that there can be no pure icons, icons bereft of symbolic commentary or indexical addition. Indeed, the icon with additions is, in itself, an icon without additions.

One last note on icons: in an interesting passage that I quote without comment, Peirce wrote,

> Hypoicons may roughly [be] divided according to the mode of Firstness which they partake. Those that partake the simple qualities, are *images*; those which represent the relations, mainly dyadic, or so regarded, of the parts of one thing by analogous relations in their own parts, are *diagrams*; those which represent the representative character of a representamen by representing a parallelism in something else, are *metaphors*. (EP2:274)

b. Indices

Having said so much about indices in earlier chapters, we may now be brief. Peirce's 1903 characterization of indices – 'An *Index* is a sign which refers to the Object that it denotes by virtue of being really affected by that Object' (EP2:291) – appears to be far too narrow. It reads as if all indices were effects, the objects signified being their causes. Among much else, that would exclude the paradigmatic instance of an index, which is pointing. Other passages, including many written at about the same time, are broader. For example, in 1901, Peirce defined the index as

> A sign . . . which refers to its object not so much because of any similarity . . . nor [by association] . . . as because it is in dynamical (including spatial) connection both with the individual object, on the one hand, and with the senses or the memory of the person for whom it serves as a sign, on the other. (2.305)

A spatial connection must be between two actualities, existing or occurring, but does not require one to be a cause of the other. As it is unlikely that Peirce narrowed the conception in 1903, we must read 'really affected' broadly, as meaning that the dynamic relation of index to object depends on the existence of the latter. As he wrote in 1902, 'An *index* is a sign which would, at once, lose the character which makes it a sign if its object were removed' (2.304).[4]

Sometimes, Peirce spoke of an index as signifying via the compulsive effect of its objects on it, and other times as signifying via its compulsive effect on an interpreter, as a poke in the back compels attention. Indices, he said, 'direct the attention to their objects by blind compulsion' (2.306). Perhaps there is a compulsiveness from the object through the index to the interpreter, but that compulsiveness must be in some cases extremely attenuated. I can choose not to look where you are pointing, just as you can choose not to point at the object that caught your eye.

Perhaps we are so organized that an extended index finger exerts a compulsive affect on our attention. But even if it does not, the finger defines a line in space. The line formed by the finger itself is but an icon of spatial direction; but the line as existing here, extending from this point where the pointer is standing, makes it an index of a specific direction. If instinct, custom, or verbal instruction directs attention along the line defined, that line must nevertheless have been defined. That direction

[4] It does not follow that indices provide an infallible foundation for inquiry. For one might be mistaken about something's being an index or about what it is an index of. Attention directed to an object is not yet a judgment as to its nature.

is the object of the finger qua index – qua pure index. But in this case, notice, the object is not an existing thing, except insofar as space is itself an existent, constituted by the orderly relations of 2ndness among existing things. Anything else – any object 'pointed at' – must involve additional signs combined with the pure index to form a compound sign. I am pointing at a dog only if I am thinking 'dog' as I point (otherwise I might be pointing to a direction in which we should walk, or to the color of the dog's coat); and I succeed in pointing that dog out to you only if either I say, 'Dog!' or you are by some other means able to guess my intention or you share my disposition to notice the dog rather than, say, its shadow.

In the same place where he distinguished hypoicons, Peirce wrote of 'subindices' or 'hyposemes' (EP2:274, 286). As a hypoicon is an iconic sinsign, so also a subindex or hyposeme is an indexical legisign, such as a 'proper name, personal, demonstrative, or relative pronoun, or a letter attached to a diagram . . . but none of these is an Index, since it is not an individual'. The idea seems to be that something of a higher category may support something of lower category and thus function like the latter. However, in later writings he did not deny that sinsigns can be icons and that legisigns can be icons or indices. A 3rd can be an index, as it incorporates actuality. For example, a law of primogeniture is evidence of a feudal order, present or past, hence of a certain historical development of society; it is an effect of causes of which it is an index. But those are legisigns(G); subindices or hyposemes are indexical legisigns (non-G), of which, more later (section 3).

c. Symbols

In 1903, Peirce offered two definitions of the symbol, between which there appears to be a discrepancy. One is,

> A *Symbol* is a Representamen whose Representative character consists precisely in its being a rule that will determine its Interpretant. (EP2:274)

The other is,

> A *Symbol* is a sign which refers to the Object that it denotes by virtue of a law, usually an association of general ideas, which operates to cause the Symbol to be interpreted as referring to that Object. (EP2:292)

In the first, a symbol is said to be a rule of interpretation, while in the second (in which, by the way, we may notice that the terms 'refer' and 'denote' are used loosely for any relation of sign to object), the symbol

is not said to be such a rule. And, indeed, if the symbol *is* the rule of interpretation, then what is it that is being interpreted? One might answer, it is the symbol's replicas that are being interpreted. But those replicas cannot be replicas of a rule of interpretation (that is unintelligible). They must, instead, be replicas of some visual or auditory type, with which the given rule of interpretation is associated. The 'association of general ideas' that Peirce mentioned must be an association of the type replicated with the type denoted, for example, an association of the word 'red' with the color red. It follows that the sentence first quoted errs in being excessively concise. It follows, also, that symbols must be legisigns, that is, replicable types; they are laws that define what may count as their replicas, with which are associated other laws that determine how those replicas are to be interpreted. The second passage continues: 'It is thus itself a general type or law, that is, is a legisign'.[5]

A symbol, then, is a sign of that object that is assigned to it by a rule of interpretation. (A rule of interpretation is not itself an interpretant: a confusion of which some of Peirce's expositors have been guilty.)

If symbols are legisigns, then qualisigns and sinsigns cannot be symbols. It follows that replicas of symbols, being sinsigns, are not symbols ('it is inaccurate to call a replica of a legisign a symbol' (8.335)). A replica is related to its legisign iconically in the sense of exemplifying it; in addition, Peirce once suggested that the replica of a symbol is an index of that symbol, 'in that the image it suggests to the mind acts on a Symbol already in that mind' (EP2:295). That suggests that what the replica signifies is the symbol, and certainly replicas can be formed in order to bring the symbol to mind; perhaps every replica must do so if it is to serve its purpose. Yet, in the majority of cases, the purpose of replication is to signify what the symbol signifies. More elegantly and, I think, more accurately, we might say that it is the symbol itself that signifies, through its replica or in being replicated. Peirce also sometimes used that language, as in the long passage quoted in section 1: 'Every legisign signifies through an instance of its application, which may be termed a *Replica* of it.'

[5] In 1904, there was still some ambiguity about whether a symbol is distinct from the rule by which it (or its replicas) is to be interpreted: 'if two symbols are used, without regard to any differences between them, they are replicas of the same symbol' (EP2:317). The examples Peirce gave are 'he' and 'him', which differ, he said, merely in grammar, and 'money' and 'spondesime', which differ rhetorically (the latter was a slang term of the day). The symbols are two in design, one in meaning, and Peirce's language is wobbly, as symbols qua design are here spoken of as being themselves replicas of symbols qua meanings, contrary to assertions elsewhere that replicas are sinsigns.

Because the replica is a sinsign, it is also related to particulars in the context of its occurrence, and of these it is as many indices. Depending on the grammatical structure in which it occurs and other elements of context, a replica of a symbol may do any of various jobs. It may do no more than call the type symbolized to mind; for example, on hearing the word 'umlaut' one may think of umlauts in general. But in other cases, the replica may be interpreted as an index of particular instances of the type which the symbol replicated represents, as in, 'That's an umlaut', or to a negative instance, as in, 'That vowel needs an umlaut'. The latter combine indexical reference by particular to particular with expression of a general type, effectively characterizing the particular referred to as being of, or as not being of, that type. A fuller account of this depends on the third trichotomy, examined in section 5.

All of which is by way of saying that symbols themselves can only signify types, not individuals. 'Not only is it general in itself, but the Object to which it refers is of a general nature' (EP2:292). Conversely, replicas can signify types only via the symbols they exemplify. A sinsign, merely as an individual thing, cannot directly signify anything general.

3. Iconic, Indexical, and Symbolic Legisigns

It is often supposed that every legisign is a symbol. For a legisign is a type the replicas of which are interpretable by a general rule, and a symbol is a sign that is related to its object by a general rule of interpretation. However, there is a difference between the general case of being interpretable by a rule and the special case of having an object assigned by that rule. If the rule is to attend to a specified iconic or indexical aspect of the replica, then it does not by itself determine the replica's object. One must attend to the specified aspect to discover what the object is. Thus, 'the demonstrative pronoun "that" is a Legisign, being a general type; but it is not a Symbol, since it does not signify a general concept. Its Replica draws attention to a single Object' (EP2:295). And, in fact, different replicas of that same legisign will signify different objects. In each case, one has to notice what the speaker is looking at or pointing to, when and where he utters 'that', and/or to other words that he utters at about the same time (as in 'that dog'), and/or to what else is present that might satisfy the description appended to 'that' (a lone dog).[6] This

[6] Peirce's language suggests that 'that' as a relative pronoun is a different legisign; but we might speak, instead, of demonstrative and relative replications of the same legisign. The two uses are distinguished by context, primarily the grammatical context formed by

is a topic now much discussed, independently of Peirce; yet it bears the Peircean name 'indexicals'.

Thus there are iconic and indexical legisigns (EP2:294). Pronouns are indexical legisigns. Some unite symbolic meaning with indexical reference. 'I' and 'you' normally imply persons speaking or spoken to; 'she' normally means female; 'he', as demonstrative and as relative to a singular referring term but not as relative to a general term, normally means male; 'it' and 'that' come closer to being purely indexical.

What of iconic legisigns? Peirce provided no examples, but in a 1904 letter to Lady Welby, he mentioned an 'individual diagram' as an example of an iconic sinsign (8.335); presumably that is why the editors of the *Collected Papers* suggested, as an example of iconic legisigns, a diagram 'apart from its factual individuality' (2.258). But that gives iconic legisigns little real work to do. I propose this variant: an individual diagram, say, of an isosceles right triangle, does not replicate a corresponding legisign, of isosceles right triangle in general, but replicates a number of legisigns, of line, angle, and so on, defined by the rules of geometrical diagramatization. Those rules tell us to attend to length but not to width or color of line, to the angles formed, and so on, but not to minor wiggles. Thus Peirce in 1903: 'an iconic legisign is any general law or type, in so far as it requires each instance of it to embody a definite quality which renders it fit to call up in the mind the idea of a like Object' (EP2:294). I would add only that the requirement must be learned and that it specifies the sort of quality to look for; thus the student learns in one course about geometrical diagrams, in another about bar graphs, and so on.

A geometrical diagram may be used as an example of other features of which it is iconic, such as its color or its having been done sloppily, but, as those features are not of a type specified in the rules, special effort must be made to draw attention to them. The function of iconic and indexical legisigns is to facilitate communication by certain kinds of icon or index. They do this by causing attention to be drawn to certain aspects of their replicas automatically, to the exclusion of all other aspects. Thus, an intention to signify in exactly that way is presumed of anyone who replicates such a legisign.

There is a problem with the taxonomic system at this point. Peirce wrote as if legisigns were signs in themselves, and not merely types of

replicas of other legisigns. Relative pronouns are also indexical, as they draw attention to accompanying nouns and refer via that relation. As governed by existential or universal quantification, the nature of that reference becomes complicated. For Peirce's account of it, anticipating Hintikka's game-theoretical semantics, see Hilpinen 1983.

sign. But iconic and indexical legisigns appear to signify nothing. More precisely, there is nothing that they exist to signify. Like any legisign(G) they may, of course, be icons or indices in various unintended respects (geometrical diagrams are evidence of a certain level of culture). If, in some desperation, we say that 'it' in general means an individual and 'they' in general means plural individuals, then we are turning indexical legisigns into symbols. But even 'you', though it has a symbolic component that limits how it may properly be replicated, is not itself a symbol, for example, for person addressed. 'Person addressed' is the symbol for that, and 'individual(s)' is the symbol for what 'it' and 'they' can be used to designate.

Earlier, we encountered the converse problem, that in signifying types, it seems that it is the symbol replicated that signifies, and not the replica itself. With iconic and indexical legisigns, it appears to be the replicas that signify and not the legisign replicated. This points toward a more elegant account of legisigns: one in which the ancient problem of the One and the Many is resolved by not so drastically dividing the two (Plato's *chorismos* problem). Instead of two layers of significance, we have, in each case, one significance differently borne by legisigns and their replicas. General meaning is borne by particular signs in virtue of their general type, and particular reference is carried out by general types of sign in virtue of their particular deployment.

This alternative is suggested by some of Peirce's formulations (see above). It may seem to wreak havoc with his taxonomic system. For it means that indexical and iconic legisigns are not really signs, except in the sense of being legisigns(G). However, if, in that system, we substitute legisigns(G) for legisigns, then the slots provided for laws or types that are iconic or indexical will not be empty. And symbolic legisigns(G) must all be legisigns.

What of proper names, and also names of biological taxa (those taxa being individuals, not universals; chapter 5, note 6)? Unlike pronouns, each occurrence of such a name – say, 'Napoleon Bonaparte' or '*Homo sapiens*' – names the same thing. Of course, some names are shared by different individuals, but that is just an instance of ambiguity, such as affects general terms also. Thus, in the case of proper names as in the case of symbols, the primary bearer of significance is the legisign, not the replica. It is the legisign that is the name of such-and-so individual.

It does not follow that proper names are symbols; for they do not name their objects via a general rule. Peirce denied that individuals can be signified by any sign general in meaning (chapter 2, section 7). All instances

of a name that name the same individual or taxon are related by historical succession, hence, indexically, to original instances, as in baptismal ceremonies or introductions, that signify that individual or individuals of that taxon indexically. Although it is the legisign that is the name, it is through certain of its replicas that it came to name what it names. Rules are involved, to the effect that names signify in this way; but those rules do not specify the initial replications nor the historical chain linking later replicas to them. Thus, in 1904, Peirce wrote, 'I define an Index as a sign determined by its dynamic object by virtue of being in a real relation to it. Such is a Proper Name (a legisign)' (8.335). The parenthetical phrase is Peirce's: hence, he classed proper names as indexical legisigns (also at 8.341). The 'real relation' was explicated a year earlier: 'A proper name, when one meets with it for the first time, is existentially connected with some percept or other equivalent individual knowledge of the individual it names' (EP2:286).[7] Thus Peirce anticipated a doctrine now associated with Saul Kripke; but more on that later (chapter 10, section 4).

Proper names, then, appear to be indexical legisigns in the sense that they actually are themselves signs that exist for the purpose of signifying the individuals they name. Pronouns are indexical legisigns in a different way, by existing to be used to refer to diverse individuals, depending on replicatory context.

4. A Common Error Corrected

Many commentators have maintained that on Peirce's view the only true or genuine or complete sign is the symbol, hence, that icons and indices are not truly or genuinely complete signs. They do so on the seemingly unimpeachable ground that Peirce himself said exactly that. And so he did, but the question is, what did he mean? Others have maintained related theses, that icons and indices are but special kinds of symbol or

[7] Peirce proceeded to complicate matters: 'It is *then*, and then only, a genuine Index. The next time one meets with it, one regards it as an Icon of that Index. The habitual acquaintance with it having been acquired, it becomes a Symbol whose Interpretant represents it as an Icon of an Index of the Individual named'. As the 'it', referring to the proper name, of this account has many instantiations, it must be the legisign. So, he said that one and the same name qua legisign progresses from index to icon to symbol. It seems to me that this is an unnecessary descent into psychological description and that, as the name is still related to its object indexically, through initial replications, it is not properly classed as a symbol. Cf. a statement early in 1904: 'A *Proper Name*, also, which denotes a single individual . . . differs from an index only in that it is a conventional sign' (EP2:307).

must have a symbolic component, or that being iconic, indexical, symbolic are distinguishable but inseparable aspects of a complex whole, which is all three. Our task in the present section is to slay this hydra-headed monster.

No less a figure than Roman Jakobson, the first linguist to have discovered Peirce's relevance to that science, wrote:

> One of the most important features of Peirce's semiotic classification into icon, index, and symbol is his shrewd recognition that the difference among the three basic classes of sign is merely a difference in relative hierarchy. . . . merely the prominence of one of these factors over the others. (1990, p. 411)

Also, 'These divisions are merely three poles, all of which can coëxist within the same sign' (Jakobson 1985, p. 253). Such a view is suggested by Peirce himself: 'the most perfect of signs are those in which the iconic, indicative, and symbolic characters are blended as equally as possible' (4.448); again, 'An *icon* can only be a fragment of a completer sign' (EP2:306); and 'it would be difficult, if not impossible, to instance an absolutely pure index, or to find any sign absolutely devoid of indexical quality' (2.306). But there is also much, not only in what Peirce said, but in the structure of his semeiotic, that points to an opposite conclusion.

We have already determined that something may be more than one sign. But what of a word that signifies the same object in more than one way, as in onomatopoeia? 'Hoot' imitates what it names and thus is both icon and symbol. That, however, does not make it an iconic symbol or symbolic icon, as the two functions are mutually independent. One can name without imitating, and conversely. Were we sufficiently perverse, we could name hisses 'hoots' and hoots 'hisses' and thus explain that whatever sounds like 'hoot' is a hiss, and conversely. 'Hoot', then, would be two signs, iconically of one thing and symbolically of another. So also in its present and more natural employment, where it is icon and symbol of the same thing: it is one word but two signs, at least as signs are counted in Peirce's system.

The usefulness of onomatopoeia produces a constraint: it would be foolish to use 'hiss' to mean hoot and vice-versa. But that constraint is not built into the significance of such a word, either as symbol or as icon. A more interesting type of constraint is illustrated by personal pronouns, which combine indexical use with symbolic meaning, as noted above. The symbolic meaning constrains the indexical usage. As, for example, 'she' implies being female, it is misleading or wrong, normally, to refer to a male as 'she'. But here, too, the symbolic constraint is not essential

to the indexical use, as is clear when one uses 'she' successfully to refer to an individual wrongly supposed to be female or to things, for example, ships, no one supposes are female.

Another sort of case has to considered, that of a compound sign. For example, a sentence combines subject and predicate terms in order to express something (a truth or a falsehood) that can be represented by neither term alone. The components may be different types of sign, say, a subject that is an index and a predicate that is a symbol, and this may also give rise to the idea of 'blending'. However, Peirce maintained that the compound is a sign distinct from its component signs; it signifies how the components are to be combined – or, more precisely, it signifies how the whole is to be interpreted, complexly (see section 4). And that sign is itself a sign of just one type (Peirce named it a 'dicisign'). In contrast to Peirce's analysis, talk of blending seems to me to be both lazy and unilluminating; signs do not blend.

The least telling of Jakobson's arguments is that pronouns must be indexical symbols (1990, pp. 154, 388), as that derives from the error, already exposed, of supposing that all legisigns are symbols. Indeed, iconic, indexical, and symbolic grounds of significance are categorically different and cannot be blended. Talk of being iconically indexical or indexically symbolic, and so on, is a contradiction in terms.

What, then, of Peirce's own statements, quoted above? It is not clear from the context (4.448) what Peirce had in mind by a 'perfect' sign. Perhaps it bears on what, in the fullest sense, the function of a sign is; but a passage three years later identifies a perfect sign as one that 'involves the present existence of no other sign except such as are ingredients of itself' (EP2:545n25) – a rather special sort of perfection. In either case, this does not entail that 'imperfect' signs are not signs. In fact, the sentence partially quoted from 4.448 begins by asserting that signs may exercise any one of the three functions, iconic, indexical, or symbolic, exclusively or any two exclusively. Similarly for the notion of a complete sign, where the idea of completeness appears to refer to all that we use signs for, which Peirce in that place (EP2:306) assumed was assertion (see EP2:307). But signs that do not assert still signify. Even if their use is subordinate to a more complex process, that function is independent of the rest, and only so may it contribute what it does to the whole.

The remaining passage is more difficult to explain. We have ourselves seen that indices are normally impure; yet the impurities are additions, the pure index being an identifiable component of the complex sign. So that is all right. But what of the claim that all signs have an indexical

component? Necessarily, symbols have an indexical component, as 3rds incorporate 2nds: whether our talk is of electrons or of griffins or of logical paradoxes, we know what world, real, mythical, or ideal, is being referred to. (So, also, indices have iconic components. The extended index finger or the weathervane incorporates, in its own linearity, an icon of directionality; and only so is it able, indexically, by virtue of its existing location and occurring orientation, to indicate a particular direction.) But what of icons, especially icons of pure qualitative possibilities? Must not such an icon be devoid of indexicality? It seems to me that Peirce must have been thinking of the phaneroscopic doctrine, that while 2nds do not incorporate 3rds nor 1sts incorporate 2nds or 3rds, nevertheless, we never encounter a 1st except as a component of a 2nd, nor a 2nd except as a component of a 3rd. An icon, therefore, is always experienced as part of a larger whole. And thus the color, the odor, the shape we experience, we refer to that world of physical possibility from whence, as if trailing clouds of glory, it came. But it is we who do the referring; the qualitative icon does not itself bear that significance. Any clouds it trails are external to it.

In general, the pure index only directs attention, as in causing one's head to turn. Paraphrasing Kant's slogan, it is direction without content. Pure icons and pure symbols have content without direction. Content and direction together signify particulars, form assertions, and otherwise make knowledge possible. Therein we find the more complete or more perfect sign. But to assume that the whole alone is real is an error Peirce rejected as nominalistic (1.422; see above, chapter 3, section 4). Showing and saying, pointing and showing, saying and pointing are compound signs; as compounds, each of their components must be signs. If they were not, the whole would not signify what it does.

Umberto Eco's approach to the question differs from Jakobson's, as it derives from the Saussurian view, which Jakobson had denied, that all significance is determined by an arbitrary code, socially established. And since Peirce also did not think that all significance is arbitrary, Eco fails as an expounder of Peirce's semeiotic. However, his examples are worth noticing, as they suggest that what Peirce had thought of as not arbitrary really is arbitrary. Eco argues that pictorial art and medical symptoms are symbolic and, moreover, conventional in meaning. Let us see whether he is right.

Now, it has been shown that pictorial representation depends on rules of representation (Gombrich 1960). In medieval art, to cite an example Eco employs, a difference in size of two human figures often represents

a difference in their social status. If perspective is not an element of the style employed, then a difference in size of two figures could not rightly be interpreted as a difference in their closeness to the viewer. And so on. One has to know the style in order to see the picture properly.

All of that is interesting, but it does not justify Eco's conclusion that iconicity is 'a matter of cultural convention' or 'a codified system of expectations' (1976, p. 204). A code does not make a patch of cloth to have the color it has. Two lines, straight or curved, uniformly converging, look the way railroad tracks look as they recede into the distance. They look like that whether the painting they occur in deploys perspective or not. Only, we are unlikely to notice the resemblance, and are sure to dismiss it if we do notice it, if (a) we have not become accustomed to perspective in art or (b) we are not looking at a painting that exploits perspective. A stylistic rule (it is surely not a code) licenses a certain kind of painting and a corresponding kind of looking, promoting each. But that is all that it does. It does not make dissimilars similar; it does not create similarities.

Notice, by the way, that size is not visually similar to social status. Ergo, the painted figure's relative largeness is not a visual icon of social status. Rather, it is a visual metaphor: it is a visual icon of physical largeness, and physical largeness, in turn, is a nonvisual icon of social status because it shares with social status the property of having power over others. Relative size of human figures in medieval art is iconic qua metaphorical and therefore is iconic in a more complex way than is relative size of objects in perspectival painting. That is so, even if painting perspectivally is a more complex way of painting, more difficult to master, than is the nonperspectival style of medieval art.

Peirce's concept of an icon is subject to misunderstanding because that word is used today for any visual image, especially if highly conventionalized, that has a readily recognizable reference. That usage owes more to 'iconology' in art history and the use of 'ikon' in the Eastern Orthodox Church than it does to the Greek root of the word to which Peirce appealed. A conventionalized image has a reference that is essentially symbolic. Trousered and skirted images mark men's and women's lavatories, respectively, even in places where women are more often trousered than skirted: what matters here is not similarity – trousered women do not head for the men's room – but a rule of interpretation merely suggested by the fact that men are rarely skirted these days (see Johansen's amusing discussion, 1993, pp. 116–24). Since the images mean what they do because of a conventional rule of interpretation, they are visual symbols, not icons in Peirce's sense. More confusing still is the new journalistic

practice of calling any readily recognized person, building, and so on, an 'icon'. But that is so inexcusable and bereft of definite meaning as not to deserve further mention.

Eco also argues that medical symptoms are conventional: a symptom is a sign, he says, only 'when this association [of it with that of which it is symptomatic] is culturally recognized and systematically coded', for example, in medical textbooks, establishing a 'semiotic convention' (1976, p. 17). That is to force natural signs into a linguistic mold, assuming that language is itself conventional. But the effort is doubly mistaken. First, to have discovered a natural law or association is quite different from establishing a convention, even when conventions are devised for its representation and when that representation makes the discovery available for use in some standard practice governed by conventions. Second, laws and conventions are general, whereas Bobby's fever is a particular occurrence and is connected dyadically to its cause. Cause and effect instantiate a general law but nonetheless are dyadically, existentially connected. The symptom in a particular case is therefore an index of the disease that is its cause. To be sure, knowledge of the causal relation in general is required to identify the individual instance as being a symptom; but that neither makes the individual occurrence a symbol nor the causal law a convention.[8]

Let us turn, then, to Peirce's statements that symbols alone are 'genuine' signs, icons and indices being 'degenerate' cases of sign (see, especially, EP2:306–7). This of course follows from the phaneroscopic doctrine that 2ndness has one degree of degeneracy and 3rdness two degrees (chapter 3, section 10). 2ndness is degenerate when one of its relata is a 1st, and 3rdness is degenerate in the first degree when one of its relata is a 2nd, in the second degree when one of its relata is a 1st. Peirce held that a symbol, having a general conception as interpretant and general type as object, is a genuine sign, while indices and icons are degenerate in first or second degree, respectively.

Such claims have been taken up with enthusiasm by many commentators, who perhaps seek confirmation of a Saussurian or intellectualist bias toward symbolism over other forms of representation. But they overlook

[8] In a letter of 1904 to Lady Welby, Peirce said, 'The symptom [of a disease] is itself a legisign, a general type of a definite character. The occurrence in a particular case is a sinsign' (8.335). Unless Peirce is here confused, he meant 'legisign' in the more general sense, legisign(G); for a symptom is not established in order to be a sign. But even if he here conflated a symptom with its representation, he did not say that a symptom is conventional. And he did clearly distinguish the instance from the law instantiated.

the converse principle, mentioned above, that the whole is not all that is real. In any case, Peirce, in this context, used the genuine/degenerate dichotomy in a special sense, derived from mathematical usage, in which, to cite an example he himself cited (EP2:306, 544n24), some conic sections – a point, a straight line, an angle between straight lines – are degenerate because they can be defined more simply, that is, without reference to cones or anything equally complex. A degenerate case is a limiting case of a complex type: one that is a case also of a simpler type. But, as the degenerate case is nonetheless a case, so also degenerate signs are nonetheless signs. Thus, no particular value is connoted by the word 'genuine' in this usage: 'Of signs there are two different degenerate forms. But though I give them this disparaging name, they are of the greatest utility, and serve purposes that genuine signs could not' (EP2:306).

It should be noted that Peirce sometimes used 'genuine' in a different sense, of being irreducible to a logical complex of relations of lower order (Kruse 1991), and in that sense icons and indices are genuine signs. For, while indices and icons are grounded in less than triadic relations of the sign to its object, they are nevertheless signs only in relation to a third element, a potential interpretant (the rule-established interpretant of a symbol is also a potentiality). The grounding relation is one thing and the sign relation another. The signhood of icons and indices cannot be explicated without reference to interpretants. Thus Peirce in 1903, with reference to any sign: 'The triadic relation is *genuine*, that is, the three members are bound together by it in a way that does not consist in any complexus of dyadic relations' (EP2:273, cf. 8.332).

In conclusion: signs are never of more than one division of any one trichotomy, except when analyzable as compounds of distinct signs. The textual evidence that that was Peirce's view is strong: the classes of sign he enumerated (chapter 9, section 1) are defined as combinations of divisions of distinct trichotomies, not as combinations of divisions of the same trichotomy.

5. Rheme, Dicisign, Argument

In his 1906 'Apology for Pragmaticism', Peirce wrote,

> A familiar logical triplet is Term, Proposition, Argument. In order to make this a division of all signs, the first two members must be much widened. (4.538; cf. 8.337)

Each sign must belong to one division each of each trichotomy, and therefore if this traditional trichotomy is to be used it must be broadened so that nonverbal signs such as symptoms and visual icons have a place in it. Icons and pokes in the back are like terms: they cannot be false or inaccurate. A photograph with a caption is like a proposition: it can be false or inaccurate. So also a weathercock indicating wind direction (EP2:297): if rusty, it might stick and misinform us. Argument alone has no generalization: for nothing but a verbal sign can appeal to the interpreter's own reason (to confirm that *q* does indeed follow from *p*) as an argument does. In addition to the generalization of the ideas of term and proposition, however, Peirce also intended to deepen or correct traditional analyses of propositions and arguments.

This third trichotomy is the one most thoroughly explored by earlier logicians yet most obscurely expressed by Peirce. The obscurity has several causes, one of which is the complexity of the topic. He was addressing simultaneously a number of difficult issues, some of which had never before been discussed. Some of these are now familiar to us, such as those that pertain to distinctions between sentence and statement (or assertion) and between either of those and what is stated. Here, Peirce was ahead of his time; but in some respects his remarks are by contemporary standards glaringly deficient, and that is a second cause of obscurity. As we will not discuss those problems until the next chapter (section 2), the reader's indulgence is begged for the present (e.g., as to Peirce's casual – and, I think, inconsistent – use of 'proposition'). A third cause of obscurity is that other issues Peirce discussed are still not widely recognized. We must make an effort to see what it is that he was puzzled by. It is these issues that we examine now. A fourth cause of obscurity is that Peirce was, as so often, never satisfied by his formulations, and so we find dramatically different, and probably mutually inconsistent, accounts of this trichotomy.

The language varied. For term, proposition, and argument (or, of the first two, their generalizations), respectively, we find: 'sumisign', 'dicisign', 'suadisign' (1903, EP2:275); or 'rheme', 'dicisign' (or 'dicent'), 'argument' (also 1903, EP2:292, and 1904, 8.337); or 'seme',[9] 'pheme', 'delome' (1906, 4.538, and 1908, EP2:481,490). For the present, we shall adopt 'rheme', 'dicisign', and 'argument', as being the most familiar of the alternatives; but later, midstream in the next chapter, we switch to the 'seme', 'pheme', 'delome' terminology, for reasons given then, and ride that horse to the further shore. So also, the definitions of

9 Confusingly, 'seme' is sometimes used as a synonym for an index (EP2:274).

these terms varied. For reasons of space, and fear of exhausting the reader's patience, I will forego close examination of the key passages (EP2:275–88, 292–9, 8.337–8, 4.538, 572, EP2:482–90) and simply state what I take to be their upshot, so far as consistency or a consistent drift can be found in them, omitting much that is of interest.

It is traditional to suppose that propositions are composed of terms and that arguments are composed of propositions. However, 'Theatetus sits' is not merely a name and a verb, signifying merely an individual and a type of act. The syntax, or mode of combination of these terms, is also significant, and the object signified – a fact or putative fact – is not signifiable otherwise. An argument, similarly, is not a set of propositions but, in Peirce's view, represents a process resulting in a conclusion. The problem is to account for these differences in signification, which Peirce sought to do, or came in stages to do, by identifying them as a sign's mode of influence on interpreters (1908: 'IX. As to the Nature of the Influence of the Sign' (EP2:490)). Whereas an argument makes an appeal to the interpreter's reason, a proposition professes without appealing, and a term merely calls attention to something without further profession.

One advantage of this analysis is that it reconciles our sense that term, proposition, and argument differ in complexity with the fact that terms can be compositionally complex ('mothers who too-late regret, most bitterly, having spoiled their last-born child, to its permanent detriment') while propositions can be simple ('Fire!'). The term or rheme, whether it is an icon, index, or symbol, is simple in function: it functions like an icon, merely bringing something to one's attention ('*like* a simple sign' (EP2:490, emphasis added). The dicisign, by contrast, presents itself as indexically related to the object it portrays (EP2:276),[10] and it has that doubleness of function (EP2:275), referring and portraying, even if it is quite simple in itself. Thus, 'Fire!', uttered with sufficient emphasis in the right sort of context, is to be interpreted as an index of its cause, though, as replicating a rhematic symbol, it also is descriptive; hence, it is taken or may be taken, rightly or wrongly, as descriptive of its cause. '[I]n order to understand the Dicisign it must be regarded as composed of two such parts [index and rheme] whether it be in itself so composed or not' (EP2:276). Similarly, the weathervane is an index of the cause of its

[10] In one of his uses of 'proposition', Peirce identified it with a symbolic dicisign (e.g., at EP2:278). But only through its replications in sinsigns can such a dicisign can be interpreted as related to its object indexically.

veering but is also understood to contain an icon of that cause's direction. Finally an argument, by appealing to a law (stated or not) relating premiss to conclusion, is essentially a symbol. 'It thus appears that the difference between the Term, the Proposition, and the Argument, is by no means a difference of [compositional] complexity, and does not so much consist in structure as in the services they are severally intended to perform' (4.572).

There is a difficulty in specifying the influence that distinguishes rheme, dicisign, and argument. Sometimes Peirce specified a dicisign's influence as an index-like compulsion (4.538) and other times he distinguished the dicisign from its assertion, for example, as a sign merely capable of being asserted (8.337). But is it compulsive when not asserted? Arguments, similarly, are sometimes said to be urged, other times, to be capable of being urged. We discuss these questions in the next chapter (sections 2 and 3).

As a logician, Peirce said, of course, a great deal about arguments and their classification, and about types of dicisign, on the logical forms of which the validity of arguments turn. But much of what he said about deduction, induction, and (his discovery) abduction has been written about and is well known, and thus is neglected here.

9

More Taxa

Let us now examine the principles implicit in Peirce's semeiotic taxonomy and then proceed to his later system of ten trichotomies and sixty-six classes of sign. That, anyway, is what we should like to do; but the principles are made less certain, not more, by the later taxonomy, in which unresolved problems proliferate. The promise of a system, rigorous and comprehensive, remains a promise only.

1. Principles of Semeiotic Taxonomy

Certain principles should be evident from the preceding chapter. They are that the system of classification is to be based on the general analysis of signhood as consisting in an irreducibly triadic relation, of sign, object, and interpretant; that each aspect of this relation is divisible trichotomically according to the three phaneroscopic categories; that each sign must be of one or another division of each trichotomy; and that no sign may be of more than one division of any trichotomy. These principles do not explain why certain inter-trichotomic combinations, such as indexical qualisigns or argumentative icons, are impossible. But the forbidden combinations exceed in number those permitted, and those permitted fall into a neat pattern that is the most salient feature of Peirce's semeiotic taxonomy. Consider the facts about the three trichotomies with which we are already acquainted.

A symbol signifies by a law that relates instances of one type (the symbol) to another type or to its instances. But a type is a 3rd. Now, legisigns alone are 3rds; for sinsigns are 2nds and qualisigns are 1sts. Therefore, legisigns alone may be symbols. Similarly, an index signifies via an

existential relation, and, therefore, it must be a 2nd or something – a 3rd – in which 2ndness is implicated. For that reason there can be no indexical qualisigns. In like manner, we see that an icon can be a rheme only. For that which signifies only through the possibility that it is or that it embodies or that its instances embody can neither profess anything nor appeal to the interpreter's reason. Some indices are rhemes; for example, those that only point. But others, that incorporate another sign, may convey information or misinformation about the object indicated, and these are dicisigns; for example, a weathercock (EP2:297) or portrait with a legend (8.341). But an index cannot be an argument; for, lacking the generality of a symbol, it cannot represent a law nor, hence, the principle by which premisses yield a conclusion. An argument, then, can only be a symbol. But a symbol may be either an argument, a dicisign, or a rheme. For example, a rule of inference can also be asserted in a symbolic dicisign, or named by a common noun, hence, a symbolic rheme.

These relations between the first and second trichotomies and between the second and third trichotomies entail like relations between the first and third trichotomies. Thus, since symbols alone may be arguments and legisigns alone may be symbols, it follows that legisigns alone may be arguments. So also, since qualisigns can only be icons and icons can only be rhemes, qualisigns can only be rhemes. And so on.

In that way we obtain the ten classes of sign noted in 1903 (EP2:294–7) and 1904 (8.341). All qualisigns are icons and all icons are rhemes; hence, one class of sign may be named 'qualisign' simply. But a sinsign may be either iconic or indexical, and an indexical sinsign may be either rhematic or dicent, yielding three classes. An iconic sinsign must be a rheme. A rhematic indexical sinsign includes paradigmatic instances of indices, such as pointings and pokes in the back. Any dicent sinsign (weathercocks, photographs, etc.) must be indexical but, as being informative, must 'involve' the complexity we have discussed (chapter 8, section 4). Legisigns may be either iconic, indexical, or symbolic; if iconic, then rhematic, but if indexical, then either rhematic (e.g., a proper name; 8.431) or dicent ('Thus any given street-cry, since its tone and theme identifies the individual, is not a Symbol, but an Indexical Legisign; and any individual instance of it is a Replica of it which is a Dicent Sinsign' (EP2:297)), and if symbolic, then rhematic (e.g., a common noun)or dicent (a proposition; EP2:295, 8.341 – but we will find reason in the next section to challenge this identification) or an argument. As arguments

can only be symbols, hence, legisigns, they may be named 'arguments' simply.[1]

Thus this table, where '1','2','3' represent the respective adicities of the trichotomous divisions of A, the sign in itself, B, the sign in relation to its object, and C, the sign in relation to its interpretant:

A	B	C	
1	1	1	qualisigns
2	1	1	iconic sinsigns
2	2	1	rhematic indexical sinsigns
2	2	2	dicent indexical sinsigns
3	1	1	iconic legisigns
3	2	1	rhematic indexical legisigns
3	2	2	dicent indexical legisigns
3	3	1	rhematic symbols
3	3	2	dicent symbols
3	3	3	arguments

The pattern of adicities is striking. It could easily be extended for any finite number of trichotomies, and we would expect that it should be so extended for any additional trichotomies germane to sign analysis. We would expect such an extension, just as we would expect that an inverse square law, found to apply to the solar system, will apply to other planetary systems. It therefore requires an explanation; it must be explained by a general principle.

That principle has been the subject of much discussion in the secondary literature, beginning in 1932, with footnotes Hartshorne and Weiss appended to 2.235–41, and proceeding to Weiss and Burks 1945, Lieb 1977 [1953], Sanders 1970, Savan 1977, and Müller 1994, among others.[2] It is our turn, now, to add to the confusion.

As previously noted, Peirce began to develop a taxonomy of signs in 1903, in an extensive work written to accompany his Lowell lectures of that date. Composed largely in one month, the 'Syllabus' shows a swift

[1] Despite the seriousness with which these ten classes have been taken by many, much that Peirce said in relation to them is unclear. E.g., as all arguments are legisigns, their replicas cannot be arguments but indications or images of arguments. Peirce said that the replica of an argument is a dicent sinsign (EP2:296). But why dicent? No explanation was given.

[2] See Sanders and Müller for citations of other work and Müller for trenchant criticism of the egregious nonsense written on this topic by Max Bense and Elizabeth Walther (cf. Short 1986b, pp. 121–3, for similar criticism of Bense).

development of thought; for example, in the third section, signs are divided by two trichotomies, but in the fifth section, a third trichotomy is introduced and placed in the first position (Editors' introductions, EP2:258, 267, 289, and their 535n5). In fact, several other trichotomies are suggested in the fifth section as well, although this seems not to have been noticed. Some of this material is repeated in a letter the next year to Lady Welby, in which, however, Peirce, suggested six trichotomies (8.327–41). The plot thickens in the Logic Notebooks (MS339) for 1905–6, leading up to a letter and letter draft of December 1908, addressed again to Lady Welby; the upshot is the famous ten trichotomies and sixty-six classes of sign. In the 1908 letter actually sent, Peirce stated clearly a principle of the sort for which we are looking. It contrasts in some deceptive ways to a principle implied in the fifth section of the 1903 syllabus. These are the textual sources with which we have to work.

We will begin with the earliest of those texts. In the third section of the 'Syllabus', Peirce wrote,

> A *Sign*, or *Representamen*, is a First which stands in such a genuine triadic relation to a Second, called its *Object*, as to be capable of determining a Third, called its *Interpretant*, to assume the same triadic relation. . . . (EP2:272)

Obviously, the position as 'first' does not make the sign less than a 3rd; 'first', in this usage, is ordinal, not categorial.[3] But no explanation is given why or in what sense the sign is the first of the three relata. In at least one sense of 'first', the object is first: it is so, if we follow Peirce's frequently repeated formula that objects determine signs and signs determine interpretants. Why is that order altered here?

Not until the fifth section is the general problem of semeiotic taxonomy considered:

> We must distinguish between the First, Second, and Third Correlate of any triadic relation.
>
> The First Correlate is that one of the three which is regarded as of the simplest nature, being a mere possibility if any of the three is of that nature, and not being a law unless all three are of that nature. . . .
>
> Triadic relations are in three ways divisible by trichotomy, according as the First, the Second, or the Third Correlate, respectively, is a mere possibility, an

[3] Perhaps it was by such usage that David Savan was misled into arguing that Peirce's categories *are* ordinal, i.e., that whether something is a 1st, a 2nd, or a 3rd depends on the role it plays in a triadic relation and, hence, that its category may change depending on the context in which we examine it. See Savan 1977 and Short 1986a with Savan's response, Savan 1986.

actual existent, or a law. These three trichotomies, taken together, divide all triadic relations into ten classes. . . .

A *Representamen* is the First Correlate of a triadic relation, the Second Correlate being termed its *Object*, and the possible Third Correlate being termed its *Interpretant*. . . . (EP2:290)

The ordinal number of the correlate is thus explained by its 'simplicity' relative to the other correlates, the simpler having the lower ordinal number. Simplicity, in turn, consists in a fact, that the phaneroscopic category of one correlate limits what may be the categories of the other correlates. Concisely: if one correlate cannot be of a higher phaneroscopic category than another, then it is simpler than that other. As so defined, 'simpler than' is irreflexive, asymmetric, and transitive. The ordinal numbering of correlates depends on the general fact that no two correlates of the same relation are of the same degree of simplicity. That fact, in turn, explains why Peirce said that we obtain ten classes only, and not the twenty-seven we would otherwise expect, from combining three trichotomies. In general, where trichotomies are of items ordered by relative simplicity, the number of classes produced by n trichotomies $= \frac{1}{2}(n+1)(n+2)$.

That is the principle – we shall call it the principle of relative simplicity – implicit in 1903. It must be stressed that it does not by itself determine the order in question. It presupposes the facts alluded to, that a sign's belonging to one division of one trichotomy limits to which divisions of another trichotomy it might also belong. The principle no more than formulates the pattern that is a consequence of those facts.

So far, so good. But do the facts support Peirce's claims that the sign is simpler than the object and that the object is simpler than the interpretant (as simplicity is here defined)? And what does any of this have to do with the preceding table, which is based on trichotomous divisions of sign, sign's relation to its object, and sign's relation to its interpretant (let us denote this compactly as ABC, regardless of the ordering of the items), rather than on trichotomous divisions of sign, object, and interpretant (hereinafter AB′C′)? Suppose that ordering by simplicity applies not only to the three 'correlates' but also to other aspects of semeiosis, such as relations among the correlates. Then the table shows that a sign's relation to its interpretant is simpler (in this special sense of simplicity) than its relation to its object and that the sign in itself is more complex (in this same sense) than either relation. If we assume, further, that the ordering of ABC by relative simplicity is the same as that of AB′C′, then it would appear that Peirce erred in claiming that the sign is the simplest correlate and that the interpretant is the one most complex. Perhaps

that is the reasoning by which Hartshorne and Weiss concluded that 'the terms "First Correlate" and "Third Correlate" should be interchanged', that is, as to most simple and most complex (2.235n). Then the first correlate, or the sign, A, is the most complex and the third correlate, or interpretant, C, is the simplest.

But it is not clear what Hartshorne's and Weiss's reasoning was; for they did not refer to the principle of relative simplicity but to the seemingly quite different principle Peirce stated in 1908 (which they cited without giving its source, the principle not yet having appeared in print), that 'It is evident that a Possible can determine nothing but a Possible; it is equally so that a Necessitant can be determined by nothing but a Necessitant' (EP2:481). A possible is a 1st and a necessitant is a 3rd. The principle is that nothing can determine anything of a higher category than itself. We shall refer to this as the principle of determination. (The determination may of course be made *in* something of a higher category, e.g., a quality can be produced in an object.) A 1st can be determined by a 1st, a 2nd, or a 3rd, a 2nd by a 2nd or 3rd, and a 3rd by a 3rd, but a 3rd cannot be determined by a 2nd or 1st, nor a 2nd by a 1st. Now, insofar as 'determine' means nothing more than to limit (chapter 6, section 7), this principle is the converse of the principle of relative simplicity: X determines Y if and only if Y is simpler than X. Each is a principle by which a sign's adicity in one trichotomy limits what may be its adicity in another trichotomy.

There is nothing problematic about claiming that X limits Y and Y limits X; for limiting is a two-way street. Suppose that we have two sets, $\{a,b\}$ and $\{x,y\}$, and that members of the first set may be related by R to members of the second set; then members of the second may be related by the converse relation R′ to members of the first. Thus, if a is limited by R to x, while b is R both to x and to y, then, conversely, y is limited by R′ to b, while x is R′ both to a and to b. Facts of limitation may thus be stated in either of two ways. If ABC ordered by simplicity is C, B, A, then ordered by determination it is A, B, C. But we get the same relations of adicities in either case; their tabular representation is merely flipped over.

By the same token, if 'determine' means only to limit, then the choice of direction of determination is arbitrary. One alternative is chosen by adopting the principle that nothing can determine anything of higher category than itself. But we could as easily choose as our principle that nothing can determine anything of lower category than itself. Then the direction of determination would be the reverse. The arbitrariness of direction of determination, if determination consists merely in limiting, will become of some importance to us in a moment.

Now, if the order of relative simplicity of ABC is the same as that of AB′C′, then Peirce erred in identifying the sign, A, as the simplest correlate and the interpretant, C′, as the most complex correlate. For the facts represented in the table show that A is the most complex and C is the simplest of ABC. But is the assumption correct that ABC and AB′C′ are alike in order of relative simplicity (or, conversely, in order of determination)? We must notice an uncomfortable fact, that in the 1908 passage Hartshorne and Weiss drew on, Peirce went on to say that it is the object that determines the sign and the sign that determines the interpretant (EP2:481; for more detail, see below, section 4). That gives us B′, A, C′ as order of determination, which is compatible neither with A, B′, C′ as order of determination nor with C′, B′, A.

Peirce claimed that the preceding sequence 'follows from the Definition of a Sign'; earlier in the same letter he defined a sign as determined by its object and determining its interpretant (EP2:478). Perhaps in that definition he meant by 'determine' something different from merely to limit (in chapter 6, section 7, we attempted to determine what that additional meaning might be, and failed). However, all that matters in the present context is the order of limitation of adicities by adicities; therefore, if Peirce meant something else by 'determine', that is irrelevant to the taxonomic principle at issue. Or perhaps, since the direction of determination is chosen arbitrarily, he unwittingly assigned 'determine' different directions (a) in the determination of sign by object (since the sign cannot be of lower adicity than the object) and (b) in the determination of interpretant by sign (since the interpretant cannot be of higher adicity than the sign).

It is easy to see, then, why there has been so much controversy about the principles of Peirce's semeiotic taxonomy – especially since (amazingly) none of the participants in that controversy has stopped to consider what Peirce meant by 'determine'. Nor has the assumption that ABC and AB′C′ must be in the same order of relative simplicity (or of determination, in some definition thereof) been discussed. A further source of confusion might be the assumption that Peirce himself was clear and consistent about these matters.

The controversy pertains primarily to the extension of the order of adicities manifest in the table of three trichotomies to Peirce's later set of ten trichotomies. Weiss and Burks place the trichotomy based on the sign in itself first (qualisign, sinsign, legisign), then, in this order, trichotomies based on the immediate object, the dynamic object, immediate interpretant, dynamic interpretant, final interpretant, relation of sign to dynamic

object (icon, index, symbol), relation of sign to dynamic interpretant, relation of sign to final interpretant (rheme, dicisign, argument), and relation of final interpretant to object (1945, pp. 386–7). We note that this places AB′C′ in the same order of determination as ABC. But Lieb reversed the first three items, so that the dynamic object determines the immediate object, which determines the sign, which then determines the remainder in the same order (LW:163–4). Two reasons Lieb cited for his amendment are that it conforms to the order Peirce himself asserted (EP2:481) and that it follows the order of determination in Peirce's definitions of 'sign'.[4] That of course entails that the order of AB′C′ is not the same as the order of ABC.

But does it matter? I think we have been following a will-o'-the-wisp of system. For the principle at issue must in any case depend on an independent determination of whether one trichotomy limits the adicities possible in another trichotomy and, if so, in which direction. The matter is quite clear in Peirce's initial three trichotomies. In his later, very sketchy discussion of ten trichotomies, he himself confessed great uncertainty at a number of points, going so far as to introduce a notation of Greek letters by which to register the varying degrees of his uncertainty (EP2:483).

2. Dicisigns and Assertion

J. L. Austin complained, 'With all his 66 divisions of signs, Peirce does not, I believe, distinguish between a sentence and a statement' (Austin 1961, p. 87n1). That is correct, if the distinction has to be framed in those exact terms. However, Peirce, unlike Austin, was not explicating the ordinary usage of 'statement' and other words we employ in speech about speech. Instead, he wrote about language in the technical terms provided by his theory of signs. And part, at least, of Austin's account does appear in Peirce's theory, otherwise worded. Austin said that 'statements are made, words or sentences are used' (ibid., p. 88). Similarly, Peirce distinguished between a dicent symbol (which we know must be a legisign, specifically, a sentence or something on the order of a sentence) and its replication in a sinsign; and any replica is a use of the legisign replicated. Of course,

[4] The two remaining reasons seem to me not germane. One is that Peirce's commitment to realism 'requires that cognition conform to things, not things to cognition' (LW:164n1). Yes, but conformity and determination are not necessarily the same thing, and especially not if determination is simply limitation of adicity by adicity in one or another direction.

a sentence can be replicated without making a statement; and when a statement is made, it is not identical to the sentence-replica by which it is made. These facts bring us to other parts of Austin's account, but also to other parts of Peirce's theory.

How does one use a sentence to make a statement? One utters it assertively (Austin thought that 'assertion' and 'statement' are nearly equivalent, the former being 'perhaps . . . slightly wider'; ibid., p. 88). That can be done in various ways, depending on context. Often, the context alone makes any utterance of a declarative sentence assertoric, if it is not embedded in such other words as 'I doubt that . . .' or 'Only a fool would think that . . .'. Otherwise, a ringing tone may suffice, or the additional words, 'I assure you that . . .' or 'I know that . . .'. (There are some contexts where even these devices do not enable one to make a statement: in a dramatic production, they enable one only to pretend to be making a statement; hence an audience's uncertainty if an actor shouts 'Fire!'). But what is assertion in general? Ironically, Peirce's account of assertion in general anticipated Austin's own, rather famous thesis that saying 'I know that . . .' is a sort of 'performative' utterance, akin to saying 'I promise', making one vulnerable to censure if what is said turns out not to be true (ibid., pp. 66–71). Thus Peirce:

> According to my present view (I may see more light in future) the act of assertion is not a pure act of signification. It is an exhibition of the fact that one subjects oneself to the penalties visited on a liar if the proposition is not true. (8.337)

Comparison may be made more directly to John Searle's Austinian analysis of assertion (Searle 1969, ch. 1; cf. Brock 1981b).

Now we have two-thirds of Austin's analysis restated in Peircean terms. The last third is more problematic. A statement is not a particular assertoric utterance, since we say that John and Jacques made the same statement (said the same thing), even though their utterances were distinct (made by different persons, perhaps at different times and places, and possibly by replicating different sentences). And that which is stated – and which might also be denied, questioned, or merely contemplated – is yet something else, not identical with the statement of it, though, like the statement, it is (normally) either true or false (and the one is true if and only if the other is). Austin is chary of using the philosopher's term 'proposition' for that which is stated – on the not very compelling grounds that 'proposition' in ordinary usage is 'portentous' and that some philosophers have identified propositions with the meanings of sentences

but '[W]e never say "The meaning . . . of this sentence . . . is true"' (ibid., pp. 86–7).

If one philosopher may make 'proposition' a technical term, another may make it a technical term differently defined. There is nothing stopping so strange a person from defining a proposition as that which can be asserted – and *not* ham-handedly adding that this entity is a meaning. Peirce wrote:

> One and the same proposition may be affirmed, denied, judged, doubted, inwardly inquired into, put as a question, wished, asked for, effectively commanded, taught, or merely expressed and does not thereby become a different proposition. (EP2:312)

It is important to understand the nature of such a definition. It does not specify the entity defined in terms of its properties (e.g., as being a meaning); rather, it introduces it as an abstraction from *other*, more concrete phenomena. The proposition is *that* (whatever it may be) *which* can be affirmed, denied, and so on. So also, one might define length as *that which* can be measured in such and so ways. Peirce named such definitions 'hypostatic abstractions', about which more is said in the next chapter.

Now, propositions, so defined, need no more be identified with entities otherwise defined than Austin identified statements with anything otherwise defined. Statements and propositions, so defined, are very useful *façons de parler*, and may be no more than that (whether more or not depends on other considerations). But then there is no reason why they have to be regarded as signs. And if they are not signs, then Austin was quite wrong to have complained that Peirce failed to assign statements to a class of sign.[5]

So much for Austin's poorly aimed barb. There are, however, some genuine difficulties with Peirce's doctrine of dicisigns and assertion. One is that he used the term 'proposition' not only in the sense just explored but also, and inconsistently, to refer to a dicent symbol (EP2:275ff., 295). Thus, 'A *proposition* as I use the term is a dicent symbol. A dicent [i.e., a

[5] We could have come to the same conclusion on other grounds earlier, after quoting Peirce to the effect that assertion is not a pure act of signification. For if assertion as such does not signify, then neither do statements as such signify, though the sentences uttered in making statements do signify. And therefore statements will not be a class of sign. And therefore Austin was wrong to look for statements among Peirce's 66 classes of sign. In a moment, however, I argue that Peirce erred in denying that assertion is significant; and that is why I do not push this line of argument.

dicent symbol, not a dicent sinsign] is not an assertion, but is a sign *capable* of being asserted' (8.337). But a sentence is not what is asserted when it is used to make an assertion. For what is asserted, we have noted, may be asserted also using other sentences. Consider, too, the fact, well known to Peirce, that different replications of the same sentence sometimes assert different things. That is especially the case if the sentence contains demonstrative pronouns. 'I am in pain', asserts one thing if uttered by you; uttered by me, it asserts something else. But if the sentence is not what is asserted, then it is not what is capable of being asserted. The capacity of a sentence is, rather, that it can be used to make an assertion.

In another place, Peirce wrote, 'It is one and the same proposition every time it is thought, spoken, or written, whether in English, German, Spanish, Tagalog, or how' (MS599). This is also a hypostatic abstraction, and it can be combined with the one aforementioned: a proposition is that which can be asserted, denied, and so on, in speech, thought, or writing, in any number of different languages. But a dicent symbol is a legisign, necessarily of a given language, and, therefore, it cannot be what is spoken both in English and in Tagalog.

The equivocation Peirce appears to be guilty of is akin to another that we noted earlier (chapter 8, section 1): his sometimes identifying legisigns by their laws of formation and other times identifying them by their laws of interpretation. In this case, the equivocation is between sentence-legisigns defined by formation rules of one or another language and the result (the immediate interpretant) of applying the associated rules of interpretation to the replicas of those legisigns.

Equivocation on 'proposition' is not a serious problem for Peirce's semeiotic.[6] It can be treated as a casual blunder, easily rectified. In one sense, a proposition is a sign, namely, a dicent symbol. In another sense, it is a hypostatic abstraction from a set of (in some respect equivalent) replications (actual or possible) of diverse dicent symbols. The latter use of 'proposition' seems to me to be preferable to the former. The former is misleading, as it departs so far both from ordinary and from philosophical usage (however various and often incoherent the latter may be). Henceforth, we shall use the word 'proposition' in the hypostatically abstract sense. Lacking a reason to think otherwise, we assume that propositions in this sense are not signs.

[6] Indeed, it is easily overlooked, as it has been by the authors of many otherwise penetrating discussions of Peirce on propositions: Brock 1981a, 1981b; Hilpinen 1983, 1992; Hookway 1985, pp. 128–30; Houser 1992; and others.

But the equivalence of replicas, from which propositions are abstracted, is not easily defined. It has to be an equivalence in meaning that entails equivalence in truth or falsity; but equivalence in truth or falsity does not itself suffice, as different propositions may agree in being true or in being false. So, what additional likeness in meaning is required? It may be wisest to allow the determination of equivalence to vary from case to case, depending on one's analytic purpose. Then there is no precise definition of 'proposition' in general.[7] And why should there be?[8]

A more serious problem is whether Peirce was either correct or consistent when he denied that assertion (the act of assertion, the assertoric replication of a sentence – in contradistinction to the assertion, or statement, thus made) is significant. Assertion 'is not a pure act of signification', he said in 1904 (8.337). So also, a year earlier:

> A judgment is the mental act by which the judge seeks to impress upon himself the truth of a proposition. It is much the same as an act of asserting a proposition, or going before a notary and assuming formal responsibility for its truth, except that those acts are intended to affect others, while the judgment is only intended to affect oneself. However, the logician, as such, cares not what the psychological nature of the act of judging may be. The question for him is, What is the nature of the sort of sign of which a principal variety is called a proposition, which is the matter upon which the act of judging is exercised? The proposition need not be asserted or judged. It may be contemplated as a sign capable of being asserted or denied. This sign itself retains its full meaning whether it be actually asserted or not. (EP2:292–3)

This is an example of Peirce's rejection on the 'psychologism' of many nineteenth-century logicians, mainly German, who supposed that logic is a science of judgment, a mental act. Antipsychologism in logic is one of several parallels between Peirce and his contemporary, Gottlob Frege.

[7] Nor is it necessary to identify a hypostatically abstractive entity with anything at all – as, e.g., Frege might have identified a length with a set of equal measures. What is not necessary may nonetheless be both possible and germane to some purpose, e.g., formulating a theory in set-theoretical terms. But such a purpose is far from the spirit of Peirce's semeiotic. And therefore I abjure my own former suggestion (Short 1984; cf. Hookway 1985, pp. 128–30) that a proposition be identified as a set of replicas (viz., that set from which its abstraction is made) or, alternatively, as a type of replica (defined by common meaning rather than by legisign replicated).

[8] A formal system of propositional logic does not require a decision about what will count as a proposition, beyond certain strictures as to its truth value. We may allow the question of the reality of propositions to be mooted. A legisign comes into existence at a given time, whereas it is unclear whether a proposition comes into existence at all (is not a proposition independent of there being means for expressing it?). Furthermore, a legisign makes sounds or shapes of a given pattern ('Repent!') more likely (less unlikely) than they would otherwise be, whereas it is unclear whether propositions as such have any influence on events at all.

Peirce's variant treats judgment as being on a par with assertion. If, as he held, thought is internalized discourse, then the private act must be modeled on the public one: 'I says to myself, says I'.

We need not be deterred by Peirce's referring to a proposition as a sign; we have already acknowledged that his use of that term was inconsistent. But was he right that asserting is not significant? However much psychologism is to be avoided in logic, qua science of valid inference, is not the distinction between asserting and expressing a proposition (in our sense of that term) one that is of central importance to logic in the broader sense, as an analysis of the methods of inquiry, or to semeiotic? For, surely, it makes a difference whether a scientist, in describing an experiment, mentions a possible reading of a voltmeter or asserts that such a reading was made. Thus assertion plays a key, ineliminable role in inquiry. An assertion signifies (veraciously or not) a fact in a way that the mere expression of a proposition does not.

And, in fact, this form of signification is implied in some formulations of the concept of a dicisign. How can a dicent symbol be 'informational' (EP2:275) if it only formulates and does not assert a proposition? How can it '*profess* to refer or relate to something as having real being' (ibid., emphasis added) apart from its being employed in making an assertion? How can its (final) interpretant represent it 'as a sign of fact' (EP2:291) if it only expresses a possible fact? As replicas alone can be assertoric, these definitions are at fault. They must be modified (as we have seen – chapter 8, section 4 – Peirce in some passages did modify them) to refer to capacities rather than to actualities. But the capacities in question are those of being used to make an assertion, and are not of being asserted. Now, a capacity is defined in terms of what it is a capacity for. Hence, we must understand what it is to assert something, before we can understand what a dicent symbol signifies. It would seem, then, that assertion is itself a form of significance, a form from which the significance of a dicent symbol is abstracted.

That is not to deny the important point that Peirce, in his antipsychologistic mode, makes, that we can understand what is asserted, independently of its being asserted. It is only to say that to grasp any proposition, asserted or not, one must have an idea of the possibility of its being asserted, which presupposes acquaintance with assertion in other cases. To put the same point another way: we do not know what a possible fact is if we do not know what a fact is, and we have no acquaintance with facts except through their representation in judgments and assertions. The very idea of fact, like that of existence more generally, depends on our experience of the world's hurly-burly, of its 2ndness.

Now, if assertion is a class of sign distinct from the dicent symbols that express the propositions asserted, then what class is it? The answer is obvious, in part. Assertions are replicas of dicent symbols, and 'The Replica of a Dicent Symbol is a Dicent Sinsign of a peculiar kind' (EP2:296). But that is not the whole answer, since not every such replica is an assertion. Some (with minor variations in word form, syntax, etc.) are questions, some commands (cf. EP2:312, quoted above), some hypotheses, and so on. And thus we should expect to find a subdivision of dicent sinsigns reflecting these differences (in what grammarians call the 'mood' of a sentence and Austin called its 'force').[9] It may be that that was one of the motives behind Peirce's introduction of seven additional trichotomies of sign, to which, but for one more comment, we now turn.

It should be understood that what has been said in this section about dicent symbols has its parallels with respect to arguments. As with 'proposition', Peirce used 'argument' inconsistently to refer to what is expressed and to the legisigns used to express it (the same ambiguity is found in ordinary usage and sometimes in the writings of logicians). By 'argument' we shall mean what is expressed; so understood, the same argument may be expressed using any of a great variety of legisigns, in the same or in different languages. And arguments, so understood, may be represented without being asserted or submitted. Their being expressed, urged, submitted, questioned, labeled fallacious, rejected, and so on, is a matter of individual replications differing in 'force'. The distinction between an argument and its being submitted, like that between a proposition and its being asserted, requires another division of signs.

But an argument in this sense is not a sign. And therefore we shall now adopt the 'seme', 'pheme', 'delome' terminology in lieu of that of 'rheme', 'dicisign', 'argument'. A delome is a legisign that can be replicated in order to present, urge, or submit an argument, and that same argument can be represented similarly by quite different delomes.

3. Six Trichotomies

Already in 1903, in the fifth section of the 'Syllabus', in the second ellipsis in the long quotation in section 1 from that document, Peirce had

[9] This is not unfamiliar territory. For example, R. M. Hare distinguished the 'phrastic', e.g., 'S's being P', from the imperative or indicative 'neustic' that, attached to the phrastic, forms a command or an assertion, respectively: 'S's being P – please' or 'S's being P – yes' (Hare 1952, p. 18). To suppose that a proposition can sometimes be commanded rather than asserted, as Peirce sometimes did, and as we shall do, is to identify it with what a phrastic expresses.

intimated a second trio of trichotomies additional to that one he first mentioned. The first trio, you may recall, was by trichotomous division of each of the three correlates of a triadic relation. 'There will be besides a second similar division of triadic relations into ten classes, according as the dyadic relations which they constitute between either the First and Second Correlates, or the First and Third, or the Second and Third are of the nature of possibilities, facts, or laws' (EP2:290). Trichotomous division of a relation named 'dyadic' has occasioned some comment (2.239, editor's note), but surely all that Peirce meant here is that we should consider the three correlates taken two at a time, the question being whether the relation between a given pair is (elliptically) triadic, or is (genuinely and irreducibly) dyadic, or is (reducible to being) monadic.[10] He assumed that each trio is ordered by relative simplicity, as is evident from his claiming that each yields but ten classes of sign. The three trichotomies of sign with which he then proceeded to work was a hybrid of those two trios of trichotomies applied to the sign relation. It was of the first correlate, relation of first to second, and relation of first to third. As this also yielded but ten classes, relative simplicity appears to have ordered not only each trio but the two together. None of this has been much noticed by Peirce's commentators, but, having noticed it, we should not be surprised by the introduction of other trichotomies of sign in subsequent years.

In a 1904 letter to Lady Welby, Peirce divided signs by six trichotomies, but not the six suggested in the 1903 'Syllabus'. These six result from the hybrid trio already elaborated, as subdivided according to the distinctions newly made between a sign's two objects and three interpretants. They are: (1) what the sign is in itself (qualisign, sinsign, legisign); (2) what it is in relation to its dynamic object (icon, index, symbol); (3) 'in respect to its immediate object', a sign is 'of a quality, of an existent, or of a law' (the quality referred to must be a 1st, not a 3rd, hence, not such a quality as is named in a symbol like 'red'); (4) in relation to its 'signified' (final) interpretant (seme, pheme, delome); (5) 'a sign may appeal to its dynamic interpretant in three ways' (of which more in a moment); and (6) 'in its relation to its immediate interpretant', signs are interpretable 'in thoughts or other signs of the same kind in infinite series', 'in actual experiences', or 'in qualities of feelings or appearances' (8.334–9).

The three new trichotomies are (3), (5), and (6). If we suppose that 'interpretable in actual experience' means interpretable in actions (and it is hard to know what other form of 'actual experience' could interpret

[10] For the concepts of ellipticity and reducibility of relations, see chapter 3, section 3.

signs), then (6) is the emotional/energetic/logical division of interpretants applied to immediate interpretants alone.[11] (Logical interpretants, *avant la lettre*, are treated, as one might expect, in conformity to the pre-1907 view that they are always themselves signs and thus form an infinite series.) In the case of (3) and (6), what is being categorized phaneroscopically is not really the sign's relation to its immediate object or immediate interpretant but that object or interpretant itself; arguably, immediacy precludes difference in relation.

As (3) seems obvious, it is surprising that it was not introduced earlier. Perhaps it was supposed that semes, phemes, and delomes represent qualities, existents, and laws, respectively. (That is how that trichotomy was defined in 1903 (EP2:292), except for the qualification, 'for its interpretant'.) But in fact the one trichotomy does not repeat the other. For a law can be asserted in a pheme ('Given *p* and given *q* if *p*, then *q*') or named in a seme ('modus ponens') as well as 'urged', in a delome, as the leading principle of a particular argument. Furthermore, laws not of logic, though they may serve as leading principles of inference, cannot be accepted by reason alone, though they can be asserted and named. Furthermore, a fact is not the only sort of existent (if it is one at all), other existents being referred to by semes (e.g., proper names). But a quality (1st) can be the object of a seme only, namely, a pure qualitative icon.

The trichotomy that bears on the problem of distinguishing the various forces with which symbolic phemes may be replicated (and, hence, which bears on the problem of distinguishing assertion from the proposition asserted and from the sentence replicated in its assertion) is (5). The three ways that a sign may appeal to its dynamic interpretant (in an order the reverse of what one might expect) are:

> 1st, an argument only may be *submitted* to its interpretant, as something the reasonableness of which will be acknowledged. 2nd, an argument or dicent may be *urged* upon the interpretant by an act of insistence. 3rd, argument or dicent may be, and a rheme can only be, presented to the interpretant for *contemplation* (8.338).

That is all that Peirce in this place said about these divisions, though it echoes earlier discussions of propositions and arguments. Notice that 'urge' replaced 'assert', as naming that force with which either

[11] This is further evidence that the immediate and the emotional are not the same division of interpretants.

propositions or arguments may be expressed, and was replaced by 'submit', as naming the force with which arguments alone may be expressed. This corresponds better with the ordinary meaning of 'urge'. I can urge you to accept my argument, unexamined, just as I can urge you to take my word for something I say; but an argument, only, may be submitted to your reason.[12] But 'urge' may also be preferred as being more general than 'assertion', thus covering imperative utterances: I can urge you either to believe that S *is* P or to act so as to *make* S P.

For the sake of uniformity, emphasis in the third division should be moved from 'contemplation' to 'presented', so that, as in the other divisions, it is on the verb. Then, from these three verbs, we need to form names of kinds of sign; for the trichotomy is supposed to be of signs, and not only of what can be done with signs. Something that can be either submitted, urged, or presented is not thereby divided; for it is the same in each. So also, a legisign that is replicated in three such acts is the same in each. Let us therefore speak of *presentations*, *urgings*, and *submissions* as three kinds of sign.

As before, we have to correct Peirce in one important respect. Where he wrote that 'a dicent may be urged', the truth is that a dicisign or pheme may be used to urge a proposition. It is not the sign itself that is urged. So also, the delome replicated in a submission is not submitted; it is the argument expressed that is submitted. So also, we contemplate an object named, and not the name; to present a name to be contemplated, it itself must be named or otherwise signified.

It might seem that trichotomy (5) violates the principle that each sign must fall into exactly one division of each trichotomy. For (5) appears to subdivide sinsigns alone. Only a sinsign can be a presentation, urging, or submission. However, we have long ago seen that Peirce acknowledged that qualisigns signify only through being embodied in sinsigns and that legisigns signify only through being replicated in sinsigns. The crucial point to notice about (5) is that presentation, unlike urging and submission, does not add to the significance of the qualisign, sinsign, or legisign employed. Its only effect is to direct attention; it adds no force. Thus, when we speak of what a qualisign or legisign in itself means, we are referring to its occurrence in a presentation. Antipsychologism in formal logic is

[12] One might object that a truth of logic is a proposition that can be submitted (and is more properly submitted than urged) to the interpreter's reason. Yes, but is not submitting it the same as making an argument? '*p* is logically true' means that *p* is deducible without assuming anything empirical.

thereby satisfied, namely, by (5)'s recognizing nonassertoric expressions of a proposition.

If urgings comprise both assertions and commands, then these two must be distinguished by another trichotomy. That is perhaps the motive behind (6). Here again, we need terms for types of sign. The 'emotional', 'energetic', 'logical' terminology is far from ideal, but, as it is Peirce's, let us adopt it. If a sign's immediate interpretant is emotional, energetic, or logical, the sign itself shall be called 'emotional', 'energetic', or 'logical', respectively. Imperative utterances may then be classed as energetic urgings, while assertions are logical urgings. The immediate interpretant of the one is an action, of the other, a belief or other sign.

Peirce did not make these claims about (6) and its relation to (5); but they are I think reasonable guesses as to the import of those trichotomies, and those guesses address questions that we know Peirce had at one time in mind. To test this idea, we should have to see whether the combinations of (5) and (6), so construed, (a) provide reasonable pigeonholes for signs of important kinds, (b) exclude no signs, and (c) exhibit the familiar pattern of ordering by relative simplicity. Indeed, to explore and confirm the six trichotomies of 1904, we should have to do the same for all of them together (if successful, the yield will be twenty-eight classes of sign). Unfortunately, the hard work is left for us to do; Peirce gave no hint in his letter of how these trichotomies are either combined or ordered – with one exception to which we now turn.

In the statement of (5), its ordering relative to (4) is given. Rhemes, dicisigns, and arguments, Peirce said, may be presented; dicisigns and arguments may be urged; and arguments alone may be submitted. Restating this as above, we have: semes, phemes, and delomes may be presentations; phemes and delomes may be urgings; and delomes, alone, may be submissions. Using, as before, '1', '2', and '3' to denote the respective divisions of each trichotomy by their adicity, we obtain this table:

(4)	(5)	
1	1	presentational seme (nouns, verbs, pictures, pokes, pointings, etc.)
2	1	presentational pheme (hypothesis)
2	2	an urging pheme (assertion or command)
3	1	presentational delome
3	2	an urging delome
3	3	submissional delome

All three types of delome are ordinarily referred to as 'arguments', with the ambiguity already deplored.

So far, the order represented is that which Peirce stated in 8.338. We have no such guide to the order of trichotomy (6) with respect to the others. The reader is invited to try out the three different possibilities, putting (6) to the left or between or to the right of (4) and (5). (The aim is to produce a table with the usual pattern of adicities, as in the one following, in which each line represents an admissible type of sign; an order-candidate is eliminated when one or more lines of its table is found to represent an impossibility.) The following is the order that seems to me to work best:

(6)	(4)	(5)	
1	1	1	any work of art so far as 'pure'
2	1	1	pokes in the back, pointings
2	2	1	questions
2	2	2	commands, moral imperatives
3	1	1	nouns, verbs, adjectives, adverbs
3	2	1	hypotheses, proposed plans
3	2	2	assertions
3	3	1	the presentation of an argument
3	3	2	the urging of an argument
3	3	3	the submission of an argument

This entails that (6) is simpler than (4), which is simpler than (5), as relative simplicity was defined above (section 1). Such an order is contrary to what we would expect, as it places relation to final interpretant between relations to immediate and dynamic interpretants; but I could make no other arrangement work.[13]

Any such table must be subjected to criticism, before it can be accepted, even tentatively. But the question is less whether criticism can be met plausibly than whether the process proves illuminating. As always, we insist on the Peircean-Popperian test of fruitfulness; plausible answers are always possible, given sufficient verbal facility, and therefore plausibility counts for little.

One objection, pertaining to the first line, is that works of art can be composed of phemes, as novels, dramas, and many poems and songs are. This has already been answered (chapter 8, section 2): in verbal

[13] And, as I argue in the next section, it may be doubted whether these trichotomies are properly conceived of as relations to those respective interpretants.

art, words, many of which are symbols, are used to create icons; by the same token, phemes are used to create semes. Actors are given lines to speak, but it is the character and action thereby portrayed that elicits our aesthetic response. When an actor on stage intones, 'It is a fine day', we do not ponder the truth of whether it is a fine day. So far as a work of art urges ideas for our agreement or disagreement, whether those ideas are expressed in so many words or not, that is another matter: hence the qualification 'pure' in our table (to which qualification no further importance is attached).

Another objection pertains to the third line: how is it possible for a presentational seme to have an action as its immediate interpretant? It is this question that most persuades me that the table is correct: for it unexpectedly led to what appears to be a solution to an otherwise unsolved problem. Peirce suggested that questions be treated as a subtype of command, which is reasonable; yet the question, 'Is S P?' cannot be a subtype of the command, 'Make S P'. Questions, then, cannot be located in the fourth line. But, unlike the command and the assertion, the question 'Is S P?' does not urge something that has S's being P as content. It urges an energetic response – to give an answer or to seek one – but it does not dictate what that response should be. It is therefore a presentational replication of a symbolic pheme the immediate interpretant of which is energetic. Thus the table's third line.

A third objection pertains to the fact that the table makes all emotional signs to be presentational semes, whereas some statements are clearly meant to be interpreted emotionally, for example, moral harangues (chapter 7, section 5). This requires us to look more closely at the logical structure of admonition, reprimand, reproach, praise, approbation, condemnations, appeals for sympathy, and so on – any mode of discourse meant to elicit feeling.[14] (Will our attempt to square them with our table result in an unnatural forcing, or will it produce new insight?) Now, such discourse is normally replete with, when not consisting entirely of,

[14] Of course, such discourse serves many purposes and is meant to have several related effects, varying by case. Fixing blame has less to do with evoking feeling than establishing fact, perhaps with social or penal consequences (others' energetic interpretants of those facts), and much of condemnation and the like partakes of that character, while approbation is of the opposite practical tendency. But we may focus exclusively on the emotional effects intended, whether or not they are meant to be followed by changes in conduct by the party in question or in actions taken by other parties. In all of this, it is a nice question when the meanings of utterances end and further expected effects that are not parts of meaning begin; see below.

statements of putative fact. Such assertions are to be interpreted logically, if at all. It is the facts thus presented – the vivid picture of my ill-considered act and its long train of unhappy consequences – that are to evoke the desired emotion. The case is like that of verbal art, and a good deal of art goes into selecting and emphasizing the facts that will produce an emotional response. The emotion, if it is an interpretant at all, interprets not the words but the picture they paint: a seme, albeit a very complex one, composed of numerous facts. Utterance specifically emotive in language ('Shame on you!' – effective only when resting on a solid factual case for the prosecution) is perhaps the only part of the discourse that has a type of emotion for its immediate interpretant. But are those parts statements in the sense of being assertions? Are they phemes at all, grammatical appearance to the contrary? My aunt's masterful summation, 'You are a very bad boy', is an assertion, but it combines reference to facts already established, including a comparison of my action to accepted standards of conduct, with an appeal to emotion, and the appeal can be distinguished from the assertion, however much it depends on it.[15]

A fourth objection is that symptoms and other natural signs have no particular place in this table. A possible answer is that they are rightly distributed among different lines. The sight of a fox may be to a human a logical seme, as its immediate interpretant is a thought, while to a rabbit it is an energetic seme, as its immediate interpretant is an action. It may be that certain natural signs are sufficiently complex that they are phemes rather than semes (a fox scent carried by wind from a definite direction, indicating location of the fox). Whether their mere occurrence counts as urging is another question; it may be that urging requires an utterer and a purpose in utterance and is therefore limited to the replication of legisigns.

This sort of investigation has to be carried much further, and to include a table of six trichotomies and twenty-eight lines, before we can begin to be satisfied that we are on the right track. But we are already well beyond where Peirce, so far as can be told from the remaining record, left off.

[15] The emotivist theorists in ethics have exercised their subtlety on such cases, and we can rely on their work here. But we are not thereby committed to emotivism. For Peirce's semeiotic suggests, rather, that emotions interpret the facts that produce them as being signs of qualities, such as good and evil; and if certain facts are reliably so interpreted, then that testifies to the reality of the corresponding qualities – no less (and no more) than interpersonal agreement about what is green or blue testifies to the reality of those qualities. See Short 2002b.

And as further effort here will not contribute to the purposes of later chapters, we must drop the matter. Life is not all play.

4. Ten Trichotomies

The previously cited draft letter of December 24–8, 1908, to Lady Welby, and the letter actually sent, unconventionally written before the draft, if its date of December 23 can be trusted, is the source for Peirce's declaration that there are ten trichotomies that yield not 59,049 classes but only 66 (EP2:481). That is to say, the combinations number $\frac{1}{2}(n+1)(n+2)$ instead of 3^n, where n=10. Little explanation is given in this letter and draft for the abrupt changes of terminology, definition, and order, not only from the earlier taxonomies of 1903, 1904, and 1906, but even from the letter to the draft (or vice versa) and within the draft itself. The discussion is patchy: much is in the form of an outline to which little meaning can be attached, while a few matters are chewed over at length. For example, immediate, dynamic, and final interpretants are renamed (in the letter only) 'Destinate', 'Effective', and 'Explicit', but with no indication as to which is which. Sparing the reader much tedium, I shall summarize Peirce's discussion in those pages (EP2:478–91), without reproducing the reasoning, from slightest slivers of textual evidence, by which I arrived at some of its parts.

In the letter, Peirce listed ten trichotomies in the order by which their respective bases supposedly 'determine' each the next. Translated into the terminology to which we have become accustomed, and labeled for later purposes by upper-case letters, these bases are, in order: objects (A) dynamic, (B) immediate; (C) the sign itself; interpretants (D) immediate, (E) dynamic, (F) final; and relations of the sign (G) to its dynamic object, (H) to its immediate interpretant, (I) to its dynamic interpretant, and (J) to its final interpretant. As before, the relation of a sign to its dynamic object is the basis of the icon/index/symbol trichotomy, while the relation of a sign to its final interpretant grounds the seme/pheme/delome trichotomy, and the sign itself gives us the qualisign/sinsign/legisign trichotomy, here renamed, 'tone', 'token', and 'type'. Trichotomies (A) and (B) are unhelpfully labeled 'abstractive', 'concretive', 'collective', and 'descriptive', 'designative', 'copulant', respectively. Nothing is said about trichotomies (D), (E), and (F). (H), the relation of a sign to its immediate interpretant, gives us 'suggestives', 'imperatives' (including interrogatives, Peirce added), and 'indicatives', supplanting the 1904 trichotomy, on the same basis, of emotional, energetic, and logical signs.

('Imperative' and 'indicative' seem too narrow, but otherwise we can see the relation between new terms and old.) And (G), the relation of a sign to its dynamic interpretant, gives us 'Signs assuring their Interpretants by Instinct[,] Experience[, and] Form', which supplants the 1904 trichotomy, on the same basis, of presentations, urgings, and submissions. (Again, 'assurance by instinct' seems too narrow, but we can see how experience – the sign-utterer's, not the interpreter's – is the warrant for urging, while the form of what is submitted is that which addresses the interpreter's reason.)

Applying this sequence to the six trichotomies of 1904, as they were numbered in the preceding section (which followed the sequence of Peirce's presentation of them), we obtain this reordering: 3, 1, 2, 6, 5, 4. That contradicts our hypothesis (in section 3) that 4 occurs between 6 and 5 (whether in the order 6, 4, 5, by relative simplicity, or 5, 4, 6, by its converse). But Peirce did not in this letter provide any discussion that substantiates the ordering given in it, other than by the general principle (quoted in section 1) that a possible can determine nothing but a possible, and so on. And a quite different order (rather, two or three different orders) is proposed in the related draft letter, as follows.

In the draft, an older terminology is used (except that tones, tokens, and types become 'potisigns', 'actisigns', and 'famisigns') and the trichotomies are differently described (e.g., the basis that in the letter is the sign 'itself' is, in the draft, 'the Mode of Apprehension of the Sign itself' and that which was the immediate object *simpliciter* is now 'the Mode of Presentation of the Immediate Object'). More importantly, some trichotomous divisions of signs are shifted from one basis of division to another, and one old basis (a sign's relation to its immediate interpretant) is silently omitted, while a new one (the triadic relation of sign to its dynamic object and normal, i.e., final, interpretant) is added. The order given is in many ways different:

> The ten respects according to which the chief divisions of signs are determined are as follows: first, according to the Mode of Apprehension of the Sign itself; second, according to the Mode of Presentation of the Immediate Object; third, according to the Mode of Being of the Dynamical Object, fourth, according to the Relation of the Sign to its Dynamic Object, fifth, according to the Mode of Presentation of the Immediate Interpretant, sixth, according to the Mode of Being of the Dynamical Interpretant, seventh, according to the Relation of the Sign to the Dynamical Interpretant, eighth, according to the Nature of the Normal Interpretant, ninth, according to the relation of the Sign to the Normal Interpretant, tenth, according to the Triadic Relation of the Sign to the Dynamical Object and to its Normal Interpretant. (EP2:482–3, punctuation as in original)

The bases of division, A through J, of the letter are thus reordered: C, B, A, G, D, E, I, F, J (H is omitted). So also, the bases of division in 1904 are now reordered: 1, 3, 2, 5, 4 (6 is omitted).

However, as trichotomies of signs are shifted from one basis to another, their reordering is not identical to the reordering of their bases. In the letter, a sign's relation to its immediate interpretant grounded the division of signs into suggestives, imperatives (including interrogatives), and indicatives. That looks very much like the draft's 'Ejaculative, or merely giving utterance to feeling; Imperative, including of course Interrogative; Significative'. The draft initially assigned that trichotomy to the fifth basis (not a sign's relation to its immediate interpretant, but the immediate interpretant itself, i.e., 'the Mode of Presentation of the Immediate Interpretant'), but then transferred it to a quite different basis:

> V. As to the nature of the Immediate (or Felt?) Interpretant, a sign may be: Ejaculative, or merely giving utterance to feeling; Imperative, including of course Interrogative; Significative.
>
> But later I made this the 7th Trichotomy and for the fifth substituted – with great hesitation –: Hypothetic, Categorical, Relative. (EP2:489, but as edited in 8.369)

These are astounding changes. The differences between the ejaculative and the hypothetic, the imperative and the categorical, and the significative and the relative are fundamental. There is also a fundamental difference between the immediate interpretant and a sign's relation to its dynamic interpretant (the basis of the drafts's seventh trichotomy).

In addition, the hypothetic/categorical/relative division (reminiscent in language but surely not in conception to one of Kant's four divisions of judgment) here makes its appearance into Peirce's taxonomy of signs for the first time. The trichotomy that was formerly based on a sign's relation to its dynamic interpretant – assurance by instinct, experience, or form – is now moved, word for word, to the new, tenth basis, that relates three items rather than two.

If we designate the letter's trichotomous divisions, rather than their bases, A through J, their reordering by the draft is different from that which we previously saw, and it suffers a different omission: C, B, A, G, E, H, F, J, I (the unspecified trichotomy, D is omitted). This is on the assumption that the unspecified trichotomies of E and F are those specified in the draft and assigned to the same bases.[16] By the same token,

[16] The draft states, 'As to the Nature of the Dynamical Interpretant[:] Sympathetic or Congruentive[,] Shocking or Percussive[, and] Usual'. And for the trichotomy the draft

the trichotomies of 1904 may be seen as reordered yet again, with, as it turns out, no omissions: 1,3,2,6,4,5. (4 now appears between 6 and 5, but that hardly warrants our hypothesis.)

Trichotomies spring over their bases whilst the latter nimbly shift position. How is this chaotic game of leap-frog possible? We have to conclude that Peirce was very far from having achieved the level of rigor that his intimations of system promise. But I think that there are problems with his taxonomy more serious than that.

The teleological structure of semeiosis yields two objects and three interpretants, each of which may be divided triadically by phaneroscopic category, as may the relations among them. Much of this correctly grounds taxonomic division. However, not all of these items are germane, without redundancy, to sign taxonomy, contrary to what Peirce appears to have assumed. For the two objects are intimately related, as are the three interpretants, and, thus, much that Peirce said about the one applies to the other or to the other two. For example, signs may be divided on the basis of how they tend to produce dynamic interpretants; but that is a property of signs also manifest in their immediate interpretants, which are their respective potentialities for being interpreted (i.e., dynamically), and in their final interpretants, which are what their dynamic interpretants would ideally be (chapter 7, section 2). Similarly, when Peirce, toward the end of the draft, substituted the rubric, 'According to the Purpose of the Eventual Interpretant' for the earlier 'Nature of the Normal Interpretant', he unintentionally revealed the irrelevance of the category of the final (normal, eventual) interpretant altogether to this particular division. For all three interpretants, immediate, dynamic, and final, are relative to interpretative purpose, that of the immediate being contained within those of diverse final interpretants (chapter 7, section 2).

Peirce defined legitimate divisions – of the purpose for which a sign is interpretable, and of the nature of a sign's influence – but wrongly attributed those divisions to the final interpretant alone or to the sign's relation to the final interpretant alone. The divisions are nonetheless of the greatest importance. The latter is our old, familiar one of seme, pheme, and delome; the other – dividing signs into those in which we seek either gratification (e.g., works of art), or immediate practical import,

had earlier listed as 'the Nature of the Normal Interpretant', it later states, 'According to the Purpose of the Eventual Interpretant[:] 'Gratific'[,] 'To produce action'[, and] 'To produce self-control'.

or guidance in forming general ideas and other habits of conduct[17] – seems to me pregnant with as yet unexamined significance. But other divisions are less persuasive, perhaps because their bases are questionable. For example, there should be no division based on what dynamic interpretants are. For a response, R, to X is a dynamic interpretant of some sign, S, that X is, if and only if R actualizes S's immediate interpretant (chapter 7, section 2). Beyond that actualization, R may get S wrong, for example, by misrepresenting S's dynamic object. Ergo, R's nature, beyond that of the immediate interpretant it actualizes, does not make S to be the sign it is. If it did, then error in interpretation would be impossible. Similarly, to divide signs on the basis of what their dynamic objects are, apart from how they are represented to be in those signs' immediate objects (chapter 7, section 3), is wrong, assuming that we do not wish to make being true or false, accurate or inaccurate, and so on, into sign taxa.

5. Where We Are Now

For all the enthusiasm that Peirce's later taxonomy has elicited, with its promise of a vast system, an endlessly ramifying formal structure that applies everywhere and to everything, close examination of it disappoints. It is sketchy, tentative, and, as best I can make out, incoherent. Its importance lies not in what it contains but in the kind of project it defines. That project has not yet been adopted by any of Peirce's devotees. For it does not consist in formal elaboration of principles presumed to be apodeictic. Rather, it consists in a critical examination of proposed principles, in part by painstaking application of them to particular cases, and in their arduous reformulation, until a coherent and illuminating system is achieved. The initial three trichotomies that Peirce presented are deeply revealing, in fact essential: note, for example, the prevalence, in contemporary philosophy, of the type/token distinction and the idea of indexical reference. The need for further divisions is equally compelling. What is not compelling is the way Peirce attempted the latter within the formal framework he projected.

Having said that, it is only fair to note that Peirce did not fare worse than have later philosophers working the same territory. Austin's 1955 William James lectures at Harvard, with its rather Peircean title, *How*

[17] These are my glosses on Peirce's 'Gratific', 'To produce action', and 'To produce self-control'.

to Do Things with Words (Austin 1962a), dissolves into long lists of types of utterance classed as verdictives, exercitives, commissives, behabitives, and expositives (Lecture XII). These rebarbative (the characterization is Austin's own) neologisms make Peirce's similar production look pale and timid by comparison. And while Austin's lists are more comprehensive than Peirce's, they suffer from a lack of order and systematic connection. 'Command', for example, occurs in no particular place in a list of 52 exercitives, along with such terms as 'excommunicate', 'choose', 'claim', and 'urge', while 'reckon' is one of twenty-nine verdictives, and 'state' and 'ask' are, along with a great many other terms, several with question marks attached, expositives. Why are 'claim', 'reckon', and 'state' placed so far apart and what relation might there be between an exercitive such as 'claim' and an expositive such as 'state'?

Imperfect and incomplete as Peirce's taxonomy is, it at least suggests some possible order and some possibly illuminating connections among the items ordered. In addition, it is not limited to linguistic phenomena, and, so, it is that much more ambitious in scope and that much more revealing, potentially. Perhaps most important, it is based, however shakily, on a general theory that provides the kind of guidance Austin needed and lacked.

Nelson Goodman's 1968 book, *Languages of Art*, does extend its reach to nonverbal signs ('symbol' is Goodman's generic term). It is more successful than Austin's work but perhaps only because it does not attempt a taxonomy, despite its claim to be 'an approach to a general theory of symbols' (p. xi). Rather, it enumerates a variety of symbol systems without arranging them in any order. Connection among them is made only in one respect, at the book's conclusion, when Goodman argues that art and science have it in common that 'Truth and its aesthetic counterpart amount to appropriateness under different names' (p. 264). Appropriateness, however, presupposes purpose, a concept Goodman, in his nominalist moments, would have had difficulty accommodating. Peirce, not subject to the same embarrassment, made a like connection between art and science when he placed feeling and thought in the same teleological structure, as two kinds of interpretant, emotional or logical; for then each is equally subject to evaluation (chapter 7, section 5; chapter 12, section 8).[18]

[18] It would be a useful exercise to test Peirce's taxonomy against Goodman's sharp observations. Some have supposed that the latter has demolished, or at least supplanted, the former. But that is too quick a conclusion. See chapter 8, section 2, for one example.

The 'general theory of signs', sketched by Ruth Garrett Millikan in her 1984 book, includes some divisions of signs. We discuss it below (chapter 11, section 4), only to discover that it, too, suffers by comparison to Peirce's taxonomy. For other comparisons we have to go back further in time, to Charles Morris, Suzanne Langer, and Ernst Cassirer.

10

How Symbols Grow

Peirce spoke of meaning in a variety of ways, once proposing to make it 'a technical term of logic' defined as 'the intended interpretant of a symbol' (5.175), another time enumerating the diverse meanings 'meaning' has in ordinary usage, including that of purpose (what one 'means' to do), suggesting their interconnection (EP2:256). The pragmatic maxim for clarifying meaning (W3:266, cf. chapter 2, section 10) is limited to 'intellectual signs'. In general, it seems best to declare that Peirce gave the term 'meaning' no special place in his semeiotic and that his theory of the interpretant, in its various divisions, is his technical counterpart to the tangled uses 'meaning' has in ordinary language. (At least once, he identified a sign's meaning with its immediate object (EP2:274); but the immediate object is a function of the immediate interpretant (chapter 7, section 3).) Similar remarks apply to the term 'reference': often used and sometimes explicated, it is not a technical term of semeiotic. Roughly, a sign refers to its dynamic object, and its meaning – as distinct from what it means to someone or what someone means by it – is its immediate interpretant.

Contemporary philosophy of language has been preoccupied with meaning and reference; that is no less true when the idea of meaning is held suspect and attempts are made to get by with reference alone. Peirce's semeiotic should be tested against the arguments of Quine and others. Doctrines associated with the later Wittgenstein, that meaning is use and that a private language is impossible, appear Peircean and invite another kind of critical comparison. Yet apart from comparisons of pragmatism to verificationist theories of meaning (the latest and most thorough is Misak 1995, ch. 3), there have been few such studies. We

have room here for one only, chosen because it bears on issues addressed in the next two chapters. As it happens, it also clarifies pragmatism's meaning.

It has not been widely recognized that Peirce anticipated, in important respects, the idea of rigid designation and related theories of reference due to Donellan, Kripke, Putnam, and others. Peirce came to it by a different route and drew different conclusions. On the topic of natural kinds, however, his view is comparable to Kripke's and Putnam's, though, I shall argue, it is more nuanced and consequently more useful than is theirs. Briefly: Peirce dissociated meaning from reference less radically than they did, and therefore his analysis enables us to make better sense of theory change than Putnam and others have made.

1. Hypostatic Abstraction

In 1861, at age twenty-two, Peirce wrote, 'I believe in mooring our words by certain applications and letting them change their meaning as our conceptions of the things to which we have applied them progress' (W1:58). This presupposes that a term's reference can be fixed independently, or to some degree independently, of its meaning, which is assumed in this passage to be conceptual. That overturns the familiar view that a term's reference is determined by its meaning, that is, that it refers to that of which a concept is true. The same passage entails that the meaning of a term will change or can be changed with the growth of our knowledge of the world. And that undercuts definitions of philosophy as 'conceptual analysis', that is, as usefully employed in explicating received meanings. The point is not to understand our meanings but to change them.

Peirce never abandoned this view. It was expressed differently in 1868, when he wrote that a thought-sign's reference is determined by the preceding signs it interprets, while its meaning is to be found in the succeeding thoughts that interpret it (chapter 2, section 3). Otherwise, it was expressed off-handedly, from time to time, until, in later years, it blossomed into his doctrine that symbols grow. Meanwhile, there had to be some ancillary developments. One was the idea of indexical signification, as depending on causal or other existential relations rather than on thought or general precepts or habits of interpretation. Only so could the reference of a term be fixed independently, to a degree, of what it means. Another was the idea of hypostatic abstraction. By hypostatic abstraction, an entity can be introduced into discourse independently of

direct characterization of it. Only so could we have some idea of what we are referring to, independently of knowing what it is.

We are familiar with indexical signification. Let us turn, then, to hypostatic abstraction. The earliest intimation of this idea that I have found is in 1887–8, in 'A Guess at the Riddle'. That is the unfinished book, or sketch of a book, wherein Peirce's new system of categories is first suggested and a need for phaneroscopy is implied but not yet recognized. A section on creativity in art and in science asks what the 'genius of mind' contributes to the data with which it is given to work. What does the mind 'add' that, even if a fiction, as in art, nonetheless brings out 'affinities' altogether real but otherwise indiscernible? With respect to science, Peirce referred to 'the realistic hypostatization of relations'. 'That', he added, 'is the one sole method of valuable thought. Very shallow is the prevalent notion that this is something to be avoided' (W6:187).

Later, hypostatization is featured in connection with a distinction Peirce drew between theorematic and corollarial reasoning. In a failed application of 1902 to the Carnegie Foundation, for support for a multivolume work in 'minute logic', he wrote,

> My first real discovery about mathematical procedure was that there are two kinds of necessary reasoning, which I call the Corollarial and the Theorematic, because the corollaries affixed to the propositions of Euclid are usually arguments of one kind, while the more important theorems are the other. The peculiarity of theorematic reasoning is that it considers something not implied at all in the conceptions so far gained. . . . Euclid, for example, will add lines to his diagram which are not at all required or suggested by any previous proposition, and which the conclusion he reaches by this means says nothing about. I show that no considerable advance can be made in thought of any kind without theorematic reasoning.
>
> [I]t is the operation of *abstraction*, in the proper sense of the term, which, for example, converts the proposition "Opium puts people to sleep" into "Opium has a dormitive virtue." This turns out to be so essential to the greater strides of mathematical demonstration that it is proper to divide all of theorematic reasoning into the Non-abstractional and the abstractional. (NEMiv:49)

Jaakko Hintikka (1983) has pointed out that, among geometers, the corollarial/theorematic distinction was a commonplace in Peirce's day and that his originality lay in extending it to all of deductive reasoning. 'What makes a deduction theorematic according to Peirce is that in it we must envisage other individuals than those needed to instantiate the premise of the argument'. In respect to a formal system, the distinction can be precisely formulated. Referring to Quine's system (Quine 1951), in which all variables are bound, Hintikka writes that 'a valid deductive

step is theorematic if it increases the number of layers of quantifiers in the proposition in question'.[1]

Hintikka does not comment on Peirce's mention of abstraction. However, the deductive introduction of a new layer of quantifiers, binding variables referring to a new domain of objects, is precisely what Quine refers to as 'abstraction' (1951, §24). Consider an example. From the premiss, 'Everything red is round', we may conclude, trivially but deductively, that 'Red things are a subset of round things'. From $\forall x(Px \rightarrow Qx)$, we may infer $\hat{x}Px \subseteq \hat{x}Qx$, where the circumflex notation represents the abstraction of a set from facts about individuals: $\hat{x}Px$ is the set of all and only those items, x (in a given domain), of which Px is true. That is not theorematic reasoning, but it does exemplify the step crucial to theorematic reasoning, of introducing new entities, in this case, sets. For we may then reason about these entities, coming to conclusions that could not have been attained otherwise (e.g., sets alone have cardinal numbers). It is an ascent from first-order to second-order predicate logic. Such ascents are not limited, as in Quine's system, to the introduction of sets (and sets of sets, etc.); in a different ascent to second-order predicate logic we might have quantification over predicates.[2] In English, we go from 'The ball is red' to 'The ball has the property, redness', and, once having made that and similar steps, we may then talk about properties, making them new subjects of our discourse. How is color like pitch? What is the smallest set of properties to which all topological properties may be reduced? And so on.

In short, there is nothing difficult about this idea of abstraction, nor is it unfamiliar to students of modern logic. Students of Peirce have tended to overlook it (Roberts 1973 being an important exception), perhaps because mesmerized by his early discussion, in 'The New List', of the type of abstraction named 'prescision' (chapter 3, section 3). But in the meantime he had made his 'first real discovery about mathematical procedure', and thus in 1903 he wrote that the term 'abstraction' should not be used for prescision, as 'it is indispensable for another purpose, that of designating the passage from "good" to "goodness" and the like' (EP2:270n). Because abstraction introduces an object, Peirce often qualified it as

[1] In Hintikka 1973, the corollarial/theorematic distinction, as between trivial and nontrivial deduction, is developed in detail. In the 1983 article cited here, Hintikka joins the roster of distinguished philosophers who have found something of value in Peirce, once having thought of it themselves.

[2] Peirce's own formalization of abstraction was graphical, presented in his Lowell Lectures of 1903, in what he called the Gamma part of his system of existential graphs (4.510–29); cf. Roberts 1973, ch. 5.

'hypostatic'. The qualification testifies to a lingering tendency to refer to prescision as 'prescis(s)ive abstraction'.

Peirce said that hypostatic abstraction 'may be called the principle engine of mathematical thought' (2.364, cf. 4.234 and 5.534). 'It is by abstraction that a mathematician conceives the particle as occupying a *point*. The mere place is now made a subject of thought'; again, 'When the mathematician regards an operation as itself the subject of operations, he is using abstraction' (NEMiv:11). Hypostatic abstraction in mathematics has two features of note. (a) It is 'a necessary inference whose conclusion refers to a subject not referred to in the premiss' (4.463). The inference is necessary in the sense that, necessarily, the conclusion is true if the premise is true; in other words, it is apodeictic. (b) '[T]he new individual spoken of is an *ens rationis*; that is, its being consists in some other fact' (4.463, cf. NEMiv:11, 162 and 6.382). Thus, there would be no such thing as height if there were no truths about how tall something is (but how, exactly, the former 'consists' in the latter, Peirce did not say). (b) explains (a): as the new entity 'consists' in the given fact, then, necessarily, the conclusion is true if the premiss is.

These two conditions are not always satisfied by hypostatic abstraction outside mathematics. Even when speaking of mathematics, Peirce's favorite example of hypostatic abstraction was not mathematical; it was Molière's burlesque of Scholasticism (the passage quoted above, from NEMiv:49, is one of many such). When asked why opium puts people to sleep, the medical degree candidate responds, to applause, that it has a dormitive virtue. Peirce remarked that even in this, 'the operation of hypostatic abstraction is not quite utterly futile. For it does say that there is *some* peculiarity in the opium to which sleep must be due' (5.534). But saying that there is a cause, and a cause located in the opium, of why people who take opium have tended to become sleepy, is to go beyond the latter fact. It is to say that the sleepiness was not coincidental, though its being coincidental remains a logical possibility. Hence, the inference is not apodeictic. In addition, if opium does possess a dormitive virtue, that virtue must be a power or a constituent of the opium, the being of which does not consist in that effect that it is adduced to explain. In these two respects, such inferences differ from hypostatic abstraction in mathematics.

In empirical science, abstraction introduces an entity that is not itself abstract, necessarily. It may be a particular event or a physical substance or an individual thing. But the entity is introduced abstractly, by being defined, not in terms of properties of its own but in terms of its presumed

relation to something else, presumed to be actual. Thus, it is conceived of abstractly. For example, opium's dormitive virtue is defined by its presumed effect, the going-to-sleep of opium takers. Hypostatic abstractions of that sort are a limiting case of the mode of inference Peirce elsewhere named 'abduction', the introduction of an explanatory hypothesis.[3] They differ from abduction properly so called by not providing any explanation. Molière ridiculed Scholasticism because it was content with a non-explanatory explanation. Peirce's point was that such an abstraction is nevertheless the beginning, or could be, of an inquiry aimed at explanation. Without supposing a cause, no search for it can be made. The dormitive virtue is *that which* in opium puts people to sleep. The words 'that which' are a placeholder for further characterization. Empirical inquiry is needed to fill in the blank.

There is, in addition, another type of hypostatic abstraction found in the special sciences, ubiquitously, that is closer in nature to the purely mathematical type, as it introduces *entia rationis*. A body's center of gravity is a geometrical abstraction whose reality consists in other facts concerning where lines of gravitational force intersect (but lines of force are another abstraction!). Specific gravity is a ratio of two measures of weight, weight itself being an abstraction that depends on attraction between two bodies. Voltage consists in there being a difference in electrical potentials in two parts of a circuit, and a potential is again an *ens rationis* consisting in a fact about the electrons in a group of atoms. And so on. So also in biology, economics, and other sciences.

Often, a hypostatic abstraction of this type fills in the blank that was introduced by a hypostatic abstraction of the Scholastic type (as we will call it). Take temperature, for example. We had thermometers before we had a good idea of what temperature is. At that time, we could define temperature as *that which* is measured by a thermometer. But what is it? We did not know – other than that it is the same thing that is felt and that is causally related to a variety of phenomena. It might have been a kind of stuff. Now we know that it is an average per molecule of random kinetic energies. But that is an abstraction from a different set of facts, facts not about instrument readings but presumed facts about invisible motions.

[3] Peirce examined this form of inference in many papers over a long period, sometimes calling it 'hypothesis', sometimes 'retroduction', with significant changes over time. But this is one of the most examined aspects of Peirce's thought. See, e.g., Hookway 1985, pp. 222–8, for brief discussion and citation of sources.

These same examples teach us to be cautious about denying reality to an *ens rationis*. For all of the quantities mentioned – voltage, temperature, and so on – are consequential: all explain a range of effects. And each exists or obtains at a particular place and time (however vaguely these may be defined). Thus they have a physical reality, even if that reality 'consists' in the reality of certain other facts.

Some *entia rationis* have only the being of a mathematical abstraction; others have physical reality; and some entities introduced by hypostatic abstraction are not *entia rationis* at all, for example, the alkaloid that is opium's dormitive virtue. Corresponding to those distinctions within condition (b) of hypostatic abstraction, there is a corresponding division of condition (a). If the inference is from true premisses and apodeictic, the introduced entity is an *ens rationis*, and it will have the same mode of being as those entities represented in the premisses: ideal in the case of pure mathematics, real in the case of empirical science. If the inference is not apodeictic but is abductive, then the entity introduced may fail to be real at all (even though the premisses be true), but, if it is real, it will be as real as the effects in terms of which it is introduced, whether it is an *ens rationis* or not.

For reasons that will appear later, we shall focus on what may seem to be the least interesting type of hypostatic abstraction, the Scholastic one, hereafter SHA. An SHA introduces a logically contingent assumption that is two-fold: it asserts an existential relation to something actual (or presumed to be actual) and it asserts the uniqueness of that relationship (i.e., that only one thing stands in that relation to the given entity). An existential relation is one all relata of which must be actual particulars. Take an even simpler example of an SHA than Molière's: one not Scholastic in the historical sense but that fits our definition of SHAs. We see John lying in an attitude suggestive of violent death, and we ask, 'Who is the murderer?'. That presupposes that John in fact died a violent death, that a person is responsible, and that only one person is responsible. If no one, or if more than one, murdered John, then the SHA errs and the entity it introduces – *the* murderer – is a fiction. So also if it turns out that John is not dead. Before it is recognized as being a fiction, the entity an SHA posits remains a subject about which debate may rage, as the nature of caloric was once a topic of debate.

A fiction, notice, is not the same as a possibility that happens not to be actualized. If John is alive, Susan's having murdered him is an unactualized possibility; it is possible that she will murder him yet; but it is not possible that she was *the* murderer whom we rashly hypothesized.

As that murderer is a fiction, no one could possibly have been it. Similarly, if there should evolve a species of white horse sporting a single horn and having a tender affinity for maidens, that species would still not be of unicorns (though they will probably be so named). For the unicorn is a myth and therefore not of a species anything actual could belong to.

The reason an SHA, if it fails to denote something actual, introduces a fiction rather than a possibility is that what it purports to denote is defined by its presumed existential relation to something presumed to be actual. If those conditions do not obtain in fact, then nothing can be this entity. If the entity is a kind, such as a chemical substance or a species, then nothing can be an instance of it. An SHA, then, differs from a concept that is defined merely by a set of general properties. The latter, such as 'red rubber ball a thousand miles in diameter', can represent a possibility even if there never was such a thing actually. But an SHA represents either something actual or a fiction, and never a mere possibility.

2. The Hiddenness of Abstraction

Abstractions of the Scholastic type are more often implicit than explicit, and implicit in ways hard to discern. They are essential to the fabric of that discourse in which they are hidden. They are the warp, the long strong threads, over and under which the more visible woof is woven.

Take, for example, the idea of heat. Heat is what, in the early history of thermodynamics, caloricists and kineticists disagreed about. According to the one, heat is a substance, possibly weightless, named 'caloric'; according to the other, heat is motion, eventually asserted to be molecular. The two conceptions are very different. What, then, could be the concept of heat that caloricists and kineticists shared? What is common to stuff and motion? And if the two theorists did not share a concept of heat, then what is it that they thought they were disagreeing about? What is it, indeed, that *we* think that they disagreed about? Can we say what heat is, without implying that it is a kind of stuff or a kind of motion or some third kind incompatible with either of those?

Well, they had words in common – 'heat', *chaleur*, and so on – but what is the idea that those words supposedly express? We turn to ordinary usage, as the source of the idea as well as the terminology that is shared by opposed theorists, and we are baffled. For in ordinary usage, 'heat' and its translations refer to – what? (How odd that we can translate the word into other languages more readily than we can say what it means.)

From the point of view of caloricists and kineticists alike, ordinary usage (in any of various languages) conflates distinct items, namely, heat and temperature. The amount of heat we can get out of a log by burning it is something other than the heat (= temperature) we feel when standing in front of the fire. There is more heat but at lower temperature in a large, cold lake than there is in a small cup of hot water. Theorists in either camp agreed on an important point, that heat is one thing and the ratio of that thing to matter is another, variously called 'degrees of heat' or 'heats' (Newton in 1701) or 'temperature'.[4] But temperature was thus the ratio of what to what? Is it a ratio of stuff or of motion to matter? And matter in what sense, as volume or as mass or as number of parts (later, molecules)? (How odd that the caloricists and kineticists could share so fundamental a departure from common sense, while sharing no language in which to state it – no language, that is, except those words that they understood in incompatible ways.)

We have already supposed that temperature was initially conceived by an SHA from the use of thermometers. 'Heat' in scientific usage, we can understand, similarly, as referring, initially, to an SHA from a variety of phenomena presumed to be related. Heat is that, whatever it is, which, by being in diverse ratios to matter, in some undecided sense of 'matter', produces different degrees of temperature. It is thus an SHA from temperature, which is an SHA from the use of thermometers. But heat is also that which is transferred from a body of higher temperature to a body of lower temperature with which it is in contact, until the two are equal in temperature (this comprehends a great variety of phenomena). And heat is also that which is produced by combustion (though this forces another distinction, between latent heat and nonlatent: for in combustion, energy is not so much produced as released). These various definitions can be combined, but only on the logically contingent assumption that it is the same thing (stuff, motion, or whatever) that plays each of these roles.

But no one (so far as I know) defined 'heat' in that way. Rather, one party said that heat is a weightless substance that flows, while the other said that it is a quantity of motion of small parts that can be transferred to adjacent parts or otherwise (in gases and liquids) become differently

[4] The word 'temperature' did not, in English, acquire this meaning until 1670, after the thermometer had been invented (roughly, 1592–1669). The word existed earlier, but with other meanings, derived, through the French, from the Latin for moderation, balance, proportion, etc. (thus 'temper', 'tempered', 'temperate'). The word 'thermometer' is etymologically unrelated, having been deliberately coined (1624) from the Greek *thermos*, meaning heat in the ordinary (confused?) sense.

distributed. The hypostatically abstractive concept of heat that they shared, and that made their communication possible, is not one that they formulated. It is left up to the historian to formulate it, in order to explain how these parties could have supposed they were talking about the same thing while disagreeing about its nature. That abstraction was implicit in their shared understanding that they disagreed.

That is one reason why abstraction is often hidden, namely, that people, scientists perhaps most of all, are impatient with it. They want to say, more directly, what things are. So we have conflicting theories, with no explicit account of what it is that the theorists disagree about. For that account must be, a fortiori, hypostatically abstract. And thus we tend to formulate the subject in dispute in a question-begging way, in terms of the theory that we accept. We believe that heat is motion, and so we say that the caloricists mistook heat – a quantity of motion – for a weightless stuff. But that makes it difficult to reconstruct intellectual history. The caloricists are made to look like such fools.

So far, we have examined SHAs implicit in the transition from common sense to theory. But common sense is also rife with abstraction. Its SHAs are hidden because common sense, rather than consisting in what ordinary people explicitly have in mind, is implicit in the structure of ordinary ways of talking. What do the unlearned understand by 'heat'? Philosophers, assuming concreteness of conception, have differed as to whether the reference is to an objective quality supposed immediately felt or to the sensation itself or to the presumed cause of the sensation, which it is presumed to be like. But you and I know that the cause of a sensation of warmth is quite unlike that sensation. Do the scientifically uneducated, then, have a mistaken concept of heat? That would fail to explain the success of people in communicating ('Watch out: the stove is hot!') and otherwise getting along in the world. Surely, they do not make the gross error of attributing their own sensations to the stove; they are referring to something objective, and know that they are. But they do not know what it is – and, usually, they have never thought to ask. Their successful use of words does not require that they have ideas – comparable to those of philosophers or scientists – about what they refer to. Yet they have ideas of a sort: they know that they refer to something. Heat is *that which* we can feel and fire produces and boils water. The idea is abstract (and, far from being confused, it merely does not make those distinctions that it would be a confusion not to make in a less abstract conception). Common sense is far more abstract than is any theory of philosophy or physics.

If so, from what is the abstraction made? We have already indicated that in the case of heat it is a variety of phenomena – fire, boiling, feeling – connected in our experience. But to get more readily to the root of the matter, take a simpler case, say, red or sweet. The concept of the red or the sweet cannot be formed by abstraction from the concept of the sensations thereof; for the latter are, rather, defined in terms of the former. A sweet taste is like that of candy. First we learn to call candy 'sweet', and only later to identify the sensation ('like that of candy') and to distinguish the sensation from the quality (Mother says, 'The sweet can taste bitter if you are ill'). Someone unlearned can say, 'Red is something you can see, sweet is something you can taste'; but so also blue can be seen and bitter tasted. What is the concept that that person has of red as distinct from blue? Perhaps we should say, simply, that he has no concept. He knows how to use the word successfully in practical contexts, and that is all. Why try to make a philosopher of him, and then berate him for being such a bad one?

But there is an identifying conception of redness implicit in ordinary usage, even if we do not want to go so far as to say that this is a conception ordinary speakers have in mind. Red is *that which* we have learned to call 'red' (or *rouge* or *rot*, etc.), primarily by looking. More exactly: that is red which a speaker not color blind would not honestly deny is red (in whatever language he speaks, if it has a word translatable by 'red'), were he to look at it in light not colored. (The seeming circularities of 'not color blind' and 'not colored' can be explicated without circularity in terms of the ways ordinary speakers check one another's attributions of color.) Similarly, the concept of sweet as distinct from bitter is parasitical on ordinary usage (yet is not limited to the language in which it is expressed). The abstraction is from that usage. The talk comes first, then the idea!

And how else could it be? The alternative, ridiculed by Wittgenstein, is to suppose that infants are stuffed full of ideas and that they need only to be taught words by which to express them. Our ordinary conceptions are not presupposed by speech but, rather, are derivative from verbal practices first learned without one's having any ideas at all. (Of course, by 'idea' I mean something such as could be given a verbal definition; I am not denying that the preverbal infant has images, sensory memories, inarticulate expectations and desires, and corresponding behavioral dispositions with which parents are only too familiar.) But that is possible only if speech, at a rudimentary level, is significant independently of speakers having ideas. And that is what Peirce's mature semeiotic explains (chapter 6). For it makes significance to depend not on thought but on

purposefulness, and the purposefulness of animal behavior, including that of very young human animals, is not conceptual (chapter 6, section 6).

3. A Very Virtuous Variety of Vagueness

'Vague' has many meanings, only one of which philosophers and logicians have paid much attention to. A concept is vague in the sense of being 'fuzzy' (a technical term of logic) if there exist borderline cases of which it is neither true nor false. (Being neither true nor false is something objective; it does not consist in one's not knowing which.) As we climb from the bottom of the valley up the slope of a mountain, we will pass through a region that is neither definitely in the valley nor definitely out of it. There is no further knowledge that could enable us to decide the question, 'Have we left the valley yet?' Ergo, being a valley is vague.

To lack specificity is another meaning of 'vague', and vagueness in that sense is distinct from fuzziness.[5] Inexactness is a quantitative form of lacking specificity.[6] We shall coin the term 'inspecificity' for any sort of vagueness qua lacking specificity. A statement is inspecific if it omits desired information, as when I ask you who you voted for and you reply, vaguely, 'Oh, someone'. Concepts also can be inspecific: SHAs are so. The idea of a dormitive virtue leaves us dissatisfied. We want to know what that virtue is, more specifically.

Before we possess that more specific knowledge, we may be unable to decide whether a sample of a sleep-inducing chemical is or is not an instance of opium's dormitive virtue. But that does not make the idea fuzzy. For that sample definitely is the same kind of thing that makes opium soporific or it is not; it is merely that we cannot as yet know which it is. An SHA may be fuzzy but, generally, it is much less so than our uncertainties in applying it might suggest.

An inspecific concept is consistent with each of many mutually opposed theories; hence, it is less open to doubt than they are. We can be wrong that opium has a dormitive virtue, but we are not as likely to be wrong as we are if we venture to say what that virtue is, for example, the work of

[5] Peirce recognized the existence and importance of this type of vagueness: 5.446–50. Cf. Short 1980b, 1988b; Hilpinen 1983, p. 269; and Lane 1997.

[6] 'About an inch' is not as exact as '1.02 inches'; also, it is fuzzy. '1±.03 inches' is not fuzzy at all, yet it is less exact than – not as specific as – '1.02 inches'. Therefore, inexactness and lack of specificity are not the same as fuzziness. Notice that inaccuracy is not the same as inexactness: the less exact our statement, the less likely it is to be inaccurate.

the Devil or an alkaloid or, still more specifically, $C_{17}H_{19}NO_3$. We can be right that it has such a virtue but wrong that that virtue is an alkaloid, or right that it is an alkaloid and wrong that it is $C_{17}H_{19}NO_3$. But we cannot be right that there is an alkaloid in opium that induces sleep and wrong that opium has a dormitive virtue. A specification of what was inspecific implies the truth of the latter, but not conversely.

It follows that observations presupposing an SHA must be acceptable equally to theorists opposed in their ways of attempting to make that SHA more specific. For example, a major piece of evidence in favor of the kinetic theory of heat was Count Rumford's observation that boring out a cannon produces heat (as measured by the capacity of the hot iron to boil water) as long as the boring continues, contrary to what the caloric theory predicts. The observation of heat being produced did not presuppose either theory, as it was in terms of the inspecific idea of heat as *that* (whatever it is) *which* boils water. But either theory is consistent with that more abstract idea of heat, as each is but a different way of making it more specific. Thus caloricists could not deny, without denying their own theory, the observation that boring out a canon produces heat and, indeed, more heat than their theory predicts. And, so, the caloric theory was refuted.

Common sense ideas import logically contingent assumptions of a general nature – that boiling is an effect that can be brought about in predictable ways, that our sensations of warmth, color, and so on are effects of causes external to one's body, that there is a spatially and temporally extended world in which we are located and that exists independently of what we think about it, and so on. Therefore, common sense is theory-like in being subject, in principle, to falsification (in principle only, because, were common sense false, other than at the margins, we would not survive long enough to discover that fact). However, it is unlike theories properly so called, as it lacks their hallmark trait of specificity. Furthermore, common sense ideas are, with occasional modifications, presupposed by scientific theories, which are so many attempts to make those ideas more specific. And that is why common sense is unlikely to be contradicted by the theories of modern science, though it may be refined (e.g., the distinction made between heat and temperature). These reflections are necessitated by arguments, such as those of Paul Feyerabend (1962, 1965), that common sense is a theory refuted by modern science.

Inspecific concepts are very virtuous because they are relatively secure, because they suffice for most practical purposes, and, finally, because they enable us to make observations by which to test more specific ideas

(theories). Inspecificity makes its own overcoming possible. All of this is part of the doctrine Peirce announced (1905) as 'critical commonsensism' (5.438–62, 497–537), within which he asserted that common sense beliefs, though they have the same kind of foundation as scientific theories, are relatively indubitable because vague, and are useful for that reason; the same pages contain one of his major expositions of hypostatic abstraction.

4. Abstraction and Rigid Designation

Peirce wrote,

> A philosophical distinction emerges gradually into consciousness; there is no moment in history before which it is altogether unrecognized, and after which it is perfectly luminous. (W2:71)

Peirce's anticipation of Kripke et al. did not develop in the context of issues in modal logic.[7] If we focus on the questions being addressed and the arguments made, then Peirce did not anticipate 'the new theory of reference', as it has come to be called. If, however, we focus on the answers, arrived at in other ways, then we find that some key parts of the new theory were stated by Peirce three-quarters of a century earlier.

Following Kripke's exposition:[8] a term, T, designates rigidly when uses of it refer to nothing under counterfactual assumptions that they do not refer to under factual assumptions (the assumptions pertain to the conditions of the referent, not those of the act of referring).

Take the definite description, 'the twenty-sixth president of the United States'. It refers to Theodore Roosevelt. One might say, 'Had McKinley not been assassinated, it is likely that the twenty-sixth president would not have been the twenty-sixth president'. Here, the first occurrence of 'the twenty-sixth president' designates rigidly, the second does not.

Are there any terms that normally, as a matter of course, designate rigidly? Proper names seem likely candidates (plural bearers of the same name is a complication with which we know how to deal; chapter 8, section 3). One might speculate, 'Theodore could have been named Sam, instead'. In that sentence, 'Theodore' is being used rigidly to refer

[7] That is despite the fact that Peirce had sketched a formal system (indeed, two) of modal logic (4.510–29, 552–72; cf. Roberts 1973, chs. 5 and 6).

[8] However, I avoid formulation in terms of Kripke's 'possible worlds' semantics; see chapter 3, section 9.

to the same person who was in fact so named, under the supposition that he was differently named.

The so-called traditional theory of reference, for example, as formulated by Frege, holds that a term refers to all and only that of which a concept it expresses – its sense or meaning – is true. In other words, it refers to that which it may be used correctly to describe. Applied to proper names, we obtain the view that names are disguised descriptions or, at least, that they are associated with a description or cluster of alternative descriptions assumed to apply uniquely. It would follow that, if the only thing I know about Theodore Roosevelt is that he was the twenty-sixth president, then there are possible conditions under which my use of 'Theodore Roosevelt' would refer to Mark Hanna. To block such an absurdity and explain the evident fact that names designate rigidly, Kripke et al. devised a causal theory of reference,[9] that in some cases, reference is fixed causally, rather than by satisfaction of truth conditions. A proper name refers via a chain of uses of it, extending back to an original use – for example, a baptism or an introduction – causally connected to its referent. That view was anticipated by Peirce, in his account of proper names as indexical legisigns (chapter 8, section 3).

We should note, however, that naming ceremonies, introductions, and the like are not purely causal; often, they are not causal in relevant respect at all. When the priest says, 'I name thee Susan', we know, as a presupposition, that it is a human baby that is being named and not, for example, the baby's cranium. The particular infant being named is the one whose cranium the priest at that moment dabs with sanctified water. That physical, spatio-temporal relationship picks out the individual at issue from all other individuals – which is something that the words spoken and the thoughts thought cannot do by themselves. However, to suppose that the priest's dabbing is caused by the baby dabbed stretches a point – the point that there must be some causal process extending from the baby to the priest's visual and tactile sensations, lest he miss the mark. Rather, as in Peirce's general account of indexicality, it is the existentiality, or *haecceity*, of the relationship of dabbing to dabbed that is crucial (that, and the equally existential simultaneity of that act with the words uttered). Existential relations are various and range from causation to juxtaposition; but all are spatio-temporal and obtain, if at all, between existing or occurring particulars. I am not suggesting that anyone who has espoused the

[9] Gareth Evans 'The Causal Theory of Names', reprinted in Schwartz 1977; Putnam 1975; pp. 198–202; and others.

causal theory of reference meant to deny any of this – only that exclusive emphasis on causation is misleading.

Kripke and Putnam have used these ideas to distinguish between two classes of general term. In the case of terms for individuals and substances, but not properties, we may, following Wiggins (1980), distinguish these as natural and artifactual. We know what kinds of things chairs are, although the term is 'open-textured', that is, can acquire new applications. 'Water', 'gold', 'tiger', 'panda bear', by contrast, refer to kinds of which we may know little or nothing; the descriptions we associate with these terms, by which we normally identify their exemplars, may turn out not to be true of some exemplars and/or to be true of some nonexemplars. We can learn to our surprise that panda bears are not bears at all but are related to raccoons. But how is it possible that a term that we use can refer to something other than we think it refers to?

Water, we now know, is more or less pure H_2O. Normally, we identify as water that which is clear, quenches thirst, boils over a fire, freezes in winter, expands when freezing, and so on. Before the development of chemical theory, those were the only ways water was identified. Suppose that there are people in a place who find, there, a clear liquid that is thirst-quenching, and so on, but that it is not H_2O. Suppose that, before discovering that water is H_2O, we found out about these people. We would naturally have concluded that they had water in that place, but we would be wrong. So also, a creature that looks and acts just like a tiger is not a tiger if its genetic history is not that of the animals we call 'tigers'; and that is so even if we are unable to reconstruct the genetic history of tigers and even if we are ignorant of the relevance of genetic history to the definition of biological taxa.

We see, then, that terms for natural kinds designate rigidly. Their reference is fixed by the actual, if unknown, nature of the exemplars in relation to which we have learned to use them. As with proper names, there is, in Putnam's words, an 'indexical component' in our use of natural kind terms. Water, he says, 'is stuff that bears a certain similarity relation to the water *around here*' (1975, p. 234, emphasis in original). It is this indexical component that fixes reference and not the descriptions, or 'stereotypes', as he calls them, by which we pick out exemplars (1975, chs. 11–13 passim). In a later essay, Putnam speaks of an 'indexical criterion', which, apparently, is the verbal formulation of an indexical component (1988, pp. 33–4); we can take the words just quoted as an example. Indexical criteria differ from 'qualitative criteria', Putnam says, by citing a particular example (e.g., the water around here).

Notice that Putnam's criterion for something's being water is a hypostatic abstraction: it posits an entity – water – on the assumption that the water 'around here' – that is, the stuff around here that we have been calling 'water' (the verbal practice comes first!) – is of a kind yet to be identified. It therefore depends on its users having a notion of a natural kind, however vague, which perhaps was not the case when 'water' first began to be used. If there was an earlier concept of natural kind, that concept was transformed, as far as substances are concerned, by the advent of chemical theory and then by subsequent changes in chemical theory. Be that as it may, at some point, 'water' came to refer to a hypostatic abstraction, that is, to a stuff, albeit existing, abstractly identified, about which we could ask, 'What is it, really?'.

Any SHA designates rigidly for the same reason an indexical criterion does. Hence, the terms that SHAs define, or may be used to define, are rigid designators. At least in some cases, depending on the 'similarity relation' or other relation assumed, what is or would be thus designated are natural kinds, if they exist at all. And for any natural kind term, we can formulate an SHA that explains its use or would explain its use if speakers could plausibly be said to have that SHA in mind, at least tacitly. This has already been illustrated by such examples as heat.

But then we have a conception, the indexical criterion or SHA, that is true of all and only instances of the kind designated – very much like the traditional theory of reference supposedly overthrown.[10] Have we arrived at a contradiction?

5. Incommensurability and Meaning's 'Location'

Putnam has been wont to say that 'meanings are not in the head'. For convenience, let us restrict attention to the essay, 'The Meaning of "Meaning"' (1975, ch. 12), already quoted, in which he sets out his view at length, and to a later, comparable essay (1988, ch. 2), also quoted above, in which there are some important changes. In these essays he attacks what we have been calling the traditional theory of reference, but his ultimate target is something else, a contemporary theory in the philosophy of mind, of

[10] Cf. Haack 1992, p. 34, which, however, goes too far in denying that Peirce would have been sympathetic to the assimilation of natural kind terms to proper names; they are like proper names in designating rigidly, and that is essential to Peirce's idea of how symbols grow, hence, to his pragmatism. Haack's article covers many of the same topics covered in this chapter and in chapter 12, and to the same general effect, but with, as on this point, important differences.

which Jerry Fodor (1975, 1994) is a major proponent. According to it, there are 'mental representations' that are wholly contained within individual minds or brains, much like events in the circuitry of a computer. Joining the two issues in this way is, I shall argue, a mistake.

In the 1975 essay, Putnam cites Frege and Carnap as having expressed a version of the traditional theory. They held concepts to be public, abstract entities, not inner psychological states or acts, but Putnam is undaunted: 'However, "grasping" these abstract entities was still an individual psychological act' (p. 218), and therefore the theory is that 'knowing the meaning of a term is just a matter of being in a certain psychological state' (p. 219). The word 'just' is where the trouble lies. In an essay of 1988, he writes, '[I]nstead of the word "concept" I shall use the currently popular term "mental representation," because the idea that concepts are just that – *representations in the mind* – is an essential part of the picture' (p. 19, emphasis in original) – the picture, that is, that he means to attack, or in other words, the doctrine that reference is fixed psychologically.

In both essays, the refutation is two-fold. First, we often refer to kinds that we ourselves do not know how to distinguish (gold and fool's gold, perhaps), relying on the fact that others can. Second, natural kind terms refer via an indexical relation. The social division of linguistic labor, as Putnam calls it, and the indexical relation, or 'contribution of the environment', both lie outside the individual psyche. And thus reference is not fixed psychologically in all cases.

In 1975, Putnam chose to identify meaning as that which determines extension (p. 270); as extension is not determined psychologically, it follows that 'meaning is not in the head'. This slogan survives a switch in 1988 to a different definition of meaning, as 'what we try to preserve in translation' (p. 29). These two definitions differ; for, in the case of natural kind terms, what fixes reference is the indexical component and that is not what translation preserves (p. 38), especially not if the indexical relation is to something local (e.g., to the stuff around here, or to what English speakers call 'water'). But what is preserved is the term's extension, which is indeed outside the head, whether of speaker or hearer.

Notice, however, that this is *not* a refutation of the traditional theory of reference, at least as formulated by Frege and Carnap. For they maintain that it is concepts that determine extension, and concepts are not in the head. The fact that they must be 'grasped' in individual psychological acts makes no difference to the fact that they, themselves, are not in

anyone's head. It is what is grasped, and not the grasping, that on Frege's and Carnap's theories fixes reference. What is grasped is something linguistic in nature, rules or practices, a creation of the language we share with others. As such, there is no reason why the social division of linguistic labor could not be built into it; in fact, it is. Whether or not Frege or Carnap ever considered the point, the same could be said of the indexical component. And, therefore, what is grasped psychologically cannot be reduced to the psychology, much less the neurophysiology, of the grasping. A concept is not *just* a mental representation. While Fodor may be refuted by Putnam's argument, Frege is not. At the very least, if Frege did think only of qualitative concepts, we can entertain an extension of the Fregean doctrine, that *Sinn* determines *Bedeutung*, to *Sinne* that are not purely qualitative.

Consider the variety of concepts that we ordinarily admit to having. We have purely qualitative concepts, but also concepts of artifacts, such as chairs. The latter are not qualitative but refer to the actual world of persons who produce and use such items. Reference to the actual is built in. If we admit that we have such concepts, then we cannot consistently deny that an SHA is a concept, on the ground that it refers to something actual. The only difference between the two in respect to actuality is that the entity an SHA represents is posited as actual – it is either actual or a fiction – whereas artifactual concepts, although they refer to the actual world, do not entail that the objects conceived of are actual.

There are obstacles to accepting the idea that concepts can incorporate reference to actualities. A concept, whatever else it is, is at least something that individuals can grasp and share; but Putnam has stressed that an indexical criterion refers to an actual case, and it is unlikely that anyone would – or could – have such a criterion in mind, much less that it is shared by all who may be said to grasp the same concept. For we do not usually remember the occasion on which we first learned a word, much less the occasion on which the word was first coined, if it ever was coined, rather than having evolved gradually from related uses of cognates. However, a criterion that refers to a broadly or vaguely limned class of actualities is also indexical; for it has to be interpreted by a choice of one or another index.[11] And such a criterion is one that can be had in mind

[11] Quantifiers are indexical in the same way on Peirce's account, viz., as 'precepts' for indexical signification of items in a domain: 2.330, 336; cf. Hilpinen 1983.

and widely shared. There is no need to limit indexical criteria the way Putnam has.

The criterion 'stuff that in our experience has been thirst-quenching, etc.' works, even though the extension of 'our' widens as we learn of others who seem like ourselves in relevant respects; these are actual others, not possible others. It makes the assumption that all such stuff, or nearly all, impurities aside, is the same; but that is the sort of defeasible assumption that has to be made in any case (Putnam's 'similarity relation'). Defeasible and also refinable: what counts as an impurity will change, for example, with the development of chemical theory. The idea of a natural kind is robust enough to survive discoveries – sometimes made in the process of acquiring theoretical knowledge – of differences that force distinctions, for example, between gold and fool's gold. Again, 'The stuff called "water"' works fine, and is not limited to English-speakers, as it can be translated into any language that has a word or phrase that translates 'water'. Of course, establishing the translation depends on determining that the stuff others call by another name is what we call 'water'; but that can be done tentatively, defeasibly even before discovering water's chemical definition.

Another obstacle is that we may find it difficult to understand how what is grasped transcends the grasping. But such transcendence – the irreducibility of reference to psychology – is what Putnam himself is arguing for. Peirce's semeiotic may be of some help here. Peirce said that a concept is a symbol (chapter 2, section 10; cf. EP2:10, quoted below, in section 6). More precisely: different symbols 'express' the same concept if their immediate logical interpretants are the same (they may differ in emotional interpretant, associations, etc., as they do in sound, and thus poetry unlike science is untranslatable). To grasp a concept is to know how to interpret the replicas of one or another of the symbols that express it. To think conceptually is to replicate symbols one knows how to interpret. As knowing-how is a skill, one can grasp a concept without being able to formulate it (other than by replicating a symbol that expresses it). Grasping it consists in habits, which, Peirce said, are the 'ultimate' form of a logical interpretant (chapter 2, section 10; chapter 6, section 8). Pragmatic clarification of meaning is formulating such habits in words, in subjunctive conditionals. But habits (themselves, not their verbal formulations) govern action, and action is with respect to objects outside our heads. Therefore, grasping a concept *always* transcends the merely neurological. How am I able to think of something, an individual or a type, that is outside my head? By replicating a proper name or a symbol,

whose replicas I know how to interpret through interaction with actual things, other replicas of which have occurred or may occur in indexical relation to that individual or to individuals of that type. (Compare our general account of intentionality; chapter 6, section 10.)

It follows that there are as many kinds of concept as there are kinds of symbol and that knowing how to interpret them varies accordingly. Some symbols incorporate indexical legisigns as parts – not syntactic parts, but as implicated in how the symbol is to be interpreted (recall the way Peirce distinguished semes, phemes, and delomes, by 'the nature of the influence of the sign': chapter 8, section 5). By way of illustration, consider the concept of actuality in general or of individuality in general. No concepts are more general, but they cannot be grasped in terms of universals (cf. Peirce's failed effort to define individuality in terms of inexhaustible predication; chapter 2, section 4). But how, then? We were at pains to establish that phaneroscopy's method is not purely descriptive but consists of reminders and prescriptions pointing us toward aspects of experience (chapter 3, section 2). The concept of actuality, of *haecceity*, defined phaneroscopically (chapter 3, section 4), is impossible to grasp apart from our actually experiencing particular forces and resistances. Indexical reference to actuality is built into our concept of the actual.

Some of Putnam's arguments turn on what may plausibly be said to be *the* meaning of a term or *the* concept of its referent. For example, 'That elms are called "elms" is not part of the concept of an elm' (1988, p. 27). That seems right. But at the same time, 'tree called "elm"' expresses *a* concept – as it happens, a concept of elms! Peirce, we noted, did not propose a general theory of meaning, much less one that limited meaning to one thing; nor is it our purpose to nail down what *the* meaning of a term is or what *the* concept of something is. Our topic is how symbols grow, hence, how concepts evolve. And for the purpose of understanding conceptual change, it is important to distinguish a variety of concepts and to identify those that form, as we said earlier, the warp of thought. It is here that Putnam's, and also Kripke's and others', development of the idea of rigid designation disappoints and that the Peircean alternative is superior.

The overstatement with which I am charging those authors is that of divorcing rigid designation too radically from conception, at least in the case of natural kind terms. The key is not that we lack any concept, prior to the development of a good theory, that is true of all and only that which belongs to a natural kind. We do have such concepts, for example,

that water is all and only the stuff that is the same in kind as the stuff we have been calling 'water'. The key, rather, is that any such pretheoretical concept depends on indices of actuality (as well as on assumptions that are subject to correction in subsequent inquiry). The indexical relation is primary and irreducible to a concept, but that does not prevent concepts from being built on it.

The overstatement leads to the absurd view, described in section 2, that rival theorists can share no idea, sufficient to identify the subject, of what it is that they disagree about – or, in other words, that a scientist with a false theory does not know what he is talking about. It would follow that the historian is limited to tracing the causal/historical chain of caloricists' utterances back to motions of molecules, in order to discover that they were referring by the word 'heat' to the same thing that the kineticists were referring to by that word. But that is ridiculous as well as impossible. We know from their own writings that caloricists were talking about what boils water, and so on – hence, the same thing, identified in the same way, as what the kineticists were talking about.

There is another absurd consequence as well. Theories are developed in stages of increasing specificity, and the choice among rival hypotheses is decided by relative success or failure in making one or another hypothesis more specific. The two-fluid theory of electricity was not developed successfully. The caloric and phlogiston theories were not developed successfully. Rival theories were. But before being discarded, each of these theories introduced a topic, a postulated entity, about which there was debate. Is phlogiston of negligible weight, no weight, or negative weight? Now, on the purely causal theory of reference, in which reference can only be to what exists, all of these debates, in the sciences of heat, combustion, and electricity, were really about exactly the same thing! – namely, nothing. For we now know that none of the posited entities exists. I have sketched an alternative theory of reference in Short 1980a (cf. Hacking 1983, pp. 86–7), in which entities posited in SHAs, whether they turn out to exist or not, are objects of reference.

Finally, there is the absurdity for which Feyerabend is notorious, that every observation disconfirming a theory depends on the adoption of a rival theory – thus making theory-choice subjective. That follows from three premisses, two explicit and true and one implicit and false. The true premisses are that we cannot observe that of which we have no concept and that concepts of physical entities are invariably 'laden' with logically contingent assumptions. The false premiss is that all such concepts are

on a par, that it is never the case that one is a more specific version of the other. We have already seen (section 3) how recognition of hypostatic abstraction and its inspecificity makes better sense of theory choice – a sense that cannot be made otherwise.

These three issues are parts of a larger question, the so-called incommensurability of theories, that we address in chapter 12. I have targeted Putnam because his response to these parts of the incommensurability problem is generally regarded as the best available (e.g., Hacking 1983, p. 75, who, however, indicates the need for some improvements).

6. Pragmatism and the Growth of Symbols

> Symbols grow. They come into being by development out of other signs, particularly from likenesses or from mixed signs partaking of the nature of likenesses and symbols. We think only in signs. These mental signs are of mixed nature; the symbol-parts of them are called concepts. If a man makes a new symbol, it is by thoughts involving concepts. So it is only out of symbols that a new symbol can grow. *Omne symbolum de symbolo.* A symbol, once in being, spreads among the peoples. In use and in experience, its meaning grows. (EP2:10, 1894)

> [E]very symbol is a living thing, in a very strict sense that is no mere figure of speech. The body of the symbol changes slowly, but its meaning inevitably grows, incorporates new elements and throws off old ones. But the effort of all should be to keep the *essence* of every scientific term unchanged and exact; although absolute exactitude is not so much as conceivable. Every symbol is, in its origin, either an image of the idea signified, or a reminiscence of some individual occurrence, person, or thing, connected with its meaning, or is a metaphor. (EP2:264, 1903)

> A symbol is essentially a purpose, that is to say, is a representation that seeks to make itself definite, or seeks to produce an interpretant more definite than itself. For its whole signification consists in its determining an interpretant; so that it is from its interpretant that it derives the actuality of its signification. (EP2:323, 1904)

I have quoted at length, including much that I cannot here comment on, to show the richness of the vein we are mining; each passage is torn from a context yet more suggestive. We may note that the growth of symbols is in many dimensions: in origin, from other symbols employed in connection with signs of other kinds; in body, that is, the sound and shape of the legisign, which can become more distinct, more easily and universally recognizable or more syntactically manageable; in spreading from one group of speakers to another, as, for example, monotheism

spread to Europe from the Mideast and the ideas of science and liberty have spread globally from Europe; and in meaning.

A symbol grows in meaning through 'use and experience'. Use implies purpose. Being a legisign, a symbol has a general purpose, which is to be used – replicated – for a variety of particular purposes (chapter 8, section 1). The replicas of a symbol are not symbols but are indices, sometimes of the symbol alone (and, through it, of the type the symbol represents), sometimes of particulars (chapter 8, section 2). By the meaning of a symbol, Peirce evidently meant our understanding of the type it denotes, which determines how we are to interpret its replicas. The emphasis is not on establishing reference but on what comes next, what we are to do with the items referred to. Experience is of success or failure in particular cases – success or failure, that is, of the conduct (including thought) that ensues when replicas are taken at face value. Experience thus leads to modifications of meaning. By unexpected consequences, good and bad, we learn more about what to expect from, or that can be done with, things of the type signified. Meaning is added to meaning. The meaning of a symbol 'inevitably grows, incorporates new elements and throws off old ones'.

Let us examine the preceding in more detail. When a replica, R, of a symbol, S, is an index of a particular, P, that indexical relation is only slightly dependent on S's meaning; it consists primarily in an existential relation between R and P. If, for example, I point in the direction of a pig frothing at the mouth and exclaim, 'Watch out for that dog!', you cannot, if you understand what 'dog' means, take me to be referring to a color or a shadow – to that extent S's meaning limits what P may be. But you can see that I am referring to a pig, mistaking it for a dog: P is not of the type that S represents. Suppose, now, that we are zoölogists attempting to construct an anatomical definition of the family of Ursidae, and to that end dissect a panda bear. We had little doubt that it was a bear, but discover that its anatomy varies too greatly from those of paradigmatic Ursidae to be one of them. The meanings of 'bear' and 'panda' are both altered, with consequent changes in what particulars we will use replicas of 'bear' to refer to. However, the type denoted by 'bear' remains the same. We now know better what that type is and what it contains, but the reference of the symbol 'bear' has not changed. Correction and growth of meaning has less to do with reference than with knowledge about the things referred to.

'But the effort', Peirce added, 'of all should be to keep the *essence* of every scientific term unchanged and exact'. That cannot be essence in the metaphysical sense, which pertains to things, not to words; but what is it? He told us in 1861: 'I believe in mooring our words by certain applications and letting them change their meaning as our conceptions of the things to which we have applied them progress'. That which makes a natural kind term a rigid designator is not the theoretical identification of its referent (possibly identifying *its* essence); it is the SHA by which it was first introduced (or it is that SHA as later refined). That abstraction, in which the reality of the physical world (chapter 7, section 4) is conjecturally presupposed, is the connecting thread of all the subsequent inquiry in which we seek more specific knowledge of the referent. Reference is not so far out of 'the head' as to be free of conception altogether, but it is fixed independently, at least for the most part, of the more specific ideas we may form of the referent. And these ideas become part of the meaning of the term – not primarily in order to fix its reference, but so as to make the symbol more useful (meaningful) to us, both in our myriad practical purposes and in scientific inquiry.

'A symbol is essentially a purpose, that is to say, is a representation that seeks to make itself definite.' As purpose is always indefinite yet can be realized only in some definite way, we may expect, where there is purpose, always some growth toward definiteness. Under the influence of Peirce, some linguists have made teleologically governed change the essence of language, and not only in its symbolic parts. These are Roman Jakobson, first of all (1990 [1966]; cf. Short (1998a)); but, then, deepening and extending Jakobson's views, and making greater use of Peirce's semeiotic, Michael Shapiro (1983, 1991) and Raimo Anttila (1989). Shapiro and Anttila extend the idea even to the grammar of language, which they argue seeks to become more diagrammatic, thus accounting for much of language's evolution. Peirce's ruling passion, however, was science, and with respect to science it is growth in knowledge – symbols and the meanings of symbols – that matters primarily.

The purpose of symbols in science may be said to promote the growth of symbols: theories guide inquiry so that they may be improved therein (chapter 12, section 3). Needless to say, this growth is toward truth (not only greater accuracy but, more importantly, toward more and deeper truths); but it might better be said that the growth comes first and that 'truth' becomes defined as its ideal limit (chapter 12, section 4). Growth – its growth and the growth it promotes – is the essence of a symbol. That is

what Peirce's pragmatism means by its emphasis on use. Verifiability and fixing reference are merely subordinate concerns. Far from identifying meaning with a finite list of verification conditions, Peirce's pragmatism entails that there can be no end to their enumeration, even as a symbol is currently understood and, more importantly, that new verification conditions are added as discoveries are made through a symbol's application. A symbol's meaning lies in that potentiality. Pragmatism is not a verificationist theory of meaning.

11

Semeiosis and the Mental

Peirce's early semeiotic was a theory of mind: it identified thoughts as signs interpreting signs. The mature semeiotic retains that identification but embeds it in a wider context. There are signs outside thought that thought interprets, and there are interpretants outside thought as well, in animal behavior especially. Thought itself is interpretable in behavior. It thus becomes possible to conjecture a natural history of thought: perhaps the capacity to think evolved from more primitive forms of semeiosis, when our ancestors began to interpret signs by other signs rather than directly in action.

The mature semeiotic entails a naturalistic theory of mind also in another respect: it explains significance (chapter 6, section 10), and the intentionality of thought is a special case of significance. A sign's significance is its bearing on the purposes of possible interpreters, and purpose, which may be practical or not, is not always conscious or otherwise dependent on an individual mind. The attribution of purpose, although it grounds valuation, is an empirical hypothesis, testable against observation. And the explanation of purposeful behavior, although not mechanistic, is naturalistic. In particular, thought's significance consists in its being interpretable, ultimately, in conduct. Pragmatism, which in 1878 was a maxim for the clarification of ideas, thus became in 1907 part of a naturalistic account of thought. It accounts for how we can think of what is, that exists independently of the thought of it, and also for how we can think of what is not.

This theory is not behavioristic; it does not reduce thought to behavior, actual or potential. It entails that thoughts are distinct from their manifestations in overt behavior. While the significance of thought lies in

its potentiality for being interpreted in conduct, thought is something (a sign) that is so interpretable. It exists whether or not it is interpreted, and it is distinct from its interpretants, actual or potential. Thoughts can be formed for their own sake. Furthermore, the habits potentially changed by thought are of expectation as well as of overt behavior; in other words, they pertain as much or more to what else we may think as to what we may do, though the possibility of doing remains essential to thought's significance. Most important, behaviorism conceives of animal behavior mechanistically, whereas Peirce's teleology entails a nonmechanistic conception of all that is purposeful, including animal behavior.

The theory is not reductive in any sense, since purposefulness has the same essential structure as intentionality: each has an object independently of whether that object exists; each has thus to be defined and explained in terms of general types of possible outcome, that is, final causes. Such explanation is logically distinct from any form of mechanistic explanation. Nor is it merely a way of thinking about phenomena that can also be explained mechanistically: what it explains can be explained in no other way. But that does not mean that it presupposes an idea of the mind or that it is theological or in any way mysterious. We have seen in chapter 5 that it is of a kind with explanations well established in the natural sciences, namely, explanation of organic features by natural selection and, at a further remove, statistical explanation of the anisotropic phenomena studied in thermodynamics.

The preceding ideas are no more than implicit in Peirce's writings. For example, he did not directly address the issue of intentionality as framed by Brentano. And those implications are obscured by his broad use of the words 'mind' and 'mental'. We have used and will continue to use 'mind' and its cognates as they are ordinarily used, to refer to contents and capacities assumed to vary with the individual, humans primarily if not uniquely. When Peirce wrote of the physical world as having mental characteristics or exhibiting the workings of mind, he was not referring to an individual mind. His point, rather, was to emphasize the continuity of human mentality with nature, and especially that the human mind's way of working is of a genus not uniquely human nor unique even to individual creatures. In the way words have, his usage is as likely to mislead the reader as it is to convey the meaning intended. The same point can be made differently, by asserting commonalities between mind narrowly construed and the (in that sense) mindless physical world.

There is another stumbling block as well. Most that Peirce wrote about mind was in the 1890s (five *Monist* articles, 1891–3, and related fragments

up to 1898: see especially 6.18–25, 66–87, 102–63, 214–86, and 7.388–523), before he developed his theory of final causation, in 1902. That account of mind was thus very different from his later views, less extensively expressed.[1] It is not our purpose in this chapter to review all that Peirce said about mind, but only to bring out the bearing of his mature semeiotic on contemporary discussions. That is the apologia I would address to anyone who objects, as I imagine some will anyway, that we are reading views into Peirce's writings that he never expressed, while ignoring much that he did say about mind.

1. Contemporary Philosophy of Mind

The philosophy of mind has been especially active over the past half-century, following the demise of behaviorism. Alternatives and arguments proliferate. Physicalism, the view that anything that is, is physical, is the reigning assumption. In sections 2 through 4, we comment on one thread only of these debates. However, an initial word may be said about the range of contemporary opinion and about how Peirce's approach compares to it, overall.

Feeling, thought, consciousness, and selfhood appear not to be physical or to be explicable by the laws of physics that presumably apply to neuronal activity in the brain. However, what seems is not always what is; heat does not feel like motion, but it is molecular motion anyway. Some philosophers (from Place 1956 to Armstrong 1968) have suggested that the mental may similarly be identical in fact with neural processes. The plausibility of that view depends on the theory explored in the preceding chapter, according to which a term may refer independently of what it is

[1] The 'law of mind' of which Peirce wrote in 1892 (6.102–63) is that of association or generalization, by which ideas and feelings affect one another and spread, forming habits. It is supposed to operate outside human minds and more fundamentally than mechanical causation: thus matter is 'effete mind, inveterate habits becoming physical laws' (6.25); cf. chapter 5, note 12. There is no reference in these years to purpose or final causation or teleology, except for one pregnant passage on the personality as a 'coördination or connection of ideas' (6.155), which is also a 'teleological harmony in ideas' (6.156): see below, section 5. But it is doubtful whether the law of association can explain so much. In 1902 that law was superseded, or possibly it was glossed, by the idea of final causation: 'The psychologists say that consciousness is the essential attribute of mind; and that purpose is only a special modification. I hold that purpose, or rather, final causation, of which purpose is the conscious modification, is the essential subject of psychologists' own studies; and that consciousness is a special, and not a universal, accompaniment of mind' (7.366). Here again, 'mind' is used broadly, so that it applies beyond the confines of individual human minds, only now it is identified with final causation.

taken to mean. By 'pain' we do not ordinarily mean firings in C-fibers, but that may be what the word refers to, if it is those firings that cause us to feel pain.

However, the grammar of mental discourse does not fit very well with physical description. Is it possible to identify each mental entity, postulated in our ordinary ways of describing and explaining human behavior, with one or another neurological entity? Against that 'identity theory' there stands the position sometimes named 'eliminative materialism', that mind is not real at all. Some regard talk of things mental as a 'folk psychology' that will ultimately be replaced by a scientifically respectable language; in the latter, nothing not physical will be referred to (Churchland 1979, 1981; Stitch 1983). The idea behind such views is again one discussed in the preceding chapter, that all description, no matter how ordinary, is 'laden' with assumptions that may be challenged. That idea applies to first-person reports of one's own thoughts and feelings; they, too, involve theoretical interpretation. In arguments traceable to Wilfrid Sellars (1963, ch. 5) and Wittgenstein's *Investigations*, to Peirce's 1868–9 papers (chapter 2, section 3), and to Kant's first *Critique*, the Cartesian assumption that we have intuitive knowledge of our own minds is rejected.

Between the identity theory and eliminativism, there is a family of positions named 'functionalism' that retain the reality of the mental, but only by identifying it abstractly. The distinction between computer software and hardware is the paradigm: the same programs can be run on machines differing in construction and materials. Also, the same functional relations between input and output can be realized by different programs. If thinking, feeling, and so on are defined functionally, then they may be realized differently in different creatures, possibly in computers and robots also. In the commonly used jargon that derives from Peirce, there is no identity of mental types with physical types, but there is an identity of tokens of mental types with tokens of physical types. Different tokens of the same mental type may be tokens of different physical types. Functionalism is due to many but most of all to Hilary Putnam (in papers published from 1960 to 1973: Putnam 1975, chs. 14, 18–21), who now rejects it (1988, 1992).

Daniel Dennett dwells here but with a more complicated attitude. An eliminativist with respect to the mind postulated in folk psychology, he argues nonetheless that we are 'intentional systems', whose behavior can be understood, to a degree, and often predicted from an 'intentional stance' (1978, 1987). Taking that stance, we attribute rationally formed

beliefs and desires to the system in question, predicting that it will act accordingly. Irrational behavior limits but does not destroy the usefulness of that approach: humans *are* intentional systems, albeit imperfectly. Nothing follows about the inner workings of an intentional system, however, as programs and hardware may vary. Taking an intentional stance requires us to articulate desires and beliefs as sentences, but the hardware is something else. Furthermore, verbal articulation may impose a specificity unjustified by the behavior observed: Dennett here applies Quine's strictures about the indeterminacy of 'radical translation'. While Dennett admits that his view seems ambiguous on the question of the reality of intentionality, he insists that it unambiguously occupies a 'knife-edge' between treating intentionality as real and as a useful fiction (1987, p. 37). On the one hand, he says, the hardware, as perfectly physical, lacks intentionality, but on the other hand, there are systems, such as humans, that as a matter of objective fact behave in ways successfully described from the intentional stance.

Others of physicalist persuasion go so far as to assert the existence of the mental as distinct from neurons and the like, either as higher-level properties of neurological systems (John Searle 1983, 1992) or as immaterial but 'supervenient' upon material systems (Jaegewon Kim 1993). Either way, the mental is neither type-identical nor token-identical with the neuronal; it exists dependently on neurons but is not merely an abstract way of talking about them. The idea of supervenience raises difficult questions that Searle's view appears to avoid (1994, pp. 124–6). His telling example is of liquidity, a concept that does not apply to individual molecules, say, of water, though something's being liquid is explained, he says, by the properties of its molecules, in a 'bottom-up' causality (1994, ch. 5). So also, neurons and the mind: Searle maintains that mental states and acts are physically real, albeit they are properties uniquely of complex neurological systems irreducible to the properties of the parts of those systems. You cannot look inside the brain for its thoughts, but thoughts are as much a feature of the living brain as liquidity is of a water-fall.

This sketch of a few contemporary views, which have been argued at great length and complexly stated in thousands of pages of text, will have to suffice by way of orientation. Below, we will attend in more detail to some discussions that have taken place among functionalists (using that elastic term at its furthest stretch). It will be convenient, now, to reflect in a general way on how Peirce's view fits, or fails to fit, with contemporary thought.

The comparison is made problematic by something we noticed above (chapter 4, section 2): just as ideas of the mechanical and of matter are vague, changing, or controversial, so also there is no clear conception of the physical. Indeed, physicalism is usually supposed to be the same as materialism and/or mechanicalism, and thus it has the same defect of ambiguity that they have. The safest definition of the physical is that it is what the theories of physics postulate. But as those theories are in flux and may be expected to change dramatically in the future, that is not a definition that conveys any very specific idea about the physical world. Besides, even were quantum mechanics in its present development the last word, could we say that we understand what it represents the physical to be?[2]

If Peirce was right that statistical reasoning in thermodynamics is not mechanistic, then the physical is not the same as the mechanistic. In other words, not all aspects of systems of material particles are explicable mechanistically. The same follows if he is right that natural selection is irreducible to mechanistic explanation. Peirce's view of the mind may therefore be physicalistic by his own implied account of the physical and yet be anathema to self-described physicalists in contemporary philosophy of mind. (That is so, independently of his speculative cosmological hypothesis, of the 1890s, that matter, qua perfectly mechanical, is decayed mind (note 1 and chapter 5, note 12), which we agree to discard.)

Take, for example, Searle, whose assumption of bottom-up causation is questionable.[3] How, exactly, do the properties of molecules explain their behavior in systems? A mechanistic account may suffice for liquidity per se (I do not know), but relatively simple phenomena pertaining to liquids (eddies in water moving past constrictions, convection cells in water being heated from below) provided Ilya Prigogine examples for his account

[2] Yes, there are things of philosophical importance we can claim to understand about quantum mechanics, such as that it has made chance a fundamental feature of existence. But there is no consensus about the physical meaning of the theory's formalisms. And what are we to make of the phenomena of nonseparability?

[3] Putnam, in 1973, argued forcefully that the behavior of a complex system is not always explained bottom-up but, rather, in terms of its own parameters: it is the shape of the round peg that explains why it fits in the round hole, regardless of whether that shape is of a latticework of molecules or of a continuous stuff (1975, ch. 14). That suggests the view I am urging here (chapter 5, section 6, and the present chapter), with the basic amendment that the explanation may sometimes be teleological rather than mechanistic (or nomological; chapter 4, section 2). In the present instance, however, the question is indeed how we get to the complex whole from pre-existing or independently existing parts. Ergo, Searle's bottom-up view is initially plausible and cannot be rejected on Putnam's grounds alone.

of the emergence of molar order out of molecular chaos (chapter 5, section 4). On that account, you will recall, the higher-level properties, as Searle would call them, have to be explained as satisfying one of the types of order that are alternative solutions to the nonlinear equations that define the system. Those equations are not part of the theory of individual molecules. Explanandum and explanans, both, are at the same 'high' level. Of course, that depends on there being lower-level parts, in these cases molecules, conforming to their own laws. But the latter laws alone do not explain molar order. Non-equilibrium thermodynamics shows that one can be a physicalist about higher-level properties only by abandoning the view that bottom-up causation is necessary to physicalism. By our argument in chapters 4 and 5, that means that some physical processes have aspects that can only be explained in ways that do not conform to standard ideas of mechanistic explanation. That conclusion obtains even before we get to the phenomena of life, much less to those of mind.

Questions about what physicalism entails may be connected in some minds with questions about what is observable. In contrast to Dennett and the Continental phenomenologists alike, Peirce did not portray nonmechanical phenomena as requiring, for their observation, our taking a nonscientific 'stance' – in effect denying that they are phenomena. Either sort of phenomenon, mechanical or nonmechanical, is identified by empirically testable conjecture; either is fallibly revealed by a constancy in experience, of sequence in the one case, of result in the other (chapter 5, section 5, chapter 6, section 2). To be sure, specific purposes are harder to identify in human behavior than in that of the lower animals, and, even once identified, inferences therefrom to what is thought or desired are tenuous (see below, section 5). But those differences, while profound, do not mean that the mental is either unobservable or inexplicable.

2. Functionalism's Problem with Content

Functionalism identifies the physical reality of mind with neurological states and processes comparable to the electronic states and processes internal to a computer. You confront a pile of unpaid bills, add up some figures, and conclude that you are bankrupt; someone feeds like data into a computer, runs the program, and the computer announces that you are bankrupt. There is a difference: you draw your conclusion because you know that $2 + 2 = 4$, or, more pertinently, that $2 - 4 = -2$, whereas the computer proceeds simply as electronic cause is followed by electronic effect. But the presumption is that that difference is more apparent than

real and, like your bank balance, will disappear under scrutiny. The reason is that our brains do our thinking for us and must work something like a computer.

In either case, there is, normally, a functional relationship between input and output. Given true premisses as input, a good computer or good thinking produces true conclusions as output. This is called 'preserving truth'. That processes of cause and effect can be harnessed to the requirements of logical reasoning is no mystery: it is what hardware designers and software programmers do very well. Logicians paved the way by developing rule-governed syntactic systems and 'mechanical' decision procedures that have such desirable semantic features as that of preserving truth. Nature preceded logicians by producing brains. But how can we account for the meaning that truth presupposes? Whence the semantic interpretation of syntactic systems, programs, machines?

A platitude: what computers do has no meaning in itself. It is we humans who assign computer operations their meaning. But if our brains are themselves computers, then from where do *its* operations get *their* meaning?

Functionalism, narrowly conceived in terms of relations of proximal input (key-punches, sensations) to proximal output (speech, printouts), seems confronted with the impossible task of accounting for semantics in terms of syntax, that is, in terms of all the sequences (inferences) allowed by a given system or computer or brain. What do we mean by 'mass'? Well, from what sentences may we infer 'X has mass M'? And what sentences may we infer from 'X has mass M'? The more inferential connections there are, connecting concept C to others (mass to weight, to inertia, to acceleration, to the bending of space, and so on), either the narrower or the richer C will be. This is the translation theory of meaning that we have encountered before (chapter 2, sections 5 and 6, chapter 6, section 3) and rejected before. It does not get us out of the circle of words interpreting words. In fact, it does not get us from syntax to semantics. The popular doctrine of 'meaning holism' (chapter 12, section 1), according to which a term means what the theory or language in which it occurs makes it mean (cf. the 'theory-ladenness' of observation, chapter 10, section 3), if it is not supplemented by other ideas of meaning, is another version of the same.

Functionalism, less narrowly conceived in terms of relations of distal input and distal output, gives greater hope. Distal input may be the physical causes of sensation, distal output, bodily motions, and manipulations of other bodies. If, to take a stock example, we are so constructed or

conditioned as to say or think 'horse' in the presence of a horse, then that is what 'horse' means. If we are told, 'Look for a horse' and we engage in looking-for behavior until we encounter a horse, at which time we stop looking and, beaming, point toward the horse, then that, too, is what 'horse' means. The example is crude, but it shows how connections may be established between words and things, and that seems to capture the idea of semantics. An implication, welcome to some, decidedly not to others, is that computers similarly equipped to sense and to act – that is, robots – will not need humans to achieve meaning; they will be endowed with mentality as much as are we, perhaps more so.

There is, however, a problem. It is functionalism's problem with content. If a cause-effect relationship between presence of horses and tokenings of 'horse' accounts for the meaning of 'horse', then 'horse' must include within its meaning anything that might be mistaken for a horse, since those items, too, cause one to think or say 'horse'. It turns out that 'horse' means 'horse or cow or . . . '. Hence, one can never be mistaken in applying it. Which is not how it is, or even how it ought to be; for if no mistake is possible, then nothing is being said. Jerry Fodor has labeled this 'the disjunction problem'. One might suppose that it can be evaded by suitable amendments, but Fodor persists in rediscovering the problem in its purported solutions.

For example, Fred Dretske suggested that a term means that which caused tokenings of it during the period in which its use was learned; later, the term may be applied erroneously to other things (1981, pp. 194–5). But, recasting Fodor's argument (1990, chs. 2 and 3), if there is any learning at all, there is generalization: horses are things *like* this. The term cannot mean only those exact instances in relation to which it was learned. And then, if it is subsequently applied to cows, on what basis can we deny that cows are part of what is meant by 'horse'? What account of generalization would explain generalizing from gray mares to chestnut stallions but not to gray cows? Why would only the latter application of 'horse' be a mistake? Accounts are easy enough to devise: for example, one is shown a horse and *told* that 'horse' means any animal of the same species or any animal that looks like this one in such and so respects, or one innately grasps what sort of features are germane to defining a natural kind. But those accounts must incorporate the content at issue (e.g., concepts of species or of relevant respects), begging the question.

In response to such problems, Ruth Garrett Millikan (1984, 1993), David Papineau (1984, 1987), and Elliott Sober (1985) have argued that functionalism needs a normative or teleological concept of function

(putting the function back into functionalism, in Sober's phrase). Very roughly, a function in this sense is defined not by what happens (as when we speak of how a machine functions) or by a general relation of dependency (as in a mathematical function) but by what ought to happen. The contents – the intentional objects – of thoughts, desires and so on, are then to be found not in the actual causes or actual effects of those mental states, or in any subset of their actual causes or effects, but in what they ought to be caused by or ought to cause. At the same time, this teleological functionalism is intended to be naturalistic. Hence, 'ought' (or equivalent normative language) is defined in terms of what something has been selected to do, in the first instance by Darwinian natural selection. It follows that intentionality will not be limited to human mental functions. Thus we witness contemporary philosophers gradually groping their way back to the position Peirce formulated from about 1902 to 1907, that is, a hundred years ago. I think they have not yet made it all the way.

Referring to Millikan, Papineau, and others, Fodor (1990, ch. 3) argues that 'teleological notions, insofar as they are themselves naturalistic, always have a problem about indeterminacy just where intentionality has its problem about disjunction' (p. 70). Thus we come to frogs, flies, and BBs, a staple of this debate. Frogs snap up small dark flitting objects. Their disposition so to behave has been selected in amphibian evolution because, often enough, such objects are flies and flies nourish frogs. But may we then say that the retinal image that triggers snapping *means* 'fly' to a frog? (Must we say that it means anything? But let us defer that question to the next section.) Why not 'small dark flitting thing'? Or 'fly or BB', at least in neighborhoods where boys or philosophers test frogs by shooting BBs past them? If the image means 'fly', then frogs may mistake BBs for flies, while if it means 'fly or BB' then no such mistake can be made, but a frog may still fail to be nourished by its catch of the day. Fodor insists that it makes no difference which way we put it: 'Darwin cares how many flies you eat, but not what description you eat them under' (p. 73).

However, Fodor has switched attention to the operation of the mechanism and away from the type of outcome for which that mechanism was selected (he takes the function back out of functionalism). The mechanism at issue exists in order to catch flies; for it is that type of outcome (and not flies-or-BBs, nor small dark flitting things) that explains its selection. The type selected-for does not have the indeterminacy Fodor alleges (see chapter 6, section 3, on the degree to which interpretation's object is specific and the basis on which it is specific to that degree). Fodor's

argument fails because it neglects the fact that selection is for a type of outcome.[4] He has ignored what is essential to teleology, namely, the final cause. This is a reprise of the error made by Braithwaite, that led to Scheffler's posing him the problem of multiple goals (chapter 5, section 6).

The fault is not all Fodor's. The authors he cites appear to share the view of Wimsatt, Wright, and Brandon (chapter 5, section 6), that explanation by natural selection, though not itself mechanistic, is nonetheless consistent with mechanicalism. The same is implied about any teleological explanation modeled on natural selection. The consistency is made out this way: teleological explanations are construed as historical, which mechanistic explanations are not, but the history is of processes that were exclusively mechanical. Papineau writes that 'when we talk of some characteristic C being present *in order* to produce E, we should understand ourselves to be claiming that C is now present because of some past selection process that favored C because it produced E' (1987, p. 65). This formulation is fine, as far as it goes; indeed, it cannot be understood without taking E to have been a cause of C's selection – a fortiori a final cause, as E is a type and not a particular event. But it does not explicitly say that E is a cause, nor does Papineau say that elsewhere. And thus it is easy, in the present mechanicalist climate, to suppose that he means that the complete explanation of C is its past mechanical effects (and their effects, in turn, on the reproductive success of past organisms that possessed C, the ancestors of present-day C-possessors). Presumably, that is what Papineau means. E comes into his account only as the type of the effects of C, and not as a cause.

[4] Fodor is aware of the distinction between selection of and selection for; but even when discussing it, he reduces the latter to the former. He writes that 'appeals to evolutionary history can't distinguish selection for *being F* from selection for *being G* when G and F are necessarily coextensive: If you always get Fs and Gs together, then a mechanism that selects one *thereby* selects the other, so the utility of being F and being G always comes out the same' (p. 87n23). If by 'necessarily' Fodor refers to logical necessity, then, before the colon, we have an unproblematic truism. So, presumably Fodor means something weaker than logical necessity. A stock example he discusses later in the same note is that the heart's pumping is coëxtensive with its making a rhythmic noise. That is presumably the sort of thing he means: for whatever reason, one does not occur without the other. But it is simply wrong that evolutionary history – that is, if it is informed by the theory of natural selection – cannot distinguish one of these as that for which the heart has been selected. Heartbeat has many uses, but it is circulating the blood that is crucial and explains the heart's particular construction and location. That is what it was selected for. Selection-for is more specific than Fodor is willing to admit. After the colon, we have an outrageous non sequitur; for the utility of pumping and the utility of rhythmic sound are *not* the same, even if the beating and the pumping always occur together.

Similarly, if less clearly, Millikan writes that 'it is not then the actual constitution, powers, or dispositions of a thing that make it a member of a certain biological category' – that is, biological categories are not defined mechanistically – and continues, 'My claim will be that it is the "proper function" of a thing that puts it in a biological category, and this has to do not with its powers but with its history' (1984, p. 17). 'Proper function' is Millikan's term for the kind of function that is teleological: something belongs to a biological category according to what it is its proper function to do (as determined by its history), whether it does it or not (as determined by its mechanical powers). What is the history that accords something a proper function? 'Having a proper function is a matter of having been "designed to" or being "supposed to" (impersonal) perform a certain function' (ibid.). The scare quotes and the parenthetical qualification alert us that no Designer is intended. Millikan continues, 'The task of the theory of proper functions is to define this sense of "designed to" or "supposed to" in naturalist, nonnormative, and nonmysterious terms'. (So are they, then, in no sense normative? Then why use normative language at all?) Again, without explicit disavowal that naturalistic explanation is constrained by mechanicalist assumptions, one can only suppose, in this climate, that Millikan does not intend to challenge mechanicalism. Thus, the only causes implied by her account (taken as it is reasonable to suppose she intends it) are mechanical causes.

And that is why Fodor can with some plausibility treat teleological functionalism as yet another attempt to account for intentionality in terms of some privileged set of actual causes and actual effects, taking these to be mechanical causes and effects, that is, particulars, not types. Whereas Dretske identified that privileged set with the occasions via which something's meaning was learned, Papineau and Millikan identify it with the occasions via which an organic feature evolved or some device, such as a language, was subsequently selected, artifactually. Thus the argument that worked against Dretske's theory also works, or seems to, against their theory. To be sure, Fodor misses the point of teleology, but only because his adversaries have failed to state it adequately.

Dennett is another example of the same. He criticized an earlier expression of Fodor's views along the lines we have taken here (1987, p. 311). However, as he makes the teleological reading of Darwin all a game of 'as if' ('taking the intentional stance'), his objection cannot be taken seriously. Indeed, from the same stance, he has felt free to talk of 'Mother Nature' and what she wants and chooses, and so on, showing that, for him, all teleological talk is equally anthropomorphic. And

therefore Fodor has felt free to laugh his critique out of court (1990, pp. 75–9). You cannot defend a teleological theory and in the same breath declare that it is not literally true. Dennett's knife edge is too sharp to stand on.

Teleological functionalism has therefore not been refuted, because it has not been adequately stated. An adequate statement of it would require adopting something like Peirce's concept of final causation. In other words, it requires taking the idea of 'selection for' seriously: it requires putting the type selected-for back into functionalism.

3. On Being Simple-minded

A last comment on the debates to which Fodor is central: all the disputants seem to agree that, if any intentionality is involved in the case at all, then frogs' retinal images must mean something to them (even if, with Dennett, that is only an 'as if'). For all have worked down to frogs after beginning with traditional philosophy of mind, which is about humans, and, in particular, from the theory that human thought might be understood as 'inner representation'. In other words, all still think of intentionality essentially in Cartesian fashion, as something possessed by 'inner' episodes; only, the container has been switched from a nonspatial *res cogitans* to the nervous system. Peirce's theory is more subtle than this, and for that reason it is likely to be misunderstood. Hence this section.

Two points: first, Peirce's theory makes all significance to be relative to interpretation, actual or potential, but the interpretant does not have to be a thought or other 'inner' episode. It can be overt behavior. It must be purposeful, but purposefulness, we have noted repeatedly, does not always depend on thought, and, moreover, it is observable, even if often with difficulty and uncertainty, just as mechanical causation is observable, often with difficulty and uncertainty. Second, what is interpreted need not be internal to the interpreter. Significance begins, as it were, on the outside. It is the odor that the bear's behavior interprets as a sign of grubs. That odor is something physical. The first things to which interpretation accords significance are physical objects, events, conditions. It is only with more developed neurological systems that there might be 'inner' representations. It would therefore be a gross error to suppose that a semeiotic analysis of animal behavior entails that animals possess some sort of 'inner representations'.

In applying Peirce's semeiotic, the critical question is whether what is going on is purposeful – not merely whether it has a purpose but whether

it is done purposefully. What is done purposefully is subject to correction by the agent itself ('correction' in the course of a species' evolution or an artifact's construction does not make an individual organism or artifact act purposefully, even though it gives some of their features or actions a purpose; chapter 4, section 6; chapter 5, section 7). And what is not done purposefully is not an interpretation. Now, if (as it appears) frogs cannot correct their snapping, once having a retinal image of a certain sort, then they are not interpreting that image as a sign of anything. The image is not a representation. It is simply a mechanical effect of one thing and a mechanical cause of another, albeit the whole arrangement has a purpose and serves a purpose, for which it was selected in amphibian evolution.

To be sure, the frog's behavior can only be described and understood in terms of purposes that it has – it is fly-directed – and those purposes may fail to be fulfilled because a condition necessary to their fulfillment is missing. Sometimes, a frog snaps at a fly that is not there. But this is not intentional inexistence. Unless the frog is acting purposefully, it interprets nothing as being a sign, and in that case there is no intentionality.

Applying the same doctrine to bears, we get a different conclusion, because ursine behavior, even if largely instinctual, appears to be under the individual bear's control to some degree, as it is modified in light of failure. But whether there are any neurological episodes to which the bear similarly makes a correctable response is another matter. Its *behavior* interprets the *odor* as a sign of grubs. No inner representations of grubs, or of anything, are necessary, either as signs interpreted or as interpretants.

At what point do there begin to be inner representations? I would suggest that it is when creatures become capable of *not* acting. I hesitate, ergo, I am thinking. Alternatives are being considered, which is possible only by forming representations of them. There is a thought that I can act on, or not. It is being interpreted (depending on the sort of thought it is) as a proposition true or not, an aim desirable or not or permissible or not, a plan prudent or not. Once this power of hesitation is formed, then of course we can indulge in thought for no further purpose.

The moral is that bears are quite simple-minded, that frogs are so simple as perhaps to be mindless, and that we should not simple-mindedly suppose that all significance depends on thought or 'inner' episodes à la Descartes. Significance and interpretation antedate mind, which is why we can analyze and explain the mental as semeiosis of a higher order, namely, that which occurs when there is hesitation, when action is not

automatic and signs, instead of being interpreted in action, are interpreted by further signs.

Necessarily, signs that interpret must be produced for that purpose, as interpretation is purposeful (chapter 6, section 3). And as they interpret only by signifying, they must be produced with a purpose to signify. Typically, then, they will be replications of legisigns; for legisigns are those signs that exist to be used to signify (chapter 8, section 1). But legisigns are of many kinds, the varieties of which Peirce had begun to explore (chapter 8, section 3). Such a study potentially yields a definite and detailed theory of thought as representation. Blips on a frog's retina are not included. Overt speech, however, is included. One can think aloud. As has often been suggested, thought might first occur in that public sphere before being 'internalized' as silent converse with oneself. It follows that the mental is not essentially 'inner', hidden away from public observation.

4. Beyond Biology

Peirce has not been absent entirely from contemporary philosophy of mind. The following instances are Peircean in different ways. Together, they lead us from biology to psychology and beyond.

In many respects, Ruth Millikan has been engaged in reinventing Peirce's wheel. Despite a terminology of her own invention, differences in every detail, complexities where Peirce is simple and simplicities where Peirce is complex, the correspondence is remarkable.[5] In the case of two key concepts, even the language is the same. 'The notion of a sign makes intrinsic reference to a possible interpreter', she says, though she says it in a special context, with reference to what she distinguishes as 'natural signs', and not in the central position the axiom deserves (1984, p. 118). And, as we will see in a moment, she also speaks of significance, though of one kind only, as being 'grounded' in a relation of sign to signified. But, while thrice alluding to Peirce's theory in her 1984 book, once describing it as a 'general theory of signs', she does not avail herself of it, even when constructing such a theory, so labeled, herself (1984, chs. 5–9). And that is a pity, as the distinctions she draws might have been refined

[5] Lest it seem that I have followed her ideas in reconstructing Peirce's theory, I should mention that I first stated my view in an article, 'Semeiosis and Intentionality' (Short 1981b), three years before Millikan's 1984 book (which I believe is her first publication). Nor do I mean to imply that Millikan derived any of her ideas from Peirce; she is too evidently an original thinker.

by comparison to those Peirce drew, both in his taxonomy of signs and in his divisions of objects and interpretants.

More importantly, while Millikan does not pursue the teleological idea as deeply as Peirce did, her philosophy of mind has other metaphysical linkages that recall Peirce's. She writes that 'intentionality is grounded in external natural relations . . . between representations and represented. . . . Hence nothing that is either merely in consciousness or merely "in the head" displays intentionality *as such*' (1984, p. 93). Meaning, then, is a matter of engagement with the world. But what must the world be, if it is to play its part in that transaction? The passage quoted echoes Putnam's account of reference, but Millikan takes issue, in an epilogue, with Putnam's distinction between internal and metaphysical realism (pp. 329–32). As did we (chapter 7, section 4), she finds that intentionality requires a world more independent of its representation than Putnam's internal realism permits but less remote from representation than metaphysical realism ordains (also in 1993, ch. 10).

The identification of meaning with real-world engagement is pragmatism, and thus we may expect much of Peirce's philosophy to reappear in Millikan's, not derivatively but because like premises lead to like conclusions. From that identification, she argues against what she calls 'meaning rationalism', the assumption that we know a priori what we mean or what the contents of our own thoughts are (1984, pp. 10, 326ff.). As the knowledge in question is not a priori, strictly speaking, a better term would be 'intuition' or 'intuitive cognition' in the sense that Peirce employed in 1868, when he attacked the Cartesian assumption that we have intuitive knowledge of our own minds (chapter 2, section 3). Indeed, Millikan's critique of meaning rationalism resembles, in part, Peirce's of Cartesianism (1984, p. 327). The more important strand of her argument, however, is the one already mentioned: because the real world is our partner in meaning, the meaning of an idea depends on how it may be applied; but how it may be applied depends on other ideas with which it may be combined, not all of which come to mind immediately or, indeed, can come to mind in any finite time. Besides, we learn more about the consequences of those applications by actually making them. Hence Peirce's pragmatic maxim for clarifying our ideas, by enumerating their practical consequences, and hence, also, his doctrine, explored in the preceding chapter, that symbols grow in meaning. The common objection to Peirce's pragmatic maxim, that there is no limit to an idea's practical implications, is rather part of that maxim's point. Only on the assumption that we must already know what

we mean – the assumption Millikan dubs 'meaning rationalism' – could the endlessness of the explications it entails invalidate the pragmatic maxim.

However, Millikan's premiss, that 'intentionality is grounded in external natural relations', results in a much narrower theory than Peirce's. For, by such relations she means 'mappings' between thoughts or sentences, on the one hand, and physical conditions either actual or desired, on the other. In consequence, her 'general theory of signs' is surprisingly limited. She makes indicative and imperative sentences to be the 'paradigm' of all signs (1984, pp. 85, 115). With a nod to Peirce, she names them 'intentional icons' (p. 95), presumably because their function is to map what is or what is to be done (p. 97). But they must also be 'articulate' by having distinct parts or aspects: the sign 'divides into significant aspects', she says (p. 116). For, otherwise, the sign would not indicate to what its mapping applies. However, the narrowness of Millikan's view undercuts its own articulation. In the first place, indication is not iconic but indexical, hence, a sentence is not merely iconic. But that is a quibble. A more serious problem is that, if every sign must be articulate, then articulation is impossible. For, either the significant aspects into which a sign articulates are not signs themselves – in which case, how could they be significant? – or they, too, subdivide – and so on, ad infinitum. To solve this problem, Millikan needs to admit signs that are not articulate but that are pure indices and pure icons.

Millikan notices that sentences are produced in order to signify: 'Normally a sentence stands midway between two cooperating devices, a producer device and an interpreter device, which are designed or standardized to fit one another' (1984, pp. 96–7). We made the same point in different words about legisigns and their replication (chapter 8, section 1), sentences being one kind of legisign or replica of legisign. But, by making sentences paradigmatic of all signs, Millikan excludes signs not produced for the sake of signifying. She struggles to include them anyway, under the rubric 'natural signs', in this manner: 'A natural sign is analogous to an intentional icon in that it is the kind of thing that could be used by an interpreter in exactly the same way that interpreters use intentional icons' (1984, p. 118). Well, not exactly, since the interpretation of a sign uttered for the purpose of signifying must take that purpose into account (see the discussion of Grice in chapter 8, section 1), but a natural sign occurs without a purpose to signify. Peirce's broader conception of signs avoids this problem; it includes *signa naturalia* naturally, as it were.

Again, Millikan has a chapter on indexical expressions, that is, indexical legisigns and their replicas (chapter 8, section 3), but such expressions work only through their indexical aspects, where the latter have to be understood as Peirce understood pure indices, a category Millikan omits. In all these ways, Peirce's more encompassing semeiotic is preferable to one that takes sentences to be the ruling paradigm. If we want to use a theory of signs to build a naturalistic explanation of mentality, then that theory should employ the broadest possible conceptions – ones that apply to human languages and animal gestures and natural signs equally. Making sentences paradigmatic subverts the naturalist intention.

Unlike Millikan's work, James H. Fetzer's 1990 book is explicitly Peircean. Its second chapter contains a statement of Peirce's theory of signs that is applied, in the next chapter, to the philosophy of mind; Fetzer's thesis is that minds are semeiotic systems in Peirce's sense, not merely computational systems as in the 'artificial intelligence' program. However, that is about the extent of the similarity of Fetzer's study to ours. Fetzer's understanding of Peirce's semeiotic is through the medium of Charles Morris' sign theory. (Morris was also an influence on Millikan, whose first book is dedicated to his memory.) But the relation of Morris' ideas to Peirce's is tenuous, and thus they are a hazardous guide to the latter. For example, Fetzer takes the type/token distinction to apply to any class of sign and its instances: 'any property (or pattern) that can have distinct instances may be characterized as a type (or kind), where instances of that type (or kind) qualify as tokens. . . . This distinction applies to icons, to indices, and to symbols alike' (p. 34). And thus what is distinctive about legisigns and their replicas is lost, or, conversely, what is lost is any notion that there are signs not produced in order to signify. More importantly, while Fetzer rightly stresses interpretation and calls semeiotic systems 'causal systems of a special kind' (p. 35), he betrays no sense that Peirce's conception of that special causality was teleological. Fetzer nevertheless attempts thereby to make an important point, that research in artificial intelligence aims only at simulating, not duplicating, human intelligence, as it omits Peircean interpretation of signs (pp. xiv–xv and passim). But whether that thesis can be sustained without recognizing that interpretation presupposes purpose, and that purposefulness cannot be reduced to mechanical causation, is another question.

Our third and final Peircean exhibit is provided by Hilary Putnam, who has questioned the assumption of many analytic philosophers that science, physics especially, tells, or will tell, the whole story about what the world really is. His concern, like Berkeley's, like Kant's, is to save morality

and religion. (This has been a growing theme of Putnam's since about 1976, but we shall take his 1992 book as our text.) Now, Putnam has written authoritatively on some aspects of Peirce's philosophy and has sometimes expressed sympathy with Peircean pragmatism, but he has not associated Peirce with a critique of scientism (as he calls it). One might think, at first, that that is for good reason. No philosopher has been more besotted with science than Peirce, who went so far as to portray scientific inquiry as a model of the moral and religious life and who made philosophy itself to be a science, or set of sciences, pursued in relation to the other sciences (chapter 3, section 1).

But those same facts prompt second thoughts. Might it be that Peirce's conception of science avoids scientistic narrowness? One of his philosophical sciences is phaneroscopy, emphatically not based on physics, and three other philosophical sciences are designated collectively as 'normative'. Moreover, the very fact that Peirce wrote about science in religious, specifically Christian, and moral terms suggests that he shared Putnam's worry (apart from its specific reference to the current state of professional philosophy). Such worries were common in the late nineteenth century, when, in Matthew Arnold's simile, the Christian faith began to ebb, like a sea, from the shore of human life. Putnam's response to the problem seems the polar opposite of Peirce's – to divide rather than to meld – but the difference may be less than appears. Peirce himself decisively divided the practical from the theoretical, defending the autonomy of practical wisdom for practical purposes (RLT: ch. 1). And I shall now argue that where Peirce does differ from Putnam, it is to close a fissure – right where the philosophy of mind lies – that otherwise would be fatal to Putnam's purpose.

Putnam wishes to reject scientism without falling into the irrealism or extreme relativism that has been espoused by skeptics such as Rorty. To that end, he seeks to refute Goodman's arguments for irrealism[6] and

[6] See above, chapter 7, section 4, where Goodman is quoted. Putnam distinguishes two arguments Goodman makes, handily rids us of the first but has more difficulty with the second, which turns on there being mutually incompatible ways, within physics alone, of describing what we are tempted to call not only the same world but the same facts about that world. One cannot simultaneously treat spatial points as concrete and as limits, yet Putnam shows in convincing detail, with respect to that example and a similar one from quantum mechanics, that incompatible ways of talking enable us to describe the same 'states of affairs'. The trouble is, that does not address Goodman's argument, that it makes no sense to talk of 'the same' in these cases, as we have no neutral way of designating what is allegedly the same. Putnam is aware of this defect, as his denial that he is introducing any 'transcendent ontology' of states of affairs (1992, p. 117), makes painfully clear.

then takes aim at Derrida's exercises in 'deconstruction' (1992, ch. 7). As against Goodman, he maintains that there is one reality and that it does not consist of the ways we represent it; as against Derrida, he maintains that there is reference and truth, for example, in physics; but as against the scientistic, Putnam argues that physics is not the only 'language game' in town. Now, all of that would be unproblematic if the other language games could be set aside as noncognitive, as in the emotivist theory of ethics. But, in a sensitive discussion (1992, chs. 7 and 8) of Wittgenstein's lectures on religious belief, Putnam declines the gambit of noncognitivism. Even relativizing 'reference' and 'truth' to different language games (each thus claiming its own form of 'cognition') is rejected. As with Wittgenstein, language games are identified with 'forms of life', among which we find the moral and the religious. The suggestion, gently and tentatively made, is that each game has its meaning in relation to its respective form of life, and that we would be ill-advised to abandon all the others for the scientific life alone.

But where does this leave us? These forms of life are not sealed off from one another. It is essential to Christian doctrine that Jesus was a historical figure, to the Jewish faith that Moses was, to Buddhism that the Buddha was. Moreover, if doubt is essential to faith (Putnam approvingly quotes Kierkegaard to this effect), then there must be some connection, however dialectical, between religious discourse and factual judgment (commonsensical or scientific); both must in some sense be about the same world, otherwise there could be no cause for religious doubt and no call for faith. Nor, for that matter, would a miracle be a miracle if it did not take place in the natural world. Yet it is not clear how Putnam would explain the ways in which these different forms of life relate to one another. His theory of reference accounts by causal relation alone for different ideas' sameness of reference. It omits the possibility that people with different ideas can share an idea of the referent they have in common. Hence, it implies that they cannot agree about what it is

Painfully, because the denial is not persuasive. Here, Peirce's semeiotic can help, and also his concepts of hypostatic abstraction and vagueness (in the sense we have named 'inspecificity'; chapter 10). For example, if we examine the two ways of talking about spatial points – not in the abstract but in their application to concrete cases – then we can distinguish the indexical aspects, of either from their symbolic aspects, and in the neutral because inspecific terms of common sense, we can note that reference is being made by each to the same things. However one chooses to think of points, it is from 'there' to 'there' that the rock, meteor, or atom moved. Vagueness and indices can do the job: transcendence is put out of work.

that they disagree about (chapter 10, section 5) – or, if not disagreeing, that they apprehend differently. Yet the scientist, the God-fearing, and the artist (who sometimes are the same person) know that it is the same world that they apprehend in different ways.

The problem is exacerbated by Putnam's wholesale rejection of evolutionary explanations of intentionality (1992, ch. 2). He cites Millikan, among others, and expresses their view succinctly this way: mental representations are 'data structures' that have the function of corresponding to physical realities; they have that function, since the capacities to form them were selected in evolution just because the said correspondences were regular enough to be useful. ('Data structure', notice, includes blips on frogs' retinas that correspond often enough to flies to account for why the optical apparatus that produces them was selected.) Among other problems with this view, some of which we have already considered, Putnam asserts that it defines representation too narrowly: it fails to account for reference formed in consequence of cultural evolution. Putnam therefore sees these explanations as a form of scientism – as yet another example of the assumption that every facet of human life reduces to something that can be understood in the natural sciences. But is he then to say that our moral and religious and aesthetic natures have no explanation, or only a nonscientific explanation? As Pauline escaped one peril only to find herself in another, it appears that Putnam has escaped reductionism only to fall into dualism's clutches. It keeps us all eager for the next episode.

Putnam's critique is not entirely fair to Millikan, who, despite her persistent use of what she calls 'biological categories', expressly attributes 'proper functions' to artifacts and to human languages, as well as to things biological (pp. 28–32). Fair or not to Millikan, it in any case does not touch Peirce's semeiotic. For the latter differs from the evolutionary semantics of Millikan, Papineau, et al. in this respect: it does not make significance, in general, to be a function, whether biological or cultural, nor does it make signs, in general, to have a function (i.e., a purpose), whether biological or cultural. (Any sign is such by possibly serving a purpose, but serving is not having: the distinctions drawn in chapter 4, section 6, are crucial.)

The evolutionary account that Peirce's theory implies is one, in the first instance, of animals' capacity to act purposefully and, hence, to interpret signs; in the second instance, of the capacity of some animals to replicate legisigns, signaling one another; and, in the third instance, of humans' twin capacities to interpret signs by replicating legisigns and to form new

legisigns. Legisigns and their replicas, alone, have functions (purposes), but before there were legisigns there were signs of other kinds. Furthermore, not all the signs that humans deliberately produce are legisigns or replicate legisigns; hence, not all signify by a rule of interpretation. The fine arts may exploit linguistic or other culturally instituted rules, but works of art do not signify by rule; it might be argued that it is their essence not to signify by rule, but to make us apprehend things in new ways. It follows that, on Peirce's account, meanings are in relatively few cases determined by biological evolution. What evolution explains, instead, is the capacity to find and to create meanings. These may serve a biological or a cultural or an idiosyncratic purpose.[7]

Peirce's synechism reconciles the idea of a unitary real world with the fact of life's, and especially of human life's, irreducible richness, complete with all the tensions, conflicts, and unsolved riddles thereof. For it neither reduces one thing to another nor declares chasms of incomprehension between them. Epistemologically: inspecific concepts are shared across fundamental differences of outlook, providing enough mutual understanding to at least make tensions, conflict, and disagreement possible. Metaphysically: continuity is a *via media* between reductionism and dualism. Physical systems, we have argued, have anisotropic aspects explicable statistically, not mechanistically (chapter 5). These statistical aspects are consequential, as much as or more than are the mechanical processes that underlie them. They are, in order of occurrence and complexity: growth of entropy in closed systems, self-sustaining and self-replicating order, natural selection, and purposeful action. Purposeful action is sign interpretation; signs and significance are defined in its terms. Thought is the interpretation of signs by signs, and thus we may identify mentality with a semeiotic aspect of the living body. The mind is no more reducible to biological functions than biological functions are to molecular action, nor is it any more capable of existing apart from the body than the body is capable of existing apart from its molecules.

There is nothing in this synechistic theory to preclude mental functions that serve other than biological purposes (chapter 5, section 7). There is nothing in it to preclude the possibility that these functions will diverge into distinct forms of life, coëxisting uneasily and yet inseparably.

[7] The idiosyncratic productions of an artist or a scientist or a statesman may serve a cultural or even a biological purpose, and signs of any kind may serve an idiosyncratic purpose. The idiosyncratic and its complex interplay with the communal tends to be lost sight of in sign theories that make all significance to be a function historically determined.

5. Consciousness and Subjectivity

What of consciousness and selfhood? Peirce's semeiotic presupposes neither but sheds light on both. Space remains for a few brief remarks only, rounding out the sketch we have begun. We shall draw on what Peirce said in a variety of writings, recalling that sometimes he changed his mind and sometimes he changed his language, and that often he stretched terms so as to emphasize continuities.

Physical particles react to one another; organisms react complexly to simple stimuli (Peirce, in the argot of his time, spoke of the irritability of protoplasm). Sensory organs are specialists in sensitivity; moreover, they organize sensation. The eye's lens projects a sharp and complex image on the retina, a pair of eyes and visual cortex gives spatial depth to vision, a pair of ears finds directionality of sound. Central control over peripheral organs – focusing the lens, moving eye and head – enables the organism to seek information, reacting to what is seen by seeing it better. At what point in this progression may we begin to speak of consciousness? Is the shark conscious of other creatures swimming about? Peirce wrote, in 1905, that 'to be conscious is nothing else than to feel' (EP2:368, cf. 5.492, 6.268) and, in 1902, that 'the synechist will not believe that some things are conscious and some unconscious, unless by consciousness be meant a certain grade of feeling (6.174, cf. 7.539–52). The term is extended beyond its ordinary meaning, recognizing that what we call 'consciousness' is continuous with other phenomena not ordinarily so named; but then gradations within that broader category are admitted.

In 1907, Peirce linked consciousness qua feeling to self-control: 'If this be a correct account of consciousness, i.e., of the congeries of feelings, it seems to me that it exercises a real function in self-control' (5.493). We might therefore expect grades of feeling to correlate with grades of self-control. In another paper circa 1907, Peirce enumerated several degrees of self-control, beginning with 'inhibitions and coördinations that entirely escape consciousness' (presumably automatic reflexes and the like, that occur independently of feeling). Next, there are instinctive modes of self-control, followed by 'the kind of self-control which results from training'. 'Next, a man can be his own training master and thus control his self-control. When this point is reached much or all of the training may be conducted in imagination'. That is, one can compare a contemplated action, or its likely consequences, to what a rule prescribes or proscribes. 'But next he may undertake to improve this rule', either by critical comparison to a higher principle or, ultimately, 'by reference

to an esthetic ideal of what is fine' (5.533, cf. Peirce's account of the normative sciences; chapter 3, section 1).

To become one's own training master, one must represent possible actions and their likely consequences. It therefore requires signs produced by oneself and subject to one's own manipulation. Such signs are diagrams and replicas of legisigns. These are the medium, as well, of higher levels of self-control, as we trace logical consequences of principles and react emotionally to detailed pictures and complex narratives of what might be. As felt, such signs are the stuff of consciousness in these grades. What the neurophysiology of imagery and calculation may be is another question.

Representation of one's possible actions includes self-designation: 'Were I to do this, then what would happen?' Thus one's self appears within the field of consciousness. What prepares the way for such uses of 'I'? It must be something beyond the experience of bodily power, as that is shared by dumb brutes. In 1868, by way of combatting the Cartesian supposition that we know our own minds intuitively and other minds and the physical world only on that basis, Peirce alluded to Kant's observation that children are slow to learn the use of the first-person singular. They 'manifest powers of thought much earlier' (W2:201), hence, before they come to self-consciousness. Even before self-consciousness, they learn to rely on others' testimony as to the facts more than on their own observations of them (W2:202). They accept correction of their own judgments and they accept instruction. The ideas of false appearance and of ignorance emerge at that point, but falsehood and ignorance need a place. The self is introduced by hypothesis to accommodate them: 'Ignorance and error are all that distinguish our private selves from the absolute *ego* of pure apperception' (W2:203).

That has seemed to many to be too negative a concept of the individual ego. Richard Bernstein complained that 'such a conception of the self makes a mockery of the ideal of individual self-control. . . . If my separate existence is manifested only by ignorance and error, if I differ from my fellow man only by being a negation, then "where" and "what" is the "I" that controls and adopts ultimate ideals?' (1971, p. 198).[8] Notice, however, that this 'I' is where the mystery lies. The process of self-control is more open to inspection than is the ego that, on Bernstein's assumption, 'does it'. We can observe a child learning to bring its actions and

[8] Bernstein's 1965 essay, which contains a similar passage, is a subtle, insightful, invaluable exposition of Peirce's idea of self-control.

its thought under control. General Washington's staff watched him mastering his temper. To suppose that the adoption of ideals can only be explained as something selves do is to insist that a fairly intelligible process be explained by a mystery.

It is rather the other way about: we are to understand selfhood as an achievement, as something that comes about through ascending grades of self-control and, ultimately, by the adoption of ideals that orchestrate one's existence, that make its parts into parts of a life that can be affirmed. Of course, there has to be something individual that does the choosing and the adopting, but that is in the first instance the organism, the biological creature. Later, it is the partly formed person – that same organism acting with a degree of self-control – that has come into existence through prior acts of self-control. The person who seeks an ideal to adopt is one who still feels the anxiety of inconsistency, of being one day this person and the next day a different one, pursuing first this goal and then that one. Like Pinocchio easily led astray, he is not yet quite real, but desperately wants to be.

First, self-control – and *then* a self. Here we may draw on the passages from 1892 cited earlier (note 1) in which Peirce wrote of the personality as a 'coördination or connection of ideas' (6.155), which is also a 'teleological harmony in ideas' (6.156). Different harmonies are possible; furthermore, the construction of one's personality is never finished, and (though Peirce did not say this) it is always subject to the vagaries of others, in relation to whom it is formed. As teleological, this harmony of ideas is general and pertains to the future: it will be made more specific and otherwise modified as time goes on. 'This reference to the future is an essential element of personality. Were the ends of a person already explicit, there would be no room for development, for growth, for life; and consequently there would be no personality' (6.157; cf. Hausman 1993, pp. 173–7).

Vincent Colapietro, also responding to Bernstein, points out that the negations that distinguish individual selves do not thereby constitute those selves (Colapietro 1989, pp. 73 and 77). But what *are* selves? If selfhood is achieved through self-control and if self-control at its higher levels consists in diagramming possible actions, formulating principles, reasoning, and forming signs of oneself, then selves have no existence apart from semeiotic processes of certain kinds. That is also how Colapietro understands Peirce's theory. He proceeds to argue that the self thus understood is nevertheless an agent and a subject of thought and experience (1989, chs. 4 and 5). How is that possible? The air of paradox is

removed, I think, by recognizing that the flesh and blood body remains at the bottom of it all. Selfhood or personality is an aspect of the organization of that body's behavior. The organization in which selfhood consists is irreducible to the laws governing the operations of the body's parts. But it is still the body that acts and suffers. Colapietro (p. 95) brings together a number of passages supporting this gloss, for example, ones from 1868 that refer to 'a real effective force behind consciousness' or 'the physiological force behind consciousness' (W2:226, 226n3).

The selfhood of selves, that makes animate bodies into persons, is abstract. It resides in signs that can be shared among individuals, and most of all in legisigns. A legisign can be replicated repeatedly and by different individuals. Feeling, too, is inherently sharable; though it is not abstract, it has another form of generality (chapter 3, section 5). (Phaneroscopy presupposes shared thought and feeling, as otherwise it would be solipsistic; chapter 3, section 2.) Peirce appears to have thought of this doctrine as one whose truth is pervasively evident in everyday experience but distorted or obscured by philosophers' theories. Thus he delighted in outré expressions of it, as if challenging the reader to 'Deny this if you can!': 'the mind is a sign developing according to the laws of inference'; 'What distinguishes a man from a word? There is a distinction, doubtless. The material qualities . . .'; 'consciousness, being a mere sensation, is only part of the *material quality* of the man-sign'; 'the word or sign that a man uses *is* the man himself' (1868: W2:240–1). But even one's 'material quality' is not absolutely individuating: 'My metaphysical friend who asks whether we can enter into one another's feelings . . . might just as well ask me whether I am sure that red looked to me yesterday as it does today' (1903: 1.314). '*Esprit de corps*, national sentiment, sym-pathy [sic] are no mere metaphors' (1892: 6.271). Peirce likened the individual self to a society and, conversely, a society to a person: 'a person is not absolutely an individual. His thoughts are what he is "saying to himself," that is, is saying to that other self that is just coming into life in the flow of time'; but, by the same token, 'the man's circle of society. . . is a sort of loosely compacted person' (1905: 5.421). He called the philosophical theory of egoism, 'the metaphysics of wickedness': 'your neighbors are, in a measure, yourself. . . . the selfhood you like to attribute to yourself is, for the most part, the vulgarist delusion of vanity' (c. 1892: 7.571).

But all of this anti-egoistic emphasis gives greater substance to Bernstein's objection. The importance of Colapietro's 1989 book, which is about Peirce's semeiotic theory of the self, is to have shown how anti-egoism may be reconciled to ideas of personal autonomy and of the

'inwardness' of experience. The key is Peirce's account of self-control as depending, in its higher grades, on feeling and, especially, on experiments carried out in the imagination. The feelings are those in which images and fancies and unspoken discourse consists; they are signs. Colapietro writes, 'This capacity to withdraw from the public world is, at bottom, the capacity to refrain from outward action. . . . [I]nward reflection is the indispensable instrument of human rationality . . . voluntary inhibition is the chief characteristic of human beings' (p. 115). He quotes Peirce's 1907 remark (from the MS318 of which we have made so much) that 'Every sane person lives in a double world, the outer and the inner world, the world of percepts and the world of fancies' (5.487). Later in that passage, Peirce wrote that 'fancied iterations' in the inner world 'produce habits'. Inwardness is a semeiotic phenomenon.

In Colapietro's account of it, inwardness issues in autonomy. But autonomy is publicly observable in the pattern of a person's actions, in their evident control by principles and by a steadiness of purpose. Thus the idea of inwardness is separable from any Cartesian assumption of intuitive self-knowledge and from any assumption of impenetrable privacy. Inwardness makes privacy – hidden motives, unspoken thoughts, deceitful plans – possible, but it does not depend on privacy, nor does it presuppose a self essentially isolated from all other selves.

There is, still, a puzzle about the self. We use the word 'I' ambiguously, sometimes denoting oneself as physical ('I went down to the Peiraeus yesterday') and sometimes denoting an entity distinct from the body ('I moved my arm'). The first usage we can understand in terms of the theory being developed here: that selfhood consists in a semeiotic grade of an organism's self-control, so that references to the self are references to that organism as so controlled. The second usage, by contrast, seems responsible for a particularly egregious bit of folk psychology, the ghost inside the machine. One might think of eliminating such locutions from scientifically respectable discourse. But there must be some good reason why we talk that way. Peirce, I think, had a subtler view of it.

In the same manuscript in which he wrote of the grades of self-control, there is a long digression on hypostatic abstraction (5.534). The connection is that 'thinking is a kind of conduct . . . itself controllable . . . by thinking about thought'. To think about thought is to make what is thought about something into a subject of thought. That which is thought about X is predicated of X. To make that thought into a new subject, Y, about which we may think, is to turn a predicate into a subject. But turning a predicate into a subject is hypostatic abstraction (chapter 10, section 1).

The higher grades of self-control thus depend on a specific linguistic strategy, that of hypostatic abstraction.

Now, self-conception began, on Peirce's account, when the child first discovered error and ignorance. To discover error is to turn what is thought into the subject of another thought, forming a sign of a sign: 'That's wrong!' The abductive introduction of the self, as *that which* is in error and ignorance, is another hypostatic abstraction, an abstraction from an abstraction. In mathematics and in physical science, as we have seen (chapter 10, section 2), abstractions are made from abstractions; no less so in self-consciousness. So, what 'I' denotes, primarily, is a hypostatic abstraction. But, unlike opium's dormitive virtue, the ego is not an entity abstractly defined that may subsequently be identified with something concrete. There is no basis, in this usage of 'I', on which to identify the ego with anything physical. The ego is an *ens rationis*. For that reason, I am not identical with my body, not even my body qua self-controlled. But neither am I anything apart from what my body does. Like any *ens rationis*, an ego's being consists in facts about other things, in this case, facts about the control, by rules and ideals, of the physical movements of a particular human body. The essential point to notice, however, is that these facts would not obtain but for that abstraction having been made. For representation of self is an essential step taken in achieving higher levels of self-control.

We noted earlier that *entia rationis* can be consequential and therefore real. Statistical averages matter in social, political, economic life; differences in temperature matter, and temperature is a mean; centers of gravity, voltages, and so on, are all *entia rationis*. What could be more consequential than these? Perhaps only persons. What makes a greater difference in the world than persons? That does not mean that persons are entities that exist independently of bodies; as *entia rationis* they cannot exist independently. But it does mean that persons are realities irreducible to anything explicable by the laws of physiology, much less physics. What persons do can be understood only at the level of personal and social existence, in terms of their purposes, their thoughts, and the forms of social, political, economic, religious, aesthetic, scientific, and moral life into which they enter.

12

The Structure of Objectivity

The practice of knowing precedes its theory. It could not have been otherwise. How, for example, could we have known to rely on the evidence of our senses, before having already done so, with some success? Because practice precedes theory, theory has some evidence, of what works and what does not, to build on – even though what counts as 'working' or as 'success' is part of what is at issue. It is the thesis of this chapter that experience's precedence to principle never ceases, that there is no point after which methods cannot be modified by the results of their employment. More modestly, the thesis is that that was Peirce's view, in which he anticipated contemporary antifoundationalism. But he was in advance not only of his but of our contemporaries, as he knew how to avoid the subjectivism or relativism, the irrationalism and irrealism to which antifoundationalism is popularly supposed to lead. A further thesis of this chapter, then, is that objective inquiry is possible sans foundations and, indeed, in no other way.

Peirce did not speak of foundations in the present sense. He sometimes referred to his own work as laying foundations for further inquiry, but he did not mean an irrevisable foundation (chapter 3, section 1). The term 'antifoundationalist' not yet having come into vogue, he called himself a 'fallibilist', meaning that he regarded none of his beliefs as being immune to possible revision. Also, he eschewed the word 'objective', as in his day its use evoked what he took to be the false dichotomy between subjective experience and objective existence. Where we might speak of an inquiry as being objective, he spoke of it as converging over the long run on a single answer to a single question. We shall nevertheless discuss his ideas in the argot of today. It will take two sections to define the key terms and

set up the question, how objectivity is possible without foundations. We will then examine Peirce's concepts of science and of truth, as it is in their terms that his answer to that question must be sought.

1. Antifoundationalism

One might suppose that we can reflect on our cognitive practices, discern their *raison d'être*, and on that basis purge them of irrelevant accretions and accidental limitations, prove their principles, and free them of error. So, in the seventeenth century, thought Francis Bacon and René Descartes. They were guided by the same metaphor: to rebuild the edifice of knowledge on new and solid foundations. Hence the term 'foundationalism', though that was coined only in the twentieth century. Seventeenth century-foundationalism appears to have had two motives. One, that of achieving autonomy, is moral and came to the fore in the period of the Enlightenment: through self-consciousness, we take control of the principles of our own practice, freely choosing them for reasons we clearly apprehend. The other motive, to achieve certainty, to avoid all risk of error, is epistemological and was clear from the beginning. The idea of the foundationalists was that inquiry should be truth-conserving, that it should start out with none but truths, whether truths of reason or truths of observation, and be so structured that no error may be allowed to slip in at any subsequent step. Any false step, they feared, will make all subsequent results uncertain.

Antifoundationalism, the denial that there are or ever can be secure foundations of inquiry, is the prevailing view today. It has diverse sources, of which we shall notice a few. One, alluded to earlier (chapter 10, section 3), is that every observation presupposes a theory, even if the 'theory' is only common sense. It follows that there can be no rock-solid empirical data on which to base or against which to test our theories. Peirce anticipated that view. Let us take a few paragraphs to set out his argument.

To begin with, the data against which we test a theory, or that otherwise serve as premisses of inference, must be in the form of judgments (or statements, etc. – something propositional), not sensory images, much less sensations. For a test entails the possibility of contradiction, and contradiction is a logical relation obtaining between propositional items only. Inductive support for a conclusion is also a logical relation, albeit nondeductive, and thus can be provided by propositional items alone. Now, in his 1903 Lectures on Pragmatism, Peirce said that a 'perceptual

judgment' is the 'first judgment of a person as to what is before his senses' (5.115). It is perceptual judgments, then, that provide us our data.

Those judgments, being first, are not derived from or based on other judgments; they are not conclusions drawn from premisses. Nor can they be justified by comparison to what, in a nonpropositional sense, we sense. Peirce denied that a judgment bears any resemblance to the sensory image, or percept,[1] that elicits it (5.115). For perceptual judgments, though they are of particulars, import general ideas – 'contain general elements' – 'so that universal propositions are deducible from them' (5.181).

Those general ideas are not assumptions made in making observations, necessarily; rather, they are assumptions introduced by the perceptual judgments themselves. They are built into the meanings of the terms used. For example, to observe that the stove is hot is to imply that others can feel it to be so as well, that it can warm cooler substances placed on it, and so on. For that is what 'hot' means; the use of that term implies that there is such a thing as heat, which behaves in predictable ways. And thus our observations have testable implications. Without such assumptions, data would not be data: they would be insignificant, having no implications.

But whence these ideas, if they bear no resemblance to the percept? They are introduced by conjecture, as hypotheses. Now, Peirce identified a form of inference by which hypotheses are introduced, sometimes naming it 'abduction' (chapter 10, note. 4). The general pattern of abduction is this: at any given time, our minds are stocked with a variety of ideas of kinds of things and how they behave; we confront a puzzling phenomenon, represented in the judgment, E; in a flash, we bring the thought of E together with an idea of what, C, would explain it; E, and E if C, are the premisses of our inference; the conclusion is C; that conclusion does not follow deductively, for E may have a different explanation instead, or none; that C obtains is a hypothesis only. Perceptual judgment, however, is not the conclusion of an inference.

Peirce held that perceptual judgment is the result of an abduction-like process in which experience elicits ideas by which it is interpreted,

[1] To avoid the suggestion of its being intended to represent something, Peirce referred to the image as a 'percept' (7.615–77). Interpreted in a perceptual judgment, a percept becomes a 'percipuum'. Unlike a percept, a percipuum can be erroneous. Error aside, a percipuum *is* the perceived world our judgments represent, though only insofar as that world appears to us in the percept. To claim that we can compare our judgments with 'the facts' and that 'what we see' justifies our observation of it, is to get things backwards; it is only with the percipuum – the percept as judged – that a judgment can be compared.

but the experience is not first formulated in a judgment, E, before it is interpreted, C. C is the first judgment. Thus, the process is not deliberate; its principles are not willfully employed. 'Abductive inference shades into perceptual judgment without any sharp line of demarcation between them; or, in other words, our first premisses, the perceptual judgments, are to be regarded as an extreme case of abductive inferences, from which they differ in being absolutely beyond criticism' (5.181).

Perceptual judgments are beyond criticism at the moment of occurrence; for they occur uncontrollably, given one's attention and what one is prepared to observe. They are acts, sometimes involving concerted effort, but the result is involuntary. I may choose to look; I may choose what to look for, say, a color; there are just so many colors that I can name; but I do not choose what I see, that is, which color name in my repertoire I replicate.

Those judgments, however, may be criticized subsequently, when they or propositions deduced from them can be compared with other observations (Bernstein 1964). When the deliverances of sense are found to conflict, then adjustments must be made. We look for a flaw in the process: carelessness in observing or astigmatism in the observer or an optical illusion in the circumstances, and so on. Finding none, we may conclude that there is some error in the assumptions introduced. In that case, it is our theories (possibly including common-sense beliefs) that must be revised.

But any such revision alters our capacity to judge perceptually, as it alters the stock of ideas by which we can identify what we observe. The empirical basis on which theories are supposed to rest thus turns out to be vulnerable to changes in theory. We have built on sands prone to shifting, and not on solid foundations. Notice: perceptual judgments are revisable not because they are superficial, but because they are not. They do not record mere sensation or the mythical subjective episodes of the 'sense data' theorists; rather, they purport to represent their own physical causes.

Another source of antifoundationalism may be approached by reflecting on Quine's curiously truncated naturalism.[2] He suggested basing epistemology on psychology and related sciences that would explain, for example, how physical processes impinge on our nervous systems so that

[2] Quine's slogan 'epistemology naturalized' applies as well to Peirce, who anticipated Quine's view that theory of knowledge can be responsibly pursued only on the basis of what we know, and not out of a feigned ignorance (chapter 2, section 3); but Quine went further and identified epistemology with natural science, and in that respect Peirce's view differs.

we are able to glean information about the physical world (Quine 1969, ch. 3). In response to objections that he scanted the normative questions that had been at the heart of theory of knowledge, he later asserted that there is still a place for 'normative epistemology', namely, as 'a chapter of engineering: the technology of anticipating sensory stimulation' (Quine 1990, p. 19). The idea is that we can establish, on the basis of past experience, rules for how to go about developing theories that yield accurate predictions. That science's aim is to establish theories yielding accurate predictions was, for Quine, neither an assumption nor normative: 'But when I cite predictions as the checkpoints of science, I do not see that as normative. I see it as defining a particular language game. . . . the game of science, in contrast to other good language games such as fiction and poetry' (ibid., p. 20).

Predictivity, however, was not an ideal in Aristotle's conception of science, which is instead taxonomic: one kind, S, is understood when it is made to stand under a broader kind, P – 'All S is P'. And yet taxonomy was not intended to be fiction, nor can it be mistaken for poetry or, indeed, for anything but an attempt, just as modern science is, to explain, understand, or comprehend phenomena and to know the truth about the world. To be sure, by definition, a theory is empirical only if it is intended to agree with observation. But agreement can be in any of many different forms, depending on the kind of theory; accuracy of prediction is but one of these forms. And whether predictivity is a measure of understanding is another question: what counts as understanding or explanation? What kind of (empirically true) theories should we seek?

A theory is a representation, but there are many kinds of representation and as many kinds of understanding or explanation: taxonomies, mechanical models, line diagrams in geometrical optics, bar graphs and flow charts in economics, differential equations, statistical regularities and probabilistic laws, narrative histories or genetic explanations in parts of geology and biology, functional explanations, and so on. Not all of these are closely tied to predictive power. And we do not always feel that such power equals explanation: do Feynman diagrams in quantum electrodynamics explain anything or are they merely very effective devices for calculating predictable results from known conditions? These are normative questions, and not successfully swept under a rug of definition – a strategy that is in any case inconsistent with Quine's rejection of the analytic/synthetic dichotomy.

The preceding examples are drawn in part from the work of Thomas Kuhn (1970 [1962], 1977) and Stephen Toulmin (1961), who showed

that the growth of theoretical knowledge has been intimately connected with a corresponding evolution of method and aim. 'Aim' is Toulmin's word, though he also refers to the kinds of explanation science aims at as 'ideals of natural order'. Kuhn spoke of 'values' (1977, ch. 13). Aims, ideals, values: this raises a problem that Quine evidently was trying to evade. Aim and method determine the rules by which we are to choose between competing theories. If those rules are up for grabs, too, then how can theory choice be justified? How can a rule be justified? Modern science began by banishing values, as well as the experience of colors, tastes, and odors, to the realm of the merely subjective (in that sense of the word Peirce avoided). It is an embarrassment to have to admit that scientific method itself depends on values – worse yet, on values that have changed and might change again. Is science subjective, at bottom?

A third source of antifoundationalism, combined in Kuhn's thought with the preceding historical approach, is the assumption of 'meaning holism', the doctrine that a word's meaning is determined by its place in a web of words, or that a theory gives meaning to the terms that occur within it. It follows that a term occurring in two theories cannot mean the same thing in both – at least, not if the theories are opposed and there is no deeper theory from which each has borrowed its language. The stock example is 'mass' in Newton's and Einstein's theories, respectively. In one, mass is a quantity invariant with respect to the observer's inertial frame, in the other it is not invariant. It is therefore not the same thing and, rather than saying that Newton and Einstein disagree about the nature of mass, we should say that they are talking about different things. But if they are talking about different things, then they do not disagree. If they do not disagree, then there is no reason to deny that both theories are true. We may go on being Newtonians if we please, and by the same reasoning we would not be in error to adopt astrology or Druidism. We may do so, even while allowing ourselves to benefit from the technological advances of modern science. We can inhabit different worlds: one when jetting to the next Arch-Druidical conference, another when worshipping oak trees and committing human sacrifice.

In chapter 10, we mentioned some parts of the alleged problem of 'incommensurability' of theories; now we have all the parts in hand. As rival theories, at least if they are basic, have no meanings in common, there is no language in which to express a contradiction between them. For the same reason, they are about different topics; if each is about the world as a whole (as physical theories often are), then they are about different worlds. In addition, as the data against which theories are to be

tested and the standards by which they are to be evaluated change with changes in theory, there is no standpoint, neutral between opposing basic theories, by which to decide between them. In all these ways, basic theories are not commensurable: they have no measure in common, either as to what they mean, what they are about, what they are aimed to accomplish, how they are to be judged, or against what they are to be judged. One can learn to view 'the world' through each of several theories; one can thus in a manner compare the theories; and one might in consequence choose among them; but there appears to be no ground on which one could justify the choice.

The subjectivism that is supposed to follow from antifoundationalism is most often arrived at via this problem of incommensurability. In particular, Kuhn's 1962 book, *The Structure of Scientific Revolutions,* became notorious for seeming to entail relativism – a view of it on which his initial critics (e.g., Scheffler 1982 [1967]) and his many epigones agree, though Kuhn himself insisted that his relativism is limited and does not entail irrationality or irrealism (Kuhn 1991, 2000). You may object that all of this is a step back into the dark ages of the 1970s, that long-ago time. For the philosophy of science, growing ever more mercurial, has flowed into other channels since then. Yes, but that is not because the old issues were resolved; people merely got tired of discussing them. And the result is that an irrationalist or subjectivist view of science has been allowed to prevail elsewhere in the academy, for example, in so-called science studies (Bruno Latour is a prominent example; see his 2004 book) or, curiously, among literary theorists. Relativism puts political ideology beyond the possibility of rational critique; antifoundationalism has therefore found a welcome, in certain quarters. And, so, we shall dare to reexamine some 'old' issues.

2. Objectivity

Another source of antifoundationalism is the view of the idealist philosophers, going back, through Bradley, Bosanquet, and Hegel, to Plato, that the test of truth is not agreement with antecedent data but is to be found in a mutual coherence of ideas – one we hope to achieve at the end of inquiry. Principles and data, alike, are subject to correction as we try to fit them, and the conclusions we draw from them, together, in one system. That which fits in a maximally encompassing coherence is true; that which does not is doubtful, and if its negation fits or its explanation as erroneous fits, then it is false. Inquiry is truth-creating rather than

truth-conserving, and error, far from fatal, paves the path. Antifoundationalism thus formulated evades relativism. Or, at least, it enables us to proceed in faith that a final coherence is possible and, hence, that relativism is false.

Such a view is not limited to the purely ideational coherence associated with Platonic dialectic. It may be argued that the ideas that are maximally encompassing are those that permit the greatest variety and depth and exactness of observation. But observations can only be accepted tentatively: those that do not fit with a coherent body of theory-plus-observation must be rejected as erroneous, just as theories failing that test are rejected. And thus we arrive at Peirce's conception of reality, as that on the representation of which inquirers would eventually agree, though they come to it by different processes of thought from different kinds of experience (chapter 2, section 4). Peirce called that conception 'realist', but it was derived from the idealist philosophers.

The idealist theory of truth as coherence also had some influence, about two decades after Peirce's death, on the Vienna Circle philosophers. Hence Otto Neurath's and Karl Popper's watery alternatives to the foundation metaphor. Neurath wrote of science as a ship continually rebuilt afloat: theory sails where it may through seas of experience, picking up on the way the materials it needs. Popper wrote of science as a platform built over a swamp and supporting itself by driving down piles: ideas find their own support by making observations possible. But this shift, away from the foundationalist demand that we begin with true premisses, puts a premium on our adopting the right methods – those that promote self-correction, leading eventually to sustainable conclusions. The term 'objectivity' was therefore given new employment, as denoting a property of inquiry rather than a property of theories. Objectivity, as that term has come to be used, is distinct from truth. And that is how we shall now define it.

By objectivity of inquiry, we shall not mean freedom from error, or from values, or from passion, or from private motivations, or from social pressures and institutional constraints, or from guesswork. Instead, in calling an inquiry 'objective', we shall mean that it is of such a nature (conducted in such a spirit, employing such methods, with such data being available to it) that, were it continued indefinitely and by an indeterminate number of inquirers, uncoerced opinion would eventually converge, irreversibly, on a single answer to a single question; that answer may then be called 'true', and the changes made in opinion en route to it may then be called 'corrections of error'. Notice that some questions asked may fail to

be among those on answers to which opinion would converge eventually: the correction of error includes the elimination of mistaken questions. We can weaken the definition by deleting the requirement of sufficient data being available: then an objective inquiry may sometimes be such that it will never converge on an answer to a well-formed question, but would so converge were sufficient data available.

This definition allows for the possibility that error and passion and subjective judgment and group bias and the rest are essential to inquiry and, hence, to their own eventual correction. It also allows for the possibility that convergence on answers to single questions will occur within a diverging stream of inquiry in which a growing number of questions pertain to a widening array of topics. The definition entails that an inquiry, if objective in the stronger sense, is progressive, but it does not entail that progress is monotonic: there are many ups and downs on the way up a mountain. The definition does not violate the principle of fallibilism, as it does not entail that we can know, for certain, when we have reached the point at which agreement about the answer to a given question will never subsequently be reversed. Nor can we know for certain whether an inquiry we are engaged in *is* objective: we must hope that it is, and it can be so in fact, without our knowing for sure that, if continued, it would eventually converge on answers to at least some questions.

We have not defined objective inquiry as progress toward the truth (much less, as beginning only with true premisses): instead, truth can be defined as that toward which objective inquiry progresses or would progress were sufficient data available. That may be taken as a gloss on Peirce's famous 1878 account of truth as the final opinion: 'The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth' (W3:273). If it is so taken, many of the standard objections to that account are evaded.[3] Convergence is not the same as consensus, which may be coerced. As noted above, convergence of opinion is not smoothly progressive and it is consistent with unrestricted fallibilism and with the divergence of inquiry in other respects. There is no reason to assume that convergence on the answer to a theoretical question must be in steps progressively more exact, as in measurement.

[3] The authors responsible for such criticisms are numerous and notable, ranging from Ayer to Quine. I regret that I lack room here to do their arguments justice. For careful, albeit incomplete, discussions of the issues they have raised, see Skagestad 1981, ch. 5; Hookway 1985, ch. 2, and 2000, ch. 2; Misak 1991, 1995, ch. 3, and 2000, ch. 2; Delaney 1993; and Migotti 1999.

There is no implication that convergence must in fact be achieved in order for there to be a truth.

However, that truth is independent of actual convergence (and of there being data sufficient thereto) presupposes Peirce's later understanding of the reality of possibility (chapter 3, section 9) rather than his modal theory circa 1878 (chapter 2, section 9).

3. Peirce's Concept of Science

The understanding of modern science has lagged its reality by two or three centuries, maybe four. One instance of this is the very late date at which Aristotle's subject/predicate logic was superseded, in large part through Peirce's work, by a logic of relations. Aristotle's logic, restricted to monadic predicates, fits Aristotelian science, which is taxonomic. But Galileo's kinematics, to trace the origins of modern science no further back than that, is relational; the laws of motion relate distance to time, their ratio to time, and so on, one quantity varying as the other varies. All the laws of modern physics are relational. And that requires a logic other than Aristotle's for its analysis. Without a logic with which to represent the form of the new theories, it is unsurprising that Leibniz, Hume, and Kant forced them into the ill-fitting categories of a substance/attribute metaphysics, with absurd results.

The lapse of time between Galileo's kinematics and a formal logic of relations is, at its narrowest measure, about 240 years. Larger and more subtle differences between modern science and its ancient and medieval precursors took longer to grasp. Peirce was, I believe, the first to form an adequate idea of modern science. I will argue that his conception remains in advance of that which prevails today, even among philosophers of science. The subject unfortunately requires a brief review of facts probably known to the reader.

Aristotle's *epistēmē*, Greek for trustworthy knowledge, was translated by Cicero, in the first century B.C., as *scientia*, and that was the origin of the modern term 'science'. Aristotle conceived of *epistēmē* as a finished system of knowledge, and therefore science was identified with systematic knowledge. 'Science' did not begin to be used to denote a form of inquiry until the nineteenth century. So far as the kind of inquiry that came to the fore in the Scientific Revolution had a name, it was 'natural philosophy' or 'experimental philosophy'. Not that anybody thought inquiry unnecessary; Aristotle discussed its methods at length. But the idea was that it should be gotten through as soon as possible, so that we

can enjoy its fruits, which are systematic knowing. And that idea lingered during those centuries when, in practice, theories began to be adopted, not because proven true, but despite their obvious flaws and limitations, if they seemed to provide a basis for a future stream of discovery. After the Scientific Revolution, theory became a means of inquiry rather than its end, though at first no one seemed to notice.

To be sure, theories are still intended to be improved in the course of inquiry, and in that sense theory, that is, better theory, remains an aim of inquiry. But the supposition that theories could achieve final form in a finite time gradually faded away. More importantly, the conception of a good theory changed. That a theory be coherent and stand up to empirical tests is no longer enough. It must also be productive of further discoveries. Indeed, the promise of progress can trump defects of coherence and accuracy, which may be supposed to be temporary flaws.[4] All of this had a decisive consequence: the transformation of intellectual ambition, from wanting to attain a final truth (as shown by unhorsing every rival in the philosophical jousts), to wanting to contribute to an on going process, to be continued by others. Competition did not cease, but its rules changed. Proof of success is mostly posthumous. The significance of a theory lies in what can be done with it.

Yet it was not until 1840 that William Whewell coined the term 'scientist' for one engaged in this type of inquiry. Thus inquiry was recognized as a profession or permanent way of life, rather than a route traveled temporarily. But then the definition of science becomes problematic. How does that form of inquiry differ from other forms? A science can be distinguished by its subject or data, if these may be defined by current theory (astrophysics, for example, studies the chemical composition of celestial bodies by means of spectroscopic observation). That is fine for many purposes, but it begs our present question, as it presupposes that we know which theories are scientific. Without begging the question, there is nothing in subject or data by which to distinguish science from other inquiries, no matter how mundane. That would seem to leave aim

[4] The dynamic nature of modern science is still not generally appreciated; hence the tedious debates about biological evolution and natural selection, in which one side insists that neo-Darwinian theory has flaws (lacunae, unresolved problems, etc.) and the other side insists that it is as well established as any theory in science. Both claims are right and both are irrelevant. What is relevant is that research guided by neo-Darwinian theory is the only game going in large parts of contemporary biological science. As long as that theory is productive of further discovery, and no rival is proposed that promises to be at least equally fruitful, it will continue to define what science, in that department, is.

and method as the distinguishing features wanted. And indeed it is in such terms that science is usually defined today, as for example by Quine (section 1). But we have already seen why such definitions must fail.

In contrast to all those who would define science by its method, Peirce wrote in 1893 that 'the method of science is itself a scientific result':

> What is science? The dictionary will say that it is systematized knowledge. Dictionary definitions, however, are too apt to repose upon derivations; which is as much as to say that they neglect too much the later steps in the evolution of meanings. Mere knowledge, though it be systematized, may be a dead memory; while by science we all habitually mean a living and growing body of truth. We might even say that knowledge is not necessary to science. The astronomical researches of Ptolemy, though they are in great measure false, must be acknowledged by every modern mathematician who reads them to be truly and genuinely scientific. That which constitutes science, then, is not so much correct conclusions, as it is correct method. But the method of science is itself a scientific result. It did not spring out of the brain of a beginner; it was an historic attainment and a scientific achievement. So that not even this method ought to be regarded as essential to the beginnings of science. That which is essential, however, is the scientific spirit, which is determined not to rest satisfied with existing opinions, but to press on to the real truth of nature. (6.428)

Is this talk of 'spirit' too nebulous to serve the purpose? It is even more nebulous than it seems; for the scientific spirit cannot be defined in terms of 'the real truth of nature' that it presses on to. We shall see later (section 4) that truth must be defined in terms of inquiry. Hence, the spirit that issues in scientific inquiry must come first. Yet the force of the passage is precisely that nothing more definite will do. We are constrained by it to admit that there is such a spirit, of being dissatisfied with conclusions and 'pressing on', definite enough to account for the historical fact of what we today call 'science'.

Peirce's concept of science is broad, including history and philology and philosophy; but we have seen that already, in his architectonic.[5] The idea of science that he framed in the present passage does not entail that its methods are always such as to produce objective inquiry. However,

[5] His architectonic defines science even more broadly than that, as it allows for practical science. In the latter, one does not 'press on', after having attained a conclusion that is good enough for the practical purpose. Practical science nonetheless satisfies our definition of 'objectivity', which requires only that there *would be* a convergence of opinion *were* inquiry to be continued indefinitely. An engineer, testing the strength of a bridge, employs objective methods, but he does not wish his inquiry to be prolonged indefinitely. Cost/benefit calculations determine the cut-off point: more accuracy is not needed and greater certainty we cannot afford.

the clear implication is that the spirit that presses on will be dissatisfied with methods that appear not to be objective, that is, not to yield an inquiry that converges on answers to at least some questions. This is not a simple matter of either/or, nor is it one easily decided: methods and types of theory may vary in the number and severity of the tests that they make possible and, thus, in the degree to which they promote objectivity. Consider, for example, the Scientific Revolution.

Many have claimed, citing the writings of Francis Bacon, that the aim of modern science is practical, to produce labor-saving machines, cure disease, and so on, and that that is why it seeks mechanistic laws and makes experiments, eschewing the Aristotelian ideal of understanding. But inquirers' motives are more intellectual than philanthropic. Thus it would be plausible to speculate that the shift in aim and method was motivated by a purely intellectual desire for objectivity. For there are more, and more exacting, ways of testing a theory that yields quantitative predictions than there are ways of testing a qualitative taxonomy.[6] Of course, there have to be theories that do stand up well to the tests; an aim seeming impossible is soon dropped. Thus, initial skepticism about the new, mechanistic ideal of understanding was reasonable. Who but a genius like Galileo could have foreseen the success of the new methods? But once success was achieved, modern science became intellectually compelling, even to mediocre intellects, and not of utilitarian value merely. The greater testability of the new theories, as well as their greater concreteness and wealth of detail, are reasons, not at all practical, why no one can honestly reject modern science today.

Modern science is dynamic at two levels: its topic is how things come to be and how they behave; and its modus operandi prizes continued inquiry, if fruitful, over established theory. Peirce's account provides a third layer of dynamism, in which the modus operandi of science is seen as evolving under pressure from the endlessly dissatisfied scientific spirit. Part of that evolution is in the social form that science takes. The shift to quantitative methods and research productive of concrete results spawned specialization as its essential form, requiring, in turn, societies, journals, and the like, through which communication within and between specialties occurs. In other passages, Peirce emphasized the communal aspect of science (e.g., W2:270–2); the long passage quoted presents a broader view, within which a community of inquirers is seen to be,

[6] Nor is taxonomy omitted; but it is transformed by its conjunction with dynamic laws of change, as in the Periodic Table, or with genetic history, as in contemporary biology.

like method, as one of the products of science rather than as part of its definition.

4. A Fixation on Truth

The preceding concept of science, as a mode of inquiry evolving under pressure from its animating spirit, is expressed in Peirce's oft-read and least-understood article, 'The Fixation of Belief' (W3:248–57, 1877). Its meaning has been debated and its 'argument' derided by many who, in my opinion, mistake its form and its intent (Short 2000a). My gloss of it is, in brief, as follows.

Peirce described four methods of fixing belief, as if in historical progression, proceeding from tenacity to authority to the a priori method to the scientific method. Tenacity – sticking with what one believes, because one believes it – sometimes works in the short run but fails eventually. For we find that we cannot continue fixing belief that way; we are perturbed by others' disagreement with us. In its failing, we recognize a 'social principle' that makes others' opinions germane to our confidence in our own. So we turn to authority as guaranteeing agreement among many. Authority also works up to a point but fails. In its failing, we discover that the social principle is really a principle of nonarbitrariness or impersonality. We discover that what we really are after is a belief independent of anyone's will. The a priori method seems at first to satisfy that desideratum, but then it turns out that what some find it 'natural' to believe, others do not. We are left, finally, with no other recourse than to seek to fix our beliefs by making them to be determined 'by nothing human, but by some external permanency – by something upon which our thinking has no effect. . . . Such is the method of science' (W3:253–4). Here, 'science' is used narrowly, for a form that inquiry eventually takes, and 'method of science' is used broadly, for all the ways in which we might seek to subject belief to impersonal tests.

Peirce called the assumption of an 'external permanency' the 'fundamental hypothesis' of the scientific method (W2:254). It is the first expression of his later realism (chapter 7, section 4), the intelligibility of which could not have been defended before he developed his phaneroscopic categories of 2ndness and 3rdness, about two decades later.

The four methods have been criticized for being bad history (Murphey 1961, pp. 164–5), even though Peirce began 'Fixation' by citing some historical examples of the discovery of methods, as if to underline the fact that he did not intend his later talk of 'the' four methods to be

historical.[7] Clearly, they are an ahistorical idealization and are meant to be such. And Peirce has been criticized for making bad arguments against the first three methods and for the scientific method (ibid.; Ayer 1968, pp. 32–3; Hookway 1985, pp. 48–9). But the alleged arguments are so extraordinarily bad that it is implausible that they were intended to be arguments.[8] Instead, 'Fixation' describes a process of discovery, through successive stages of dissatisfaction, by which we (or the spirit that in 1893 Peirce named 'scientific') come to realize what would satisfy us. Thus we, as inquirers, discover our purpose, or what it is that we are truly aiming at. So far as a claim is being made, it is not supported by any argument but has to be tested against the reader's own imaginative reproduction of the process described: would *you* be satisfied by tenacity, and so on?[9]

Readers of 'Fixation' have been puzzled by its apparent contradictions. Peirce began by denying that inquiry aims at truth or at anything more than fixing belief (W3:248). Yet he concluded by affirming a method that entails an endless postponement of belief's fixation. Worse yet, such postponement is for the sake of leading us toward 'the one true conclusion' (W3:254). But those contradictions are not Peirce's. Rather, they illustrate the process he is describing, which involves redefinitions of aim. And that is the fundamental thrust of 'Fixation': to show that the aims of inquiry are not fixed but evolve – not arbitrarily, but in one predetermined direction. 'Fixation' was written against fixation.

The evolutionary theme of 'Fixation' can be restated as an account of truth. In the same essay, Peirce said that we call 'true' whatever we believe (W3:248), which, of course, is a truism. But if we may vary in the

[7] Besides, there is abundant evidence that Peirce had a deep interest in and an unusually large fund of exact knowledge about the history of science: Eisele 1985; Dauben 1995. He could not have thought his four methods were historical.

[8] Besides, Peirce was careful to ascribe virtues to each of these methods. Similarly, Kuhn analyzed scientific inquiry as depending for its success on individual subjectivity and institutional authority. It is an advantage that some individuals tenaciously stick to their theories against evidence seemingly contrary, and it is another advantage that institutions such as journals conservatively slow the flow of new ideas to an assimilable rate. Misjudgments and egregious excesses of course occur, but they do not affect the eventual outcome of inquiry, whatever the damage to individual inquirers. The a priori method remained a crucial element in Peirce's theory of science: he argued that we have an instinct for the right hypothesis ('a power of guessing right': 6.530) – only so could our otherwise improbable rate of success in selecting good hypotheses from among the innumerable alternatives be explained (1.80–1, 5.591–604, 7.219–20, and elsewhere; cf. Rescher 1978, ch. 3).

[9] This style of 'argument' is like that of phaneroscopy; chapter 3, section 2. See Smyth 1997, ch. 5, for a different defense of 'Fixation', one of great interest, though I think flawed (Short 2000b).

means we adopt for fixing belief, then it follows that our conception of truth will vary accordingly. By the method of tenacity, the truth is what *I* believe; by the method of authority, it is what *he* tells us to believe; by the a priori method, it is what *we* find it natural to believe (truth is what is 'agreeable to reason', as many philosophers have argued); and by the scientific method, it is what experience would eventually compel us to believe. Not until we adopt the last method do we conceive of truth as impersonal, as independent of what anyone actually believes.

This gloss is confirmed by a comment Peirce made circa 1906: 'My paper of November 1877, setting out from the proposition that the agitation of a question ceases when satisfaction is attained with the settlement of belief, and then only, goes on to consider how *the conception of truth gradually develops* from that principle *under the action of experience*' (5.564, emphasis added). So also, in the sequel to 'Fixation', Peirce wrote, 'Now, as we have seen in the former paper, the ideas of truth and falsehood, *in their full development*, appertain exclusively to the scientific method' (W3:272, emphasis added; cf. Migotti 1999, p. 84). These references to development seem to have been missed by most readers of Peirce.

It follows that Peirce's 1878 statement, quoted earlier – 'The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth' – was not intended to be a theory of truth in the contemporary sense. That is, it was not presented as an analysis of a concept that is presumed to be static. For Peirce did not hold that 'truth' was always defined in terms of inquiry's progress. Insofar as he produced an analysis of a static, unchanging concept of truth, it is the one with which, in 'Fixation', he began: we call 'true' that which we believe. For that will hold regardless of which method of fixing belief we adopt. But is that properly called a concept of truth, at all? Is it not rather the way the word 'true' is used, to which use one or another (or no) concept may be attached?

According to contemporary deflationary theories of truth, there is nothing to say about what the (unchanging) 'concept' of truth entails, except that *p* is true if and only if *p*. Where Peirce spoke of belief, deflationists often speak of assertion: to assert that *p* is true is to assert that *p*, and nothing more. Some add an account of why it is useful to have the predicate 'is true' in our language (e.g., so that one can generalize, 'Everything Quine says is true'); but that is another matter.[10]

[10] For deflationary and related theories of truth (minimalist, disquotationalist, etc.), see, e.g., Horwich 1990 or Soames 1999, ch. 8; Misak 2000, pp. 57–67, discusses Peirce in relation to these theories.

Some deflationists or near-deflationists allow for additions to the basic schema, for example, accounts of what warrants assertion, which may vary by case (Wright 1992). That is something like Peirce's account of methods of fixing belief. Thus we may define 'truth' in diverse ways by hypostatic abstraction, namely, as that set of assertions (or beliefs) that would be warranted (or fixed) by a given method. Peirce's account of how 'the conception of truth gradually develops' may be seen to be a variant of that strategy, stressing rough-grained difference in time rather than fine-grained difference in topic. He could admit the latter as well, of course.

In Peirce's view, the impersonal conception of truth is a cultural development. It is not simply what 'truth' means, and that's that. We can adopt the goal that it defines, or not. To suppose that Peirce's convergence theory of truth contradicts deflationism is to fail to see that the two theories address different questions. The deflationist theory attempts to elucidate a rock-bottom, unchanging conception of truth: what the word 'truth' means across all its employments. Far from disagreeing with that theory, Peirce anticipated it. His convergence theory, by contrast, elucidates a particular conception of truth, one implicit in scientific inquiry and that he wishes not only to elucidate but to recommend. If there is a real disagreement between Peirce and the deflationists, it is over which question it is more important to ask.

We shall proceed under the assumption that impersonal truth is our goal and that truths are opinions on which inquiry conducted in a scientific spirit would converge eventually. As indicated earlier, there is no guarantee, a priori or at all, that any inquiry is convergent, that is, objective. That objectivity is possible – hence, that impersonal truth exists – must remain an article of faith. Faith, however, has its intellectual supports, as follows.

5. How Theories Are Tested

If observations presuppose theories, then how can they be used to test theories? If observation O presupposes theory T and conflicts with theory T′, on what rational basis can we decide to retain T and abandon T′, rather than the reverse? Or do observations decide among theories in some more complicated way? This section is about how theories are tested empirically. Why empirical testing confers epistemic warrant on the theories that survive testing is another question, deferred to the next two sections.

To suppose that basic theories are incommensurable is to assume that a theory that does not provide all of its own meanings must derive some

of them from a theory more basic. A physiologist, for example, might use ideas from physics, of mass and voltage, in describing how the heart works. The picture suggested is of theories branching off from a single stem, all equally specific, with a one-way flow of concepts, presumably by capillary action, from the bottom to the top. That picture is destroyed by recognition of inspecific concepts (chapter 10, section 3), particularly scholastic hypostatic abstractions (chapter 10, section 1). Those concepts are neither lower down nor higher up; they do not fit into the picture at all. And yet it is they that introduce the topics about which theories, no matter how basic, may disagree; and it is in their terms that one may make observations that are germane to either of two opposed theories, even if basic.

We have seen in chapter 10 how that works; let us restate it more formally. Let a theory, T, be any general idea, even if inspecific, that entails logically contingent general propositions, including commonsense beliefs. And let T′ and T″ be theories properly so-called, that is, relatively specific ideas about that which is identified abstractly in the relatively inspecific terms of T. If T′ and T″ are different ways of making T more specific, then each entails T, perhaps with some marginal corrections, though T entails neither. Hence, their common subject matter can be formulated, vaguely, in terms of T, and T can be assumed in making observations germane to both. An observation is germane to a theory when it agrees with or contradicts a proposition that can be derived from that theory, usually in conjunction with other assumptions; let us speak of these derived propositions, with admitted looseness, as those that the theory entails. As T′ is a specification of T, some propositions that it entails, in its language, must have more vaguely formulated counterparts in the language of T. If T′ entails *p* and *p* has a vaguer counterpart, *q*, in T, then T′ entails *q* also. Suppose that there is an observation, couched in the language of T, of E in conditions C. If T′ entails that E if C, while T″ entails that not-E if C, then the observation decides between T′ and T″. For T″ cannot be affirmed while denying T; and, given T, we have an observation that contradicts something T″ entails.

That is often referred to as an *experimentum crucis*. Of course, crucial experiments are not so simple and definitive as that, since there are any number of ways a theorist may defend a theory apparently disconfirmed. If he cannot identify a flaw in the experiment or a misstep in the observation, he may question one of the other assumptions made. Thus the Duhem-Quine thesis, that theory choice is always underdetermined by the evidence (about which more below). And a theory may also be rescued

by amending it. Nor is the disconfirmation of T″ a decisive proof of T′. But one thing a defender of T″ cannot do is reject T.

The example we relied on in chapter 10, of caloric versus kinetic theories of heat, might be thought suspect, as each theory was framed in terms of Newtonian mechanics. In fact, that makes no difference to our analysis, which was of an observation Count Rumford made that presupposed nothing but a common-sense idea of heat. But let us take another example, one undeniably of theories fundamentally at odds, the Aristotelian and Galilean theories of falling bodies. To be sure, as Feyerabend and Kuhn have both emphasized, Aristotle and Galileo ask us to look at the world in very different ways; one cannot translate the claims of either into the language of the other. 'Falling' means one thing for Aristotle – the action of terrestrial matter seeking its natural place at the center of the cosmos – and another for Galileo – a modification of inertial motion. However, when a Galilean and an Aristotelian disagree about where an object dropped from the mast of a ship (moving at uniform velocity) must strike the deck, they are not talking past one another. For a particular instance of falling can be identified without assuming either theory, namely, in the inspecific terms of common sense. The Aristotelian, the Galilean, and an uneducated sailor can all agree about what happened. It is such instances that the two theories conceptualize differently. This obvious fact would not have been missed but for the assumption that every concept of falling must be specific and therefore incompatible with every other concept of falling.

We cannot avoid saying the obvious, when important thinkers have argued themselves into the ridiculous position of denying that Aristotelians and Galileans could observe the same thing, as if each had been a prisoner of his own theory, blinded by it to what others can see. Of course, it is much more exciting to deny the obvious than it is to affirm it. But the real interest of the topic is in identifying the assumptions, in part about language, that have led Feyerabend, Kuhn, and others into absurdity, with, as indicated earlier, disastrous consequences elsewhere in the academy. Those assumptions are more easily identified when we have an alternative theory in mind, in this case, Peirce's.

It might be objected that I have implied that common sense is irrefutable, contrary to fallibilism. But that is false. It is possible for an observation couched in the language of T to conflict with an implication, determined by T, of another observation couched in the language of T. If one or the other observation cannot be dismissed as due to some error of carelessness, defective instruments, and so on, then T cannot be

retained. Of course, with T's demise, the observations must also be withdrawn, unless they can be reformulated in vaguer terms. If T is common sense, we may be unable to say what refuted it; we would be rendered speechless; but refuted common sense would be. Suppose, for example, that the cool ends of iron bars no longer grow warmer as the hot ends cool, that fires felt to be hot now freeze instead of boil water, and that there has ceased to be agreement among our feelings of warmth. The very conditions on which the ordinary meanings of 'hot', 'cold', and so on, depend would no longer obtain and thus the observations just stated would have to be revised, we know not how. But certainly the assumption that there is such a thing as heat could not be sustained.

In Peirce's view, common sense is refutable in principle but is unlikely to be refuted, though it may be refined (5.498). Hence, it is relatively immune to theory change. And thus, granted the relevance of inspecific observations to theories more specific in meaning, common-sense assumptions are sufficient to provide observations against which to test theories.

In fact, such tests form but a small part of how theories are evaluated. Contrary to what Popper implied, experiment is rarely aimed at refuting conjectures. Most of it is devoted to applying, extending, and refining theories, for example, by measuring postulated constants. That is something Peirce well knew, as he earned his living doing exactly that kind of work. Imre Lakatos argued that theories are tested more by their relative success in being thus refined, extended, applied than they are by crucial experiments (1970, 1978). Rival research programs (Lakatos' phrase), informed by rival theories, are pursued simultaneously by different groups of researchers. Tenacity, authority, and apriorism all come into play. Choice is made eventually as one program flags while the other continues robust, so that money and personnel flow, for whatever private reasons, wither they will do the most good (a process sometimes distorted by factors extraneous to pure inquiry, such as political or commercial imperatives). Choosing the theory that grounds the more 'progressive' research program is rational even if the reason is not evident to the chooser. Choice may of course be revised. All of this goes a long way toward resolving the problem of underdetermination of theory choice by evidence, as it means that the choices individuals make, more or less subjectively, may differ and, in differing, contribute to the progress of inquiry, while the choice made by the scientific community as a whole (i.e., the choice in effect made when individual inquirers no longer differ in their opinions) does not have to be made immediately and, when

made, is made on the basis of a great quantity of evidence pertaining to many issues.

It is essential, not only to crucial experiments but also to Lakatosian research programs, that specific theories entail less specific ideas. It is never the case that measurements, observations, and so on, are made wholly in terms of advanced theory, as if common-sense beliefs were wholly in abeyance. Thus there is a fallback position available when theory fails: 'Well, we thought we would find evidence of gravitons in this experiment, but the blips on the screen are not consistent with those the theory predicts. We don't know what those blips mean'. That is why research is not question-begging. A scientific theory is never hermetically sealed in its own juices.[11]

6. Why Observe?

Why is support by empirical test any reason to accept a theory? One answer might be that observations come to us warranted and that they transfer their warrant to the theories with which they agree. We could maintain such a view even while admitting that observations are subject to correction in light of advances in theoretical knowledge. For we could claim that they have a *partial* warrant in sense experience: they are acceptable because based on sense experience but only if they cohere, through the medium of theory, with other observations. That is one formulation of the view known as 'weak foundationalism', the major contemporary alternative to antifoundationalism.[12]

Peirce is sometimes claimed to have had such a view (Delaney 1993, p. 89), but it seems to me that he did not (Short 2000b). As noted earlier (section 1), he said that perceptual judgments occur uncontrollably (5.115–6). Thus, at the time of their formation, the question of warrant does not arise. And when, on reflection, it does arise, it cannot be satisfied by citing those judgments' causes in sensory experience. A

[11] This is a remark about the nature of empirical inquiry in general; it has nothing to do with Niels Bohr's narrower and more interesting claim that the observation of quantum phenomena must be in Newtonian terms.

[12] Another alternative, similar to weak foundationalism, has been provided by the Peircean philosopher Susan Haack, who dubs it 'foundherentism' (Haack 1993), combining 'bottom-up' support from sense experience with coherentist 'top-down' requirements. Peirce's pragmatic theory of knowledge should be tested against Haack's arguments and against much other work in contemporary epistemology as well, especially Nicholas Rescher's recent work (e.g., 2001).

perceptual judgment, Peirce said, is based neither on inference from another proposition nor on comparison to a percept, and he suggested no third possibility. In another passage of the same date (1903), he wrote: 'The perceptual judgment professes to represent the percept', but 'there is no warrant for saying that [it does] . . . other than the *ipse dixit* of the perceptual judgment itself' (7.628). To say that it has no warrant but itself is to say that it has no warrant.

Our inclination to say that observations are warranted by sense experience is perhaps due to a failure to distinguish the question, 'What justifies this judgment?' from the question, 'What justifies our relying on this judgment?' For the latter might be answered not by citing an individual warrant for the judgment – that is, something particular on which it, individually, is based – but by citing the reliability of judgments of its type. We know from experience that perceptual judgments tend to be reliable – that is, that they tend to agree with one another – especially when they are formed under normal conditions (seeing in a good light) with due care (looking steadily with attention to the features at issue) by good observers (sober, sane adults or persons who have been trained in the relevant branch of observation). Hence, we are warranted in relying on such observations. Indeed, that warrant is built into the confidence with which we form perceptual judgments under certain conditions; and its absence is built into the hesitancy and hedging with which we make similar judgments under other conditions. Perhaps the confidence with which a judgment is made is mistakenly supposed to reflect the existence of an individual warrant for it.[13]

[13] Other sources of the assumption of individual warrant are worth examining. First, there is the common assertion, 'I think it because I see it', or words to that effect, as if the seeing were distinct from the thinking. But so far as the seeing *is* distinct from the thinking, it is a nonpropositional awareness (a percept) and – going over the same ground again – it can justify the thought neither logically nor by resemblance. Whereas, if the seeing is a seeing-that, then it is propositional and not distinct from the thought (at most, it is what Peirce named a 'percipuum', the percept as interpreted in a perceptual judgment; see note 1). Second, there is the grammatical similarity of perceptual judgments to evidentiary assertions. You assert, 'The cat is on the mat', and are asked, 'How do you know?' You might give a reason ('It always is') or cite someone's testimony ('Susan said it is') or cite evidence ('I can see its shadow'), any of which may be a premiss for an inference to your conclusion. But if you respond, 'I see that it is', you are not citing evidence but are merely repeating your claim in a form that indicates that no evidence is necessary (such responses are often made rather testily, for that reason). Finally, there is our experience of examining and reexamining what we perceive, forming and correcting judgments thereby. Suppose we see an owl in a tree. 'That's a barred owl.' 'What makes

Peirce once used the example of a blind man hearing a murder being committed and a deaf man seeing the deed done (W2:468–9). That, and not progressively more exact measurement, is his paradigm of convergence of opinion: the sensations from which each man began are as different as they can be, and yet the two agreed about the facts. That agreement testifies to the reliability of judgments of those types. Notice, however, that the agreement is about things that exist or occur independently of anyone's perceptual experience. Our experience of the reliability of observation is possible only because agreement is mediated by ideas of the physical world and other persons. If our judgments were only about the contents of sensory images, the question of agreement, whether between different observers or between observations made in successive moments by the same observer, could not arise.

This defense of relying on observation assumes that observations have been made and relied on over many thousands of years – and well before anyone ever thought of asking about their justification. But so they have. Our naturalistic view of the human mind entails the same. First the practice of knowing, then its theory.

In fact, the experience mankind has had observing the world provides a much deeper defense of observation's reliability than we have so far acknowledged. For, as well as much else, we agree on theories about the physical world and animal physiology and human psychology that explain how observation works. Therefore, we agree on how agreement in fact came about. The explanation entails that the observations of the physical world that agree reliably map general terms onto physical particulars of general types. Thus reliability means not only agreement among observations (or among observers) but also agreement of observation with reality. To be sure, that is reality *as* it is represented in our theories, and we base our theories on empirical evidence. But the circle is not so tight as it seems: for there is no guarantee a priori that empirical methods would yield an explanation of their own success.

Obviously, there are important gaps in that explanation as it currently exists, some of which Peirce's semeiotic addresses. Here are two. (1) How do perceptual judgments manage both to refer to particulars and to

you think so?' 'I can see the heavily streaked breast – no! they're shadows of branches.' The judgment in question, even if at first it was perceptual, comes to be based on or corrected by more detailed perceptual judgments. We form these judgments by looking carefully. But looking carefully is how we form reliable judgments; it is not a basis for the judgment.

describe them? Therein lies the nexus between observation and theory, the particular and the general: it is the home of hoary old philosophical puzzles. The heart of a Peircean analysis of perceptual judgment, yet to be made, would be his doctrine that a symbol's tokens are not symbols but are indices (chapter 8, section 2). When occurring in a perceptual judgment, such a token is an index of its physical cause yet elicits a response governed by the immediate interpretant of the symbol tokened; the latter interpretant, in its ultimate form, is a set of general dispositions, explicable in conformity to the pragmatic maxim, that are subject to growth and modification (chapter 10). (2) How is the 'agreement' of linguistic entities, such as judgments and theories, with physical and social reality to be understood? That dreaded monster, the correspondence theory of truth, raises its frightful head, but perhaps it may be domesticated – or, rather, naturalized – by Peirce's semeiotic taxonomy, in which there are signs of markedly different kinds, differently related by real relations to markedly different kinds of object. Correspondence may thereby be divided into kinds, and which one we think obtains will depend on our theories in physics and physiology, and so on, as well as on a semeiotic analysis of the judgments and theories in question. Truth-as-correspondence, in this version of it, is a scientific theory. But such applications of Peirce's semeiotic are work that remains to be done.

It will be said that, as all of these theories and explanations are scientific, they are 'internal' to science and cannot warrant the methods of science. But, as indicated a moment ago, it did not have to turn out that way. The contingent fact of science's success, in this and in other respects, argues for taking its theories realistically, that is, as representations of a world that exists independently of its being represented. Such an argument, sometimes named 'scientific realism', is not internal to science, as it turns on facts about the history of science. Earlier, we said that such a view is part of Peirce's realism and we defended it against the idealist objection that such realism is meaningless (chapter 7, section 4). The upshot is this: the convergence-to-date of empirical inquiry on certain theories gives one reason to suppose that there is a reality more or less as those theories represent it to be; among much else, these theories explain why science has thus succeeded; and, from that explanation, we can draw a prediction that further inquiry will, at a minimum, not entirely erase the results so far attained.

There is one problem, however. The idea that scientific inquiry is convergent has been questioned. Hence the next section.

7. Realism, Not Relativism

Peirce is easily made a straw man by attributing a simplistic idea of convergence to him. We have evaded some misconstructions. Convergence on the answer to a question can occur within a diverging array of questions; convergence does not mean that all of inquiry will come to a focus on a single point. Nor does convergence have to be like the approach of successive approximations to the exact measure of a quantity; it can be agreement on the facts by investigators who proceed from diverse starting points and employ different methods, as in the blind man and deaf man example. This idea of convergence is in fact stated at length and very clearly in no less prominent a place than the essay, 'How to Make Our Ideas Clear' (W3:273).

Hilary Putnam applied the label 'convergence' to a historical series of theories in which each contains its predecessor as a limiting case (1978, pp. 19–22). We obtain Galileo's law of falling bodies from Newton's law of gravity by ignoring the increasing acceleration during fall that Newton's theory predicts, and we obtain Newton's physics similarly from Einstein's by ignoring the very slight differences of their respective predictions within ordinary parameters. This idea of convergence is not Peirce's, however, nor does it bear that name as gracefully as does his. Theories containing and contained do not converge; strands of inquiry that lead from separate starting points to the same conclusion do converge.

Problems for Putnam's view nonetheless pose problems for Peirce's. Theories' empirical adequacy and instrumental value – measurements, predictions, technological applications – are commonly distinguished from theories' representative value, and Kuhn pointed out that cumulative growth in the first is not always matched by like growth in the second. For example, Newton's ideas of matter and space cannot be construed as a limiting case of Einstein's; they are too different. Or think of the history of theories of light: from Descartes to the present, there has been a growing body of empirical knowledge accompanied by a wildly diverse sequence of ideas about what light is (like a stick, longitudinal waves, stream of particles, transverse waves, wave-particle duality: the last of these by no means contains the preceding, as it is not a fusion of undulatory and corpuscular models, which cannot be fused, but is an abandonment of models altogether). But if there is no cumulative growth in scientific representations of reality, then what reason is there to suppose that inquiry is convergent in Peirce's sense, other than on the answers to instrumental questions? Putnam's argument for continuity of reference through

changes of theory unfortunately does not entail continuity of conception of the referent (chapter 10, section 5), and therefore it does not adequately rebut Kuhn.

Kuhn, who, contrary to what is often said, never denied that theory choice is, in the long run, rational, was limited to viewing it as a calculation of which theory is 'better for doing whatever it is that scientists do', whether that is to 'solve puzzles (my view [said Kuhn])' or 'improve empirical adequacy (Bas van Fraassen's)' (2000, p. 96, emphasis deleted; the reference is to van Fraassen 1980). The problems or the evidence will not be the same for different theories, but we can decide which theory, or research program, is doing better in solving *its* problems or garnering *its* evidence. Such comparison is essentially quantitative and is independent of any assumption that the two theories are about some of the same things.

Toward the end of his career, Kuhn wrote: 'the position I'm developing is a sort of post-Darwinian Kantianism' (2000, p. 104). The Darwinian allusion reflects Kuhn's view that scientific progress entails proliferation of specializations, with different kinds of theories and techniques, much as in biological evolution speciation occurs so as to fill every ecological niche. He continued: 'Underlying all these processes of differentiation and change [in our theories], there must of course be something permanent, fixed, and stable. But, like Kant's *Ding an sich*, it is ineffable, undescribable, undiscussable'. That is the sort of view Putnam had in mind when he defined 'metaphysical realism' as, among other things, the doctrine that reality 'transcends complete formalization in any one theory' (1978, p. 125).[14] Its alternative, he says, is the realism 'internal' to science, that reality is as science describes it. But that can only mean that reality is as it is described in current theory. And if Kuhn is right, then Putnam is mistaken in supposing that current descriptions will be retained, even with modifications, in later theories.

A choice between metaphysical and internal realism is forced on us by the assumption Putnam shares with Kuhn, that a concept is either specific or empty. Thus, either we have a specific theory of reality (or of heat, of light, etc.) – namely, whatever scientific theory is current at the time – or we have no idea of it at all. Either we assert that there is something of which we have no conception – a *Ding an sich* – or we limit our assertions of existence to the entities that current science postulates.

[14] The condition (a) mentioned but neglected in chapter 7, section 4.

Fortunately, we have discovered that this is an error: there are inspecific ideas that are not empty and that enable theorists who disagree to communicate (chapter 10). They also enable theories about complex physical systems to survive revolutions in microphysics. Consider an example. Theories of psycho-physical parallelism between wavelengths of light and sensations of color were originally formulated on the assumption that light is a transverse wave in an ether. The ether hypothesis gave way, and then, more radically, the wave theory was replaced by wave-particle duality. Quantitative laws linking wavelength to sensation to color judgment were unaffected by these fundamental revolutions. That is possible only because a less specific idea of light, as passing from one point to another and as carrying energy measurable as a wave, is common to those more specific theories and suffices for psycho-physical laws. (How light is produced and how it produces effects in retinal cells is another matter.)

We might reasonably suppose that another revolution in basic physics – say, adopting string theory – would also leave a great deal of the scientific superstructure (geology, meteorology, physiology, etc.) in place. But it is in terms of these latter theories that the reliability of observation is explained and predicted. It follows that we have all the convergence (in Peirce's and Putnam's senses, both) that we need to justify relying on observation.

Does it follow that inquiry, as long as it is pursued, will continue to converge, in Peirce's sense? I think not. Peirce rightly warned us not to regard any question as being beyond possible investigation (W3:274–5), but that does not mean that every question can be settled. Especially in the case of historical questions, evidence may be no longer available ('buried secrets' Peirce called them, in the passage just cited), even though we can never be sure with respect to any given question whether a way of investigating it will not be found some day. Unavailability of evidence is one possible limitation on inquiry. Others are poverty of imagination and failure of *il lume naturale*.[15] Peirce attributed our success in choosing, often enough, the right hypotheses to investigate, from among the vast number of alternatives, on our having an 'instinct for the truth' (1.80–1). That instinct, if accounted for by Darwinian evolution, may be limited

[15] The question of how to direct our limited research resources, so as to maximize the chances of success, is one that Peirce addressed under the rubric 'the economics of research'. Though I have neglected the topic here, it pertains to the objectivity of inquiry as much as does observation's warrant. For inquiry is not passive; it is an active interrogation of nature, and everything depends on asking the right questions. See Rescher 1978, ch. 3.

to the world so far as we have been adapted to it for practical purposes (5.591, 7.508). The very large, the very small, the very distant and very fast, may be very different from what we have an instinct for comprehending. 'As we advance further and further into science, the aid which we can derive from the natural light of reason becomes, no doubt, less and less' (7.220). Peirce in that manner anticipated the surprising nature of the new theories in physics that were being developed toward the end of his life, unbeknownst to him.

8. How Aims Are Tested

There is still a question about objectivity that clamors for attention. Peirce held that the methods of science are discovered and that its aim evolves. If that process is arbitrary, then science is subjective after all; for there would be no reason to expect eventual agreement about which questions should be answered, much less on what the answers are. Are there grounds on which choices of aim and method may be justified? That may seem an impossibly ambitious topic to address so late in this book, especially as Peirce said so much about science but so little about values and moral inquiry. In fact, we have hardly been talking of anything else. We come to this topic not with a paucity but with a superfluity of Peircean materials with which to address it.

First, there is his teleology, according to which some possibilities have a greater power of becoming actualized than do others (chapter 5, section 7). If that supposition is correct, then values have an objective existence, a reality, consisting in their power to win adherents. This power is final, not efficient; otherwise, the doctrine would reduce to that of Thrasymachus, that might makes right. False ideals are perennial temptations but are self-defeating or ultimately disappointing. The power to become established as ideals is, of course, not the same as the power to be achieved; we know, for example, that hopes for peace and justice are thwarted repeatedly, often because circumstances make them incompatible.

Second, there is Peirce's architectonic, in which he claimed that possibilities create their defenders through a contemplation of them in which they evoke feelings of unqualified admiration (chapter 3, section 1).

Third, in 'Fixation', Peirce showed, in an idealized way, how that process works – construed both as a choice of method and as an evolution of aim, that is, an evolution in our conception of truth. To that, we added the historical example of Galilean science being chosen over Aristotelian science (section 3). We saw that, in making that choice, the criteria by

which it was made were clarified and affirmed: inquirers discovered that severity of tests, concrete detail, and continued growth of knowledge are important to them. In the process, the concepts of science and of objectivity changed, not arbitrarily but irreversibly, for reasons that we can articulate now but that could not have been articulated, nor appreciated, before the choice was in process of being made (cf. Kuhn 1977, chapter 13).

Fourth and last, there is Peirce's mature semeiotic, which is remarkable perhaps most of all for its surprising combination of two features. On the one hand, it admits emotional and energetic interpretants as well as logical interpretants, but on the other hand, it embraces all interpretants in the same structure of objectivity. Significance is relative to interpretative purpose, but it is a factual question whether an interpretant serves its purpose. Thus, for every dynamic interpretant there is a final interpretant of which it may fall short, and for every immediate object there is a dynamic object from which it may vary. It follows that feelings and actions and the signs they interpret are as much subject to correction as are scientific theories. Further experience of that which we have found admirable, and of its alternatives, may alter our opinions, later emotional interpretants correcting earlier ones. By such 'collateral experience', as Peirce called it (chapter 7, section 3), we make distinctions between what appears valuable and what really is valuable. In cases where inquiry into value is convergent, the values are real; in cases where such inquiry appears convergent, we have reason to suppose that the values are real, just as we have reason to suppose that the physical world is real.

Obviously, the preceding needs to be worked out in detail (Short 2002b is a very tentative beginning, building on remarks by Richard Robin and David Savan). Peirce's semeiotic points toward extensive parallels between the interplay of emotion, action, and principle, on the one hand, and the interplay of sensation, experiment, and theory, on the other. (In matters of art rather than of morality, action is less salient.) Principle or theory can correct judgments prompted by emotion or sensation, even while the former are tested by their relative success in wresting coherence from the latter. Contrary to theories that make aesthetic and moral experience cognitive, but only by committing one or another version of the naturalistic fallacy, this parallel treatment entails that they are cognitive in their own right. As Goodman wrote, 'in aesthetic experience the *emotions function cognitively*. The work of art is apprehended through the feelings as well as through the senses' (1968, p. 248, emphasis in original). But we need Peirce's semeiotic to make sense of that claim.

This chapter has argued that objectivity in the natural sciences not only is paralleled by but also is dependent on objectivity in the formation of values. But values are objective only in being final causes, accounting for our adoption of them. It is fortunate that if Peirce is correct, teleology has been reintroduced by modern science. For otherwise there would be a deep incoherence in the claim of science to be objective: it would not leave room for the objectivity of its own choice of methods and aim.

9. Objectivity and Freedom

Foundationalists suppose that foundations are self-authenticating. Once cobwebbed thought is swept away, the foundations of inquiry will be embraced in clear-eyed vision of their rightness. They will be embraced freely, without obedience to any authority or in consequence of any blind compulsion. This is the Enlightenment ideal of freedom as autonomy. Autonomy is not merely absence of external control; it is self-control, which consists in a principle of control being adopted, not arbitrarily, but because it is seen to be right. In Kant's formulation, the idea of the moral law alone – just the idea of it – evokes a feeling of reverence that demolishes self-love. And thus the dialectic of moral life is set up, between inclinations rooted in flesh and moral duty grounded in reason. Freedom depends both on there being that dialectic and on our choosing morality over inclination. But this depends on the moral law not being arbitrary. It is truth that shall make you free.

In all of this, we see parallels to Peirce's thought, which is unsurprising, as his philosophical education began with Schiller and Kant (but origins are less important than destinations). The two forms of motivation, by inclination and by reason, are efficient and final causation, respectively; the adoption of an ideal is through the feeling of admiration that its contemplation creates in us. Peirce, too, opposed autonomy to self-love, regarding the philosophical theory of egoism as 'the metaphysics of wickedness'; he portrayed scientific inquiry as a model of morality, because it requires the researcher to subject the dearest products of his own mind to impersonal tests and thus to the judgment, ultimately, of an indefinitely extended community. Finally, nothing could be more important to Peirce than autonomy, as the very existence of one's self consists in self-control (chapter 11, section 4), just as the existence of a community of inquirers consists in its struggle to find the right methods and the right aim to which to subordinate thought.

However, Peirce denied that there is anything self-authenticating: in words used in 1868, he denied that there are intuitive cognitions. It follows that freedom and personal existence are possible only if there is a form of inquiry that is objective – inquiry that, in absence of foundations or despite our not having truths with which to begin, nevertheless converges, over the long run, on answers to at least some questions. It follows that freedom will always be imperfect and that one's self and one's community with others will always be under construction; for the substitution of inquiry for certainty means that we are always on the way, that there is no time at which we will have arrived. It is in that respect that pragmatism departs from Kantianism.

In addition, precisely because there is nothing self-authenticating – precisely because we have no knowledge a priori of how to inquire – there can never be a time when we will know, for sure, that we are proceeding in the right way or even that there is a right way to proceed. We can only go by the evidence we have so far acquired, in faith that there is an impersonal truth, that is, a final opinion toward which an ideal inquiry would tend. The evidence that supports that faith is extensive and compelling and yet conceivably erroneous. It is shot through with uncertainty, unanswered questions, unresolved problems, and vague formulations. The deeper we push our inquiries, the more exciting but less satisfactory they become. We have seen that theories in basic physics are far less certain and far more obscure than theories – one might better say, knowledge – in other departments of science, despite the fact that these latter take ideas and techniques from physics for granted: a paradox removed, or reduced, only by recognizing the part that inspecific ideas play in all of science. Philosophical clarification of the methods of science likewise is less certain than are the methods themselves. Philosophy has lagged science, in some respects by centuries. The same applies to mathematics and morals, in which much is known despite philosophers' inability to discover what makes mathematical or moral propositions true.

Is this situation deplorable? In a letter of 1905 to William James, Peirce wrote,

> There is, however, nothing more wholesome for us than to find problems that quite transcend our powers and I must say, too, that it imparts a delicious sense of being cradled in the waters of the deep, – a feeling I always have at sea. (8.263)

Bibliography

With one or two exceptions, I have cited classic works, from Plato through Kant, by volume, book, chapter, section, or in other ways that are standard among scholars regardless of the edition or translation read; this I have done even when quoting a translation, though then I have given the name of the translator. These works are not listed in this bibliography, which is neither a complete record of the works I have consulted nor a guide to further reading on the topic. It lists only the editions of Peirce's writings cited in the text (according to that system explained in a note in the Preface) and the works I have cited by author or editor and date (the date is not always that of first publication; when it is relevant to note the date of earlier publication, I have included it in square brackets).

Peirce's Writings Cited in the Text

Collected Papers of Charles Sanders Peirce. Vols. 1–6, Hartshorne and Weiss, eds. Vols. 7–8, Burks, ed. Harvard UP, 1931–5 and 1958.

The Essential Peirce: Selected Philosophical Writings. Vols. 1–2, various editors at the Peirce Edition Project. Indiana UP, 1992 and 1998.

The New Elements of Mathematics by Charles S. Peirce. Vols. 1–4, Eisele, ed. Mouton, 1976.

Reasoning and the Logic of Things: The Cambridge Conferences Lectures of 1898, by Charles Sanders Peirce. Ketner, ed., with an Introduction by Ketner and Putnam and Commentary by Putnam. Harvard UP, 1992.

Semiotic and Significs: The Correspondence between Charles S. Peirce and Victoria Lady Welby. Hardwick, ed. Indiana UP, 1977.

Writings of Charles S. Peirce: A Chronological Edition. Vols. 1–6, various editors at the Peirce Edition Project, Indiana UP, 1982–2000.

Other Works

TCSPS = *Transactions of the Charles S. Peirce Society*

Albert, David Z. 1994. The Foundations of Quantum Mechanics and the Approach to Thermodynamic Equilibrium. *British Journal for the Philosophy of Science*.

2000. *Time and Chance*. Harvard UP.

Aldrich, Virgil. 1958. Picture Space. *Philosophical Review*.

Alston, William. 1956–7. Pragmatism and the Theory of Signs in Peirce. *Philosophy and Phenomenological Research*.

Anscombe, G. E. M. 1965. The Intentionality of Sensation: A Grammatical Feature. In Butler, ed., *Analytical Philosophy: Second Series*. Basil Blackwell.

1975 [1971]. Causality and Determinism. In Sosa, ed., *Causation and Conditionals*. Oxford UP.

Anttila, Raimo. 1989. *Historical and Comparative Linguistics*. John Benjamins.

Apel, Karl Otto. 1980. *Towards a Transformation of Philosophy*. Adey and Frisby, trans. Routledge and Kegan Paul.

1981. *Charles S. Peirce: From Pragmatism to Pragmaticism*. Krois, trans. U Massachusetts P.

1995. Transcendental Semeiotic and Hypothetical Metaphysics of Evolution: A Peircean or Quasi-Peircean Answer to a Recurrent Problem of Post-Kantian Philosophy. In Ketner, ed.

Aquinas, Thomas. 1951. *Philosophical Texts*. Gilby, ed. Oxford UP.

Armstrong, D. M. 1968. *A Materialist Theory of Mind*. Routledge and Kegan Paul.

Audi, Robert, ed. 1999. *The Cambridge Dictionary of Philosophy*, 2nd ed. Cambridge UP.

Augustine. 1873. *Oeuvres completes de Saint Augustin*. Traduites en français et annotées par Péronne, Écalle, Vincent, Charpentier et Barreau. Librarie de Louis Vives (Paris), Tome 6.

1958. *On Christian Doctrine*. J. W. Robertson, Jr., trans. Bobbs Merrill.

Austin, J. L. 1961. *Philosophical Papers*. Oxford UP.

1962a. *How to Do Things with Words*. Harvard UP.

1962b. *Sense and Sensibilia*. Oxford UP.

Ayala, Francisco. 1968. Biology as an Autonomous Science. *American Scientist*.

1970. Teleological Explanations in Evolutionary Biology. *Philosophy of Science*.

Ayer, A. J. 1968. *The Origins of Pragmatism*. Macmillan.

Beckner, Morton. 1959. *The Biological Way of Thought*. Columbia UP.

Bergman, Mats. 2000. Reflections on the Role of the Communicative Sign in Semeiotic. TCSPS.

Bernstein, Richard. 1964. Peirce's Theory of Perception. In Moore and Robin, eds.

1965. Action, Conduct, and Self-control. In *Perspectives on Peirce*. Berstein, ed. Yale UP.

1971. *Praxis and Action*. U of Pennsylvania P.

Boler, John F. 1963. *Charles Peirce and Scholastic Realism: A Study of Peirce's Relation to John Duns Scotus*. U Washington P.

Boorse, Christopher. 1976. Wright on Functions. *Philosophical Review*.

Braithwaite, R. B. 1953. *Scientific Explanation.* Cambridge UP.
Brandon, Robert N. 1996. *Concepts and Methods in Evolutionary Biology.* Cambridge UP.
Brentano, Franz. 1973 [1874]. *Psychology from an Empirical Standpoint.* McAlister, ed., translation of the 1924 edition, Oskar Kraus, ed., of *Psychologie vom empirischen Standpunkt* (1st ed., 1874), including Brentano's supplementary essays of 1911 and Kraus's notes. Humanities Press.
Brock, Jarret. 1977. Draft of a Critique of Greenlee's *Peirce's Concept of Sign*. TCSPS.
1981a. An Introduction to Peirce's Theory of Speech Acts. TCSPS.
1981b. Peirce and Searle on Assertion. In Ketner et al., eds.
1997. The Development of Peirce's Theory of Proper Names. In Houser et al., eds.
Bunge, Mario. 1959. *Causality: The Place of the Causal Principle in Modern Science.* Harvard UP.
Burch, Robert W. 1991. *A Peircean Reduction Thesis.* Texas Tech UP.
Burger, Ronna. 1984. *The Phaedo: A Platonic Labyrinth.* Yale UP.
Burks, Arthur W. 1948–9. Icon, Index, and Symbol. *Philosophy and Phenomenological Research*.
1964. Peirce's Two Theories of Probability. In Moore and Robin, eds.
1977. *Chance, Cause, Reason: An Inquiry into the Nature of Scientific Evidence.* U Chicago P.
1988. Teleology and Logical Mechanism. *Synthese*.
Castañeda, Hector-Neri, ed. 1967. *Intentionality, Minds, and Perception.* Wayne State UP.
Charlton, W. 1970. *Aristotle's Physics, Books I and II.* Oxford UP.
Chisholm, Roderick. 1952. Intentionality and the Theory of Signs. *Philosophical Studies* 3.
1957. *Perceiving: A Philosophical Study.* Cornell UP.
Churchland, Paul M. 1979. *Scientific Realism and the Plasticity of Mind.* Cambridge UP.
1981. Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*.
Colapietro, Vincent. 1989. *Peirce's Approach to the Self: A Semiotic Perspective on Human Subjectivity.* SUNY P.
Copleston, Frederick, S. J. 1962. *A History of Philosophy.* Vol. 2, Part II. Doubleday.
Cummings, Naomi. 2000. *The Sonic Self: Musical Subjectivity and Signification.* Indiana UP.
Dauben, Joseph W. 1981. Peirce on Continuity and His Critique of Cantor and Dedekind. In Ketner et al., eds.
1995. Peirce and the History of Science. In Ketner, ed.
Delaney, C. F. 1993. *Science, Knowledge, and Mind: A Study in the Philosophy of C. S. Peirce.* U Notre Dame P.
Deledalle, Gérard. 1987. *Charles S. Peirce, phénoménologue et sémioticien.* J. Benjamins.
2000. *Charles S. Peirce's Philosophy of Signs: Essays in Comparative Linguistics.* Indiana UP.
Dennett, Daniel. 1978. *Brainstorms.* Bradford Books.
1987. *The Intentional Stance.* MIT P.

Derrida, Jacques. 1976. *Of Grammatology*. Spivak, trans. Johns Hopkins UP.
De Tienne, André. 1996. *L'analytique de la représentation chez Peirce*. Publications des Facultés Universitaires Saint-Louis (Brussels, Belgium).
Dewey, John. 1960. *On Experience, Nature, and Freedom: Representative Selections*. Bernstein, ed. Bobbs Merrill.
Dijksterhuis, E. J. 1986 [1950]. *The Mechanization of the World Picture*. Princeton UP.
DiLeo, Jeffey R. 1997. Charles Peirce's Theory of Proper Names. In Houser et al., eds.
Dray, William. 1957. *Laws and Explanation in History*. Oxford UP.
Dretske, Fred. 1981. *Knowledge and the Flow of Information*. MIT P.
Eco, Umberto. 1976. *A Theory of Semiotics*. Indiana UP.
1995. Unlimited Semeiosis and Drift: Pragmaticism vs. 'Pragmatism'. In Ketner, ed.
Eco, Umberto, et al. 1986. 'Latratus Canis' or: The Dog's Barking. In *Frontiers in Semiotics*, Deely, Williams, Kruse, eds. Indiana UP.
Ehrenfest, Paul, and Ehrenfest, Tatiana. 1958 [1912]. *The Conceptual Foundations of the Statistical Approach in Mechanics*. Moravcsik, trans. Dover.
Eisele Carolyn. 1985. *Historical Perspectives on Peirce's Logic of Science: A History of Science*. 2 vols. Mouton.
Emmeche, Claus. 1991. A Semiotical Reflection on Biology, Living Signs, and Artificial Life. *Biology and Philosophy*.
1998. Defining Life as a Semiotic Phenomenon. *Cybernetics and Human Knowing*
Emmeche, Claus, and Jesper Hoffmeyer. 1991. From Language to Nature: The Semiotic Metaphor in Biology. *Semiotica*.
Fetzer, James H. 1990. *Artificial Intelligence: Its Scope and Limits*. Kluwer Academic.
Fetzer, James H., and Donald E. Nute. 1979. Syntax. Semantics, and Ontology: A Probabilistic Causal Calculus. *Synthese*.
Feyerabend, Paul. 1962. Explanation, Reduction and Empiricism. In *Scientific Explanation, Space and Time*. Feigl and Maxwell, eds. U of Minnesota P.
1965. Problems of Empiricism. In *Beyond the Edge of Certainty*. Colodny, ed. Prentice Hall.
Fisch, Max H. 1986. *Peirce, Semeiotic, and Pragmatism*. Ketner and Kloesel, eds. Indiana UP.
Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Oxford UP.
Fitzgerald, John J. 1966. *Peirce's Theory of Signs as Foundation for Pragmatism*. Mouton.
Fodor, Jerry. 1975. *The Language of Thought*. Thomas Y. Crowell.
1990. *A Theory of Content*. MIT P.
1994. *The Elm and the Expert*. MIT P.
Frede, Michael. 1980. The Original Notion of Cause. In *Doubt and Dogmatism: Studies in Hellenistic Epistemology*. Schofield, Burnyeat, Barnes, eds. Oxford UP.
Freeman, Eugene, ed. 1983. *The Relevance of Charles Peirce*. Monist Library of Philosophy, Hegeler Institute, La Salle, Illinois.
Gallie, W. B. 1966 [1952]. *Peirce and Pragmatism*. Dover.
Gentry, George V. 1952. Habit and the Logical Interpretant. In Weiner and Young, eds.

Ghiselin, Michael T. 1969. *The Triumph of the Darwinian Method*. U California P.
1974. A Radical Solution to the Species Problem. *Systematic Zoology*.
Girel, Mathias. 2003. The Metaphysics and Logic of Psychology: Peirce's Reading of James' *Principles*. TCSPS.
Gombrich, E. H. 1960. *Art and Illusion*. Phaidon Press.
Goodman, Nelson. 1965 [1955]. *Fact, Fiction, and Forecast*, 2nd ed. Bobbs-Merrill.
1968. *Languages of Art*. Bobbs-Merrill.
1978. *Ways of Worldmaking*. Hackett.
Greenlee, Douglas. 1973. *Peirce's Concept of Sign*. Mouton.
Grice, Paul. 1989. *Studies in the Way of Worlds*. Harvard UP.
Grünbaum, Adolph. 1963. *Philosophical Problems of Space and Time*. Knopf.
Haack, Susan. 1992. 'Extreme Scholastic Realism': Its Relevance to Philosophy of Science Today. TCSPS.
1993. *Evidence and Inquiry*. Blackwell.
1998. *Manifesto of a Passionate Moderate: Unfashionable Essays*. Chicago UP.
2003. *Defending Science – Within Reason: Between Scientism and Cynicism*. Prometheus Books.
Habermas, Jürgen. 1995. Peirce and Communication. In Ketner, ed.
Hacking, Ian. 1983. *Representing and Intervening*. Cambridge UP.
1990. *The Taming of Chance*. Cambridge UP.
Hanson, Norwood Russell. 1958. *Patterns of Discovery*. Cambridge UP.
Hare, R. M. 1952. *The Language of Morals*. Oxford UP.
Harré, R., and E. H. Madden. 1975. *Causal Powers*. Rowman and Littlefield.
Hartshorne, Charles. 1952. The Relativity of Nonrelativity: Some Reflections on Firstness. In Weiner and Young, eds.
Hausman, Carl R. 1993. *Charles S. Peirce's Evolutionary Philosophy*. Cambridge UP.
Hegel, G. W. F. 1969 [1812/31]. *Science of Logic*. Miller, trans. Allen and Unwin.
Hempel, Carl G. 1962. Deductive-Nomological vs. Statistical Explanation. In *Minnesota Studies in the Philosophy of Science*. Feigl and Maxwell, eds. Vol. 3. U Minnesota P.
1965. *Aspects of Scientific Explanation*. Free Press.
Hempel, Carl G., and Paul Oppenheim. 1948. Studies in the Logic of Explanation. *Philosophy of Science*.
Herzberger, Hans. 1981. Peirce's Remarkable Theorem. In *Pragmatism and Purpose: Essays Presented to Thomas A. Goudge*. Sumner et al., eds. U Toronto P.
Hilpinen, Risto. 1983. On C. S. Peirce's Theory of Propositions as a Precursor of Game-Theoretical Semantics. In Freeman, ed.
1992. On Peirce's Philosophical Logic: *Propositions and Their Objects*. TCSPS.
1995. Peirce on Language and Reference. In Ketner, ed.
Hintikka, Jaakko. 1973. *Logic, Language-Games, and Information: Kantian Themes in the Philosophy of Logic*. Oxford UP.
1979. Quantifiers in Logic and Quantifiers in Natural Languages; Quantifiers vs. Quantification Theory; and Quantifiers in Natural Languages: Some Logical Problems. All three in *Game-Theoretical Semantics*. Saarinen, ed. Reidel.
1983. C. S. Peirce's "First Real Discovery" and Its Contemprary Relevance. In Freeman, ed.
Hoffmeyer, Jesper. 1996. *Signs of Meaning in the Universe*. Haveland, trans. Indiana UP.

Hollinger, Henry B., and Michael J. Zenzen. 1985. *The Nature of Irreversibility*. Reidel.

Hookway, Christopher. 1985. *Peirce*. Routledge and Kegan Paul.

2000. *Truth, Rationality, and Pragmatism: Themes from Peirce*. Oxford UP.

Horwich, Paul. 1990. *Truth*. Oxford UP.

Houser, Nathan. 1992. On Peirce's Theory of Propositions: A Response to Hilpinen. TCSPS.

Houser, Nathan, Don D. Roberts, and James Van Evra, eds. 1997. *Studies in the Logic of Charles Sanders Peirce*. Indiana UP.

Husserl, Edmund. 1970 [1900–1]. *Logical Investigations*. Vols. I and II. Findlay, trans. Routledge and Kegan Paul.

1985. *Selected Writings*, Vol. 7. Harvard UP.

Jakobson, Roman. 1990 [1966]. Quest for the Essence of Language. Reprinted as ch. 25 of Jakobson, *On Language*, Waugh and Monville-Burston, eds. Harvard UP.

Jeffreys, Richard C. 1970. Statistical Explanation vs. Statistical Inference. In Nicholas Rescher et al., eds., *Essays in Honor of C. G. Hempel*. D. Reidel.

Jeffreys, Harold. 1973 [1931]. *Scientific Inference*, 3rd ed. Cambridge UP.

Johansen, Jrgen Dines. 1993. *Dialogic Semiosis: An Essay on Signs and Meaning*. Indiana UP.

Kent, Beverly. 1987. *Charles S. Peirce: Logic and the Classification of Science*. McGill-Queens UP.

Kerr-Lawson, Angus. 1992. Stripped Down Burch. TCSPS.

Ketner, Kenneth, et al., eds. 1981. *Proceedings of the C. S. Peirce Bicentennial International Congress*. Texas Tech Press.

Ketner, Kenneth Laine, ed. 1995. *Peirce and Contemporary Thought*. Fordham UP.

Kim, Jaegwon. 1993. *Supervenience and Mind*. Cambridge UP.

Krausser, Peter. 1977. The Three Fundamental Structural Categories of Charles S. Peirce. TCSPS.

Kripke, Saul. 1980 [1972]. *Naming and Necessity*. Harvard UP.

Kruse, Felicia E. 1991. Genuineness and Degeneracy in Peirce's Categories. TCSPS.

2005. Emotion in Musical Meaning: A Peircean Solution to Langer's Problem. TCSPS.

Kuhn, Thomas S. 1970 [1962]. *The Structure of Scientific Revolutions*, 2nd ed. U Chicago P.

1977. *The Essential Tension*. U Chicago P.

1991. The Trouble with the Historical Philosophy of Science. Lecture delivered at Harvard, printed in Kuhn 2000.

2000. *The Road since Structure*. U Chicago P.

Lakatos, Imre. 1970. Falsification and the Methodology of Scientific Research Programmes. In Lakatos and Musgrave, eds., *Criticism and the Growth of Knowledge*. Cambridge UP.

1978. *The Methodology of Scientific Research Programmes*. Worrall and Currie, eds. Cambridge UP.

Lane, Robert. 1997. Peirce's 'Entanglement' with the Principles of Excluded Middle and Contradiction. TCSPS.

Latour, Bruno. 2004. *Politics of Nature: How to Bring the Sciences into Democracy.* Harvard UP.
Laudan, Larry. 1984. A Confutation of Convergent Realism. In Leplin, ed.
Lennox, James G. 1992. Teleology. In *Keywords in Evolutionary Biology*, Keller and Lloyd, eds. Harvard UP.
1993. Darwin *Was* a Teleologist. *Biology and Philosophy.*
1994. Teleology by Another Name: A Reply to Ghiselin. *Biology and Philosophy.*
Leplin, Jarrett, ed. 1984. *Scientific Realism.* U California P.
Lewis, David. 1973a. Causation. *Journal of Philosophy.*
1973b. *Counterfactuals.* Oxford UP.
Lieb, I. C. 1977 [1953]. Appendix B. In LW.
Liszka, James Jakób. 1990. Peirce's Intrepretant. TCSPS.
1996. *A General Introduction to the Semeiotic of Charles Sanders Peirce.* Indiana UP.
Lycan, W. Greogory. 1972 [1969]. On "Intentionality" and the Psychological. In Marras 1972.
Markus, R. A. 1957. St. Augustine on Signs. *Phronesis.*
Marras, Ausonio, ed. 1972. *Intentionality, Mind, and Language.* U Illinois P.
Mayr, Ernst. 1974. Teleological and Teleonomic: A New Analysis. In *Boston Studies in the Philosophy of Science.* Cohen and Wartofsky, eds. Vol. 14. Dordrecht.
Mead, George Herbert. 1934. *Mind, Self and Society.* U Chicago P.
Merleau-Ponty, Maurice. 1962 [1945]. *The Phenomenology of Perception.* Smith, trans. Routledge and Kegan Paul.
Meyers, Robert G. 1992. Peirce's New Way of Signs. TCSPS.
Michael, Emily. 1976. Peirce on Individuals. TCSPS.
Migotti, Mark. 1999. Peirce's Double-Aspect Theory of Truth. *Canadian Journal of Philosophy* (Supp. Vol. 24, Misak, ed.).
Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories.* MIT P.
1993. *White Queen Psychology and Other Essays for Alice.* MIT P.
Misak, Cheryl J. 1991. *Truth and the End of Inquiry: A Peircean Account of Truth.* Oxford UP.
1995. *Verificationism: Its History and Prospects.* Routledge.
2000. *Truth, Politics, Morality: Pragmatism and Deliberation.* Routledge.
Misak, Cheryl J., ed. 2004. *The Cambridge Companion to Peirce.* Cambridge UP.
Monod, Jacques. 1970. *Le hasard et la nécessité.* Éditions du Seuil.
Moore, Edward C., and Richard S. Robin, eds. 1964. *Studies in the Philosophy of Charles Sanders Peirce.* Second Series. U Mass. P.
Morris, Charles W. 1938. *Foundations of the Theory of Signs.* International Encyclopedia of Unified Science. Vol. I, no. 2. Chicago UP.
1946. *Signs, Language, and Behavior.* Prentice-Hall.
1964. *Signification and Significance: A Study of the Relations of Signs and Values.* MIT P.
Müller, Ralf. 1994. On the Principles of Construction and the Order of Peirce's Trichotomies of Signs. TCSPS.
Murphey, Murray G. 1961. *The Development of Peirce's Philosophy.* Harvard UP.
Nagel, Ernest. 1961. *The Structure of Science.* Harcourt, Brace.

Niiniluoto, Ilkka. 1993. Peirce's Theory of Statistical Explanation. In *Charles S. Peirce and the Philosophy of Science*. Moore, ed. U Alabama P.
Oehler, Klaus. 1974. Review of Greenlee's *Peirce's Concept of Sign*. TCSPS.
1995. A Response to Habermas. In Ketner, ed.
Owen, G. E. L. 1965. Aristotle on the Snares of Ontology. In *New Essays on Plato and Aristotle*, Bambrough, ed. Routledge and Kegan Paul.
Pape, Helmut. 1989. *Erfahrung und Wirklichkeit als Zeichenprozess: Charles S. Peirces Entwurf einer Spekulativen Grammatik des Seins*. Suhrkamp Verlag.
1990. Charles S. Peirce on Objects of Thought and Representation. *Noûs*.
Papineau, David. 1984. Representation and Explanation. *Philosophy of Science*.
1987. *Reality and Representation*. Basil Blackwell.
Parmentier, Richard. 1985. Signs' Place in Medias Res: Peirce's Concept of Semiotic Mediation. In *Semiotic Mediation: Sociocultural and Psychological Perspectives*. Mertz et al., eds. Academic P.
1994. *Signs in Society: Studies in Semiotic Anthropology*. Indiana UP.
Place, U. T. 1956. Is Consciousness a Brain Process? *British Journal of Psychology*.
Plantinga, Alvin. 1974. *The Nature of Necessity*. Oxford UP.
Popper, Karl R. 1959. The Propensity Interpretation of Probability. *British Journal for the Philosophy of Science*.
1965 [1934]. *The Logic of Scientific Discovery*, revised ed. Harper.
1969 [1963]. *Conjectures and Refutations: The Growth of Scientific Knowledge*, 3rd ed. Routledge and Kegan Paul.
1972. *Objective Knowledge: An Evolutionary Approach*. Oxford UP.
1983. *Realism and the Aim of Science*. Rowman and Littlefield.
Potter, Vincent G. 1967. *Charles S. Peirce on Norms and Ideals*. U Mass. P.
Potter, Vincent G., and Paul B. Shields. 1977. Peirce's Definitions of Continuity. TCSPS.
Prigogine, Ilya. 1980. *From Being to Becoming*. W. H. Freeman.
Putnam, Hilary. 1975. *Mind, Language, and Reality*. Philosophical Papers, vol. 2. Cambridge UP.
1978. *Meaning and the Moral Sciences*. Routledge and Kegan Paul.
1983. *Realism and Reason*. Philosophical Papers, vol. 3. Cambridge UP.
1988. *Representation and Reality*. MIT P.
1992. *Renewing Philosophy*. Harvard UP.
1995. Peirce's Continuum. In Ketner, ed.
Quine, Willard Van Orman. 1951. *Mathematical Logic*, rev. ed. Harvard UP.
1960. *Word and Object*. MIT Press.
1961. *From a Logical Point of View*, rev. ed. Harvard UP.
1969. *Ontological Relativity and Other Essays*. Columbia UP.
1990. *Pursuit of Truth*. Harvard UP.
Railton, Peter. 1977. A Deductive-Nomological Model of Probabilistic Explanation. *Philosophy of Science*.
1981. Probability, Explanation, and Information. *Synthese*.
1989. Explanation and Metaphysical Controversy. In *Scientific Explanation*, Kitcher and Salmon, eds. U Minnesota P.
Ransdell, Joseph. 1976. Another Interpretation of Peirce's Semiotic. TCSPS.
1977. Some Leading Ideas of Peirce's Semiotic. *Semiotica*.

1979. The Epistemic Function of Iconicity in Perception. In *Peirce Studies*, no. 1, Institute for Studies in Pragmaticism. Texas Tech University.
1981. Semiotic Causation: A Partial Explication. In Ketner et al., eds.
Reichenbach, Hans. 1951. *The Rise of Scientific Philosophy*. U California P.
Rescher, Nicholas. 1978. *Peirce's Philosophy of Science*. U Notre Dame P.
2001. *Cognitive Pragmatism*. Pittsburgh UP.
Reynolds, Andrew. 2002. *Peirce's Scientific Metaphysics: The Philosophy of Chance, Law, and Evolution*. Vanderbilt UP.
Riley, Gresham. 1975. Peirce's Theory of Individuals. TCSPS.
Roberts, Don D. 1973. *The Existential Graphs of Charles Peirce*. Mouton.
Robin, Richard S. 1967. *Annotated Catalogue of the Papers of Charles S. Peirce*. U Massachusetts P.
Rorty, Richard. 1982. *Consequences of Pragmatism*. U Minnesota P.
Rosenblueth, Arturo, Norbert Wiener, and Julian Bigelow. 1943. Behavior, Purpose, and Teleology. *Philosophy of Science*.
Royce, Josiah. 1968 [1913]. *The Problem of Christianity*. U Chicago P.
Ruse, Michael. 1988. *Philosophy of Biology Today*. SUNY P.
Russell, Bertrand. 1905. On Denoting. *Mind*.
Ryle, Gilbert. 1954. *Dilemmas*. Cambridge UP.
Salmon, Wesley. 1970. Statistical Explanation. In *The Nature and Function of Scientific Theories*. Colodny, ed. U Pittsburgh P.
1984. *Scientific Explanation and the Causal Structure of the World*. Princeton UP.
1990. Rationality and Objectivity in Science, or Tom Kuhn Meets Tom Bayes. In *Scientific Theories*. C. Wade Savage, ed. U Minnesota P.
1998. *Causality and Explanation*. Oxford UP.
Sambursky, Shmuel, ed. 1974. *Physical Thought from the Presocratics to the Quantum Physicists*. Hutchinson.
Sanders, Gary. 1970. Peirce's Sixty-six Signs? TCSPS.
Santaella Braga, Lucia. 1999a. A New Causality for Understanding of the Living. *Semiotica*.
1999b. Peirce and Biology. *Semiotica*.
Saussure, Ferdinand de. 1983. *Course in General Linguistics*. Harris, trans., of the 2nd ed., 1922, of *Cours de linguistique générale* [1st ed., 1916]. Open Court
Savan, David. 1952. On the Origins of Peirce's Phenomenology. In Weiner and Young, eds.
1977. Questions Concerning Certain Classifications Claimed for Signs. *Semiotica*.
1981. Peirce's Semiotic Theory of Emotion. In Ketner et al., eds.
1983. Toward a Refutation of Semiotic Idealism. *Recherches Sémiotique/Semiotic Inquiry*.
1986. Response to T. L. Short. TCSPS.
1987. *An Introduction to C. S. Peirce's Full System of Semeiotic* (rev. ed. of a monograph first published in 1976). Victoria College in the University of Toronto.
Scheffler, Israel. 1959. Thoughts on Teleology. *British Journal for the Philosophy of Science*.
1982 [1967]. *Science and Subjectivity*, 2nd ed. Hackett.

Schrödinger, Erwin. 1967 [1944]. *What Is Life? & Mind and Matter*. Cambridge UP.

Schwartz, Stephen P., ed. 1977. *Naming, Necessity, and Natural Kinds*. Cornell UP.

Schweber, Silvan S. 1982. Demons, Angels, and Probability: Some Aspects of British Science in the 19th Century. In *Physics as Natural Philosophy*, Shimony and Feshbach, eds. MIT P.

Scriven, Michael. 1962. Explanation, Prediction, and Laws. In *Minnesota Studies in the Philosophy of Science*. Feigl and Maxwell, eds. vol. 3. U Minnesota P.

Searle, John R. 1969. *Speech Acts*. Cambridge UP.

1971. What Is a Speech Act? In *The Philosophy of Language*, Searle, ed. Oxford UP.

1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge UP.

1992. *The Rediscovery of the Mind*. MIT P.

Sebeok, Thomas A. 1985. *Contributions to the Doctrine of Signs*. UP of America.

1986. The Doctrine of Signs. In *Frontiers in Semiotics*, Deely, Williams, Kruse, eds. Indiana UP.

Sellars, Wilfrid. 1963. *Science, Perception and Reality*. Routledge and Kegan Paul.

Shapere, Dudley. 1966. Meaning and Scientific Change. In *Mind and Cosmos: Essays in Contemporary Science and Philosophy*. Colodny, ed. U Pittsburgh P.

Shapiro, Michael. 1983. *The Sense of Grammar*. Indiana UP.

1991. *The Sense of Change: Language as History*. Indiana UP.

Short, T. L. 1980a. An Analysis of Conceptual Change. *American Philosophical Quarterly*.

1980b. Peirce and the Incommensurability of Theories. *Monist*.

1981a. Peirce's Concept of Final Causation. TCSPS.

1981b. Semeiosis and Intentionality. TCSPS.

1982. Life among the Legisigns. TCSPS.

1983. Teleology in Nature. *American Philosophical Quarterly*.

1984. Some Problems Concerning Peirce's Conceptions of Concepts and Propositions. TCSPS.

1986a. David Savan's Peirce Studies. TCSPS.

1986b. What They Said in Amsterdam: Peirce's Semiotics Today. *Semiotica*.

1988a. The Growth of Symbols. *Cruzeiro semiotico* 8 (Porto, Portugal).

1988b. Hypostatic Abstraction in Empirical Science. *Grazer Philosophische Studien* (Graz, Austria).

1988c. Why We Prefer Peirce to Saussure. In *Semiotics 1988*. Prewitt et al., eds. UP of America.

1992. Peirce's Semiotic Theory of the Self. *Semiotica*.

1994a. David Savan's Defense of Semiotic Realism. *Semiotica*.

1994b. On Hermeticism in Semiotics. In *The Peirce Seminar Papers*. Shapiro and Haley, eds. Vol. 2, Berghahn.

1996a. Interpreting Peirce's Interpretant: A Response to Lalor, Liszka, and Meyers. TCSPS.

1996b. Review of *The Writings of Charles S. Peirce*. Vols. 1–5. *Synthese*.

1997. Hypostatic Abstraction in Self-Consciousness. In *The Rule of Reason: The Philosophy of Charles Sanders Peirce*. Brunning and Forster, eds. U Toronto P.

1998a. The Discovery of Scientific Aims and Methods. *American Catholic Philosophical Quarterly*.
1998b. Jakobson's Problematic Appropriation of Peirce. In *The Peirce Seminar Papers*. Shapiro, ed. Vol. 3. Peter Lang.
1999. Teleology and Linguistic Change. In *The Peirce Seminar Papers*. Shapiro and Haley, eds. Vol. 4. Berghahn.
1999/2000. Peirce on Meaning and Translation. *Athanor* 10/2 (Bari, Italy).
2000a. Peirce on the Aim of Inquiry: Another Reading of 'Fixation'. TCSPS.
2000b. Was Peirce a Weak Foundationalist? TCSPS.
2002a. Darwin's Concept of Final Cause: Neither New nor Trivial. *Biology and Philosophy*.
2002b. Robin on Perception and Sentiment in Peirce. TCSPS.
2004. The Development of Peirce's Theory of Signs. In Misak, ed.
Skagestad, Peter. 1981. *The Road of Inquiry: Charles Peirce's Pragmatic Realism*. Columbia UP.
Sklar, Lawrence. 1993. *Physics and Chance*. Cambridge UP.
Smyth, Richard A. 1997. *Reading Peirce Reading*. Rowman and Littlefield.
Soames, Scott. 1999. *Understanding Truth*. Oxford UP.
2003. *Philosophical Analysis in the Twentieth Century*. 2 vols. Princeton UP.
Sober, Elliott. 1984. *The Nature of Selection*. Chicago UP.
1985. Panglossian Functionalism and the Philosophy of Mind. *Synthese*.
Spiegelberg, Herbert. 1957. Husserl's and Peirce's Phenomenologies: Coincidence or Interaction. *Philosophy and Phenomenological Research*.
1965. *The Phenomenological Movement*. Vol. 1. Martinus Nijhoff.
Stalnaker, Robert. 1968. A Theory of Conditionals. In *Studies in Logical Theory*. Rescher, ed. American Philosophical Quarterly supplementary monograph.
Stern, Paul. 1993. *Socratic Rationalism and Political Philosophy*. SUNY P.
Stitch, Stephen P. 1983. *From Folk Psychology to Cognitive Science: The Case against Belief*. Cambridge UP.
Stout, Rowland. 1996. *Things That Happen Because They Should*. Oxford UP.
Strawson, P. F. 1950. On Referring. *Mind*.
Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. North-Holland.
Taylor, Charles. 1964. *The Explanation of Behavior*. Routledge and Kegan Paul.
Taylor, Richard. 1950. Comments on the Mechanistic Conception of Purposefulness. *Philosophy of Science*.
Thompson, Manley. 1953. *The Pragmatic Philosophy of C. S. Peirce*. U Chicago P.
Toulmin, Stephen. 1961. *Foresight and Understanding*. Hutchinson.
1972. *Human Understanding*. Princeton UP.
Turley, Peter T. 1975. Peirce's Cosmic 'Sheriff'. *Journal of the History of Ideas*.
van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford UP.
Vlastos, Gregory. 1969. Reasons and Causes in the *Phaedo*. *Philosophical Review*.
von Mises, Richard. 1957 [1928]. *Probability, Statistics, and Truth*. Macmillan.
von Plato, Jan. 1994. *Creating Modern Probability*. Cambridge UP.
von Wright, Georg. 1971. *Explanation and Understanding*. Cornell UP.
Wallace, William. 1972–4. *Causality and Scientific Explanation*. 2 vols. U Mich. P.
Weiner, Philip P., and Frederic H. Young, eds. 1952. *Studies in the Philosophy of Charles Sanders Peirce*. Harvard UP.

Weiss, Paul, and Arthur Burks. 1945. Peirce's Sixty-six Signs. *Journal of Philosophy*.
Wicksteed, P. H., and Cornford, F. M. 1957. *The Physics of Aristotle*, revised ed. Heinemann.
Wiggins, David. 1980. *Substance and Sameness*. Basil Blackwell.
1986. Teleology and the Good in Plato's *Phaedo*. *Oxford Studies in Ancient Philosophy*, vol 4. Oxford UP.
Wimsatt, William C. 1972. Teleology and the Logical Structure of Function Statements. *Studies in the History and Philosophy of Science*.
Winch, Peter. 1958. *The Idea of a Social Science*. Routledge and Kegan Paul.
Wittgenstein, Ludwig. 1967 [1953]. *Philosophical Investigations*. 3rd ed. Anscombe, trans. Basil Blackwell.
Woodfield, Andrew. 1976. *Teleology*. Cambridge UP.
Woodward, James. 1989. The Causal Mechanical Model of Explanation. In *Scientific Explanation*. Kitcher and Salmon, eds. U Minnesota P.
Wright, Crispin. 1992. *Truth and Objectivity*. Harvard UP.
Wright, Larry. 1973. Functions. *Philosophical Review*.
1976. *Teleological Explanation*. U California P.

Name Index

Albert, D., 121
Aldrich, V., 216
Alston, W., 43
Anaxagoras, 99, 101
Anscombe, G. E. M., 15, 83
Anselm of Canterbury, 21
Anttila, R., 287
Apel, K.-O., xiv, 61n3, 64n4, 65, 66, 66n5, 69n7, 196n16, 214
Aquinas, Thomas of, 93, 105
Aristotle,2, 5, 22, 32, 93, 98, 100, 104, 105, 106, 124, 136, 149, 164, 167n8, 201, 204, 321, 326, 335
Armstrong, D., 291
Arnold, M., 307
Augustine, 21, 23, 34, 161
Austin, J. L., 214n2, 242, 260
Ayala, F., 141
Ayer, A. J., 325n3, 331

Bacon, F., 318, 329
Baker, V., xivn3
Beckner, S., 140
Bense, M., 237n2
Bergman, M., xv, 187
Berkeley, G., 197, 306
Bernard, C., 106n15
Bernstein, R., 51, 312, 314, 320
Bohr, N., 337n11
Boler, J., 79n10, 139n
Boltzmann, L., 121
Boorse, C., 142
Bosanquet, B., 323
Boyd, R., 197
Bradley, F. H., 323
Braithwaite, R. B., 140, 141n14, 299
Brandon, R., xii, 142, 143, 299
Brentano, F., 6, 11
Brock, J., xv, 243, 245n6
Bunge, M., 106n15
Burch, R., 74
Burger, R., 107n16
Burks, A., 115n20, 136n10, 237, 241

Cantor, G., 35n7
Carnap, R., 280
Carnot, S., 128
Cassirer, E., 262
Charlton, W., 106n14
Chisholm, R., 10, 11
Churchland, Paul, 291
Cicero, 326
Colapietro, V., 46n10, 313, 314
Comte, A., 62
Copleston, F., 105n13
Cornford, F. M., 106n14
Cratylus, 22
Cummings, R., 205n20

Darwin, C., 117
Dauben, J., 36n7, 331n7
De Morgan, A., 2
Delaney, C. F., 325n3, 337
Deledalle, G., xiv, 18
Democritus, 101
Dennett, D., xi, 125n5, 292, 295, 300, 301

Derrida, J., xvi, 45, 308
Descartes, R., 21, 33, 71, 139, 318
Dijksterhuis, E. J., 98n4
Dilthey, W., 139
Diogenes Laërtius, 23
Donellan, K., 264
Dray, W., 139
Dretske, F., xi, 297, 300
Duns Scotus, 2, 50, 78

Eco, U., xvi, 24, 45, 228
Ehrenfest, P., 121n3
Ehrenfest, T., 121n3
Eisele, C., 331n7
Emmeche, C., xvi, 177
Evans, G., 277n9

Fetzer, J., 115n20, 306
Feyerabend, P., xii, 2, 275, 284, 335
Fisch, M., xv, 28n2, 50n16, 199
Fisher, R. A., 132
Fitzgerald, J., xv
Fodor, J., xi, 280, 281, 297, 298, 300, 301
Fraassen, B. van, 342
Frede, M., 106n14
Frege, G., 48, 246, 277, 280, 281

Gadamer, H.-G., 139
Galileo Galilei, 326, 335
Gallie, W. B., 138n12
Gentry, G., 43
Ghiselin, M., 129n6
Gibbs, W., 121
Girel, M., 81n12
Gombrich, E. H., 228
Goodman, N., 197, 200, 214, 261, 307, 345
Greenlee, D., xv
Grice, H. P., 213, 305
Grünbaum, A., 104n12

Haack, S., 86n15, 199n18, 279n10, 337n12
Habermas, J., xv, 214
Hacking, I., 83n13, 284, 285
Hare, R. M., 248n9
Harré, R., 83
Hartshorne, C., 79n10, 237, 240
Hegel, G. W. F., 60, 139, 323
Hempel, C., 112, 113, 125n4
Heraclitus, 22
Herzberger, H., 74
Hilpinen, R., 223n6, 245n6, 274n5, 281n11
Hintikka, J., 223n6, 265, 266n1
Hippocrates, 21
Hobbes, T., 104n11, 140
Hoffmeyer, J., xvi, 177
Hollinger, H., 119
Hookway, C., 245n6, 246n7, 268n3, 325n3, 331
Horwich, P., 332n10
Houser, N., 50n16, 245n6
Housman, C., 313
Hull, D., 95
Hume, D., 66, 77, 83, 108, 326
Husserl, E., 2, 8, 10, 60, 70, 71

Jakobson, R., 226, 227, 287
James, W., 81n12, 83
Jeffrey, R., 113
Jeffreys, H., 112n17
Jesus of Nazareth, 20
Johansen, J. D., xv, 229

Kant, I., 2, 4, 27, 31, 60, 61, 64, 81, 107, 180, 196, 214, 292, 306, 312, 326, 346
Kent, B., 62
Kerr-Lawson, A., 74
Kierkegaard, S., 308
Kim, J., 293
Kraus, O., 11
Krausser, P., 28n2, 61n2
Kripke, S., xii, 88, 225, 264, 276, 278
Kruse, F., 90n17, 205n20, 231
Kuhn, T., xii, 2, 200n19, 321, 322, 323, 331n8, 335, 341

Lakatos, I., 336
Lambert, J. H., 60
Lane, R., 274n5
Langer, S., 262
Latour, B., 323
Leibniz, G. W., 326
Lennox, J., 103n10
Lewis, D., 88
Lieb, I. C., 237, 242
Lincoln, A., 94
Liszka, J. J., xiv, 168n9, 181n1
Locke, J., 2, 21, 66, 77

Lucretius, 140
Lycan, W., 15n4

Madden, E., 83
Malthus, T., 124
Markus, R. A., 23, 24
Maxwell, J., 118, 121, 125
Mayr, E., 141
Meinong, A., 175
Mendelssohn, F., 204
Meyers, R., 168n9
Migotti, M., 325n3, 332
Mill, J. S., 125n4
Millikan, R. G., xii, 262, 297, 300, 303, 309
Misak, C., 263, 325n3, 332n10
Mises, R. von, 112n17
Mitchell, O. H., 48
Monod, J., 141
Morris, C., xv, 19n6, 262, 306
Müller, R., 237
Murphey, M., 36n7, 48n13, 50n15, 50n16, 65, 66, 330

Nagel, E., 125
Neurath, O., 324
Newton, I., 99n5
Niiniluoto, I., 118n1
Nute, D., 115n20

Ockham, William of, 2
Oehler, K., xv
Owen, G. E. L., 106n14

Pape, H., xvi, 177
Papineau, D., 297, 299, 300, 309
Parmenides, 175
Parmentier, R., xv
Peirce, B., 1
Place, U. T., 291
Plantinga, A., 88
Plato, 21, 34, 93, 94, 99, 101, 105, 106, 175, 224, 323
Plato, J. von, 121n3
Popper, K., xii, 2, 115n20, 128, 324, 336
Potter, V., 36n7, 136n10
Prigogine, I., 134, 294
Putnam, H., xi, xii, 36n7, 118n1, 198n17, 199, 264, 278, 279, 285, 292, 294n3, 304, 306, 341

Quine, W. V. O., 13, 175, 263, 266, 293, 320, 320n2, 325n3, 328

Railton, P., 112, 114, 125
Ransdell, J., xv, 43n9, 50n15, 50n16, 168n9, 186n5
Reichenbach, H., xii, 2, 104n12
Rescher, N., 331n8, 337n12, 343n15
Reynolds, A., 118n1, 138n12
Roberts, D., 266, 266n2, 276n7
Robin, R., 345
Rorty, R., 86n15, 307
Ruse, M., 105
Russell, B., 2, 175

Salmon, W., 113
Sanders, G., 237
Santaella, L., xvi, 177
Saussure, F. de, 16, 27
Savan, D., xiv, 206, 237, 238n3, 345
Scheffler, I., 141, 141n14, 299, 323
Schelling, F. W. J. von, xii
Schröder, E., 2
Schrödinger, E., 118, 122
Schweber, S., 132n8
Searle, J., xi, 10n2, 243, 293, 294
Sebeok, T., xvi, 21, 177
Sellars, W., 292
Sextus Empiricus, 23
Shapiro, M., 287
Shields, P., 36n7
Skagestad, P., 325n3
Sklar, L., 121n3
Smyth, R., 331n9
Soames, S., xiiin2, 332n10
Sober, E., 125n5, 128, 297
Socrates, 99, 101
Spiegelberg, H., 61n2, 61n3
Stalnaker, R., 88
Stern, P., 103n9, 107n16
Stitch, S., 292
Strawson, P. F., 175
Suppes, P., 115n20

Taylor, C., 139
Taylor, R., 140
Theophrastus, 106
Toulmin, S., 2, 321
Turley, P., 139n

Vlastos, G., 103n9, 106, 106n14

Wallace, W., 106n15
Walther, E., 237n2
Weiss, P., 237, 240, 241
Whewell, W., 327
Whitehead, A. N., 2
Wicksteed, P. H., 106n14
Wiener, N., 140
Wiggins, D., 104n11, 106n14, 278
Wimsatt, W., xii, 141, 299
Winch, P., 139
Wittgenstein, L., 11, 86n14, 139, 263, 292, 308
Wolff, C., 4, 98
Woodfield, A., 104n11, 140
Woodward, J., 114
Wright, C., 333
Wright, G. von, 139
Wright, L., xii, 141, 299

Zeno of Elea, 36
Zenzen, M., 119

Subject Index

abduction, 319
abstraction, *see* hypostatic abstraction; prescinding
actuality (existence), distinguished from reality, 86–87; *see also* 2ndness
adicity, 71
antifoundationalism: 318; its sources, 318–322, 323–324; raises question of subjectivism, 323
architectonic, Peirce's: compared with Kant's, 61–62; principles of, 62; ordering of the sciences, 62–63; does not prescribe a temporal order, 63; essentially teleological, 64; inspired by problem of phenomenological description, 64
argument: Peirce's inconsistent use of the term, 248; as distinct from its expression, not a sign, 248; as expression, named 'delome', 248; *see also* rheme/dicisign/argument; seme/pheme/delome
art: judgment of vs. emotional interpretation of, 205; Goodman's comparison of to science and the comparison implicit in Peirce's semeiotic, 261; works of art are semes but may be composed of phemes, 253–254; *see also* music
artificial intelligence, Fetzer's Peircean critique of, 306
assertion: a subtype of replication of dicent symbols, 247–248; Peirce's analysis compared with Austin's, 243; contrary to Peirce's sometimes denial, is significant, 246–247; *see also* propositions; statements

categories, Peirce's system of: as formal acquire phaneroscopic meaning, 86; as phaneroscopic, 74; no proof a priori of their completeness, 74; as metaphysical (also modal or ontological), 74–75, 86–87; sometimes treated ordinally, 74n8, 238n3; *see also* 1stness; phaneroscopy; 2ndness; 3rdness
causation, final vs. efficient (mechanical): 136; attributions of either type of cause are always hypothetical and empirical, 152–153; 'ideal' causation neither one nor the other,139n; final causation is not so-called reverse causation, 104; cybernetic devices are mechanical, 140; *see also* teleology
causation, mechanical (efficient): experience of can be direct yet still fallible, 82–83; Peirce's analysis of compared with Hume's, 83; probabilistic, 115n20; idea of 'bottom-up' causation (Searle) criticized, 294–295; causal laws are always mechanistic, 108
'cause': etymology of, 106; in philosophical usage, 107; as correct translation of *aitia*, 105–106; breadth of the

'cause' (*cont.*)
conception, as objective factor responsible for an effect, 108
causes, final: as types for which selection is made, 137–138; never particular, 97; can be agentless, 102; entail value, 153–154; vs. formal causes, 102–103; in Aristotle's philosophy, 100–101; Peirce's concept of differs from Aristotle's by making chance essential, 137; *see also* purpose
chance, used here as in chaos theory, 137n11
collateral experience (collateral observation): interpretation of diverse signs of same object, requiring indices, 192–193; implicated in correction, 193–194; not restricted to individual objects, 194–195; Peirce's inconsistent use of the term, 193n11
common sense: exists not in minds but in speech, 272; imports inspecific assumptions, 275; refutable in principle but relatively immune to change, 335–336; *see also* critical common-sensism
concepts: preceded by speech, 273–274; not psychological and therefore may incorporate social and external determinants of meaning, 280–281; reference to the actual built into some, 281–283; as grasped transcend the grasping, 282–283; in Peirce's semeiotic, 282–283
conditionals, counterfactual and subjunctive: truth of, 87; Stalnaker-Lewis theory criticized, 88–89; Peirce's theory of is a phaneroscopic analysis of experience, not a logical analysis of concepts, 89; *see also* law; 3rdness
consciousness: as feeling, 311; as feelings involved in self-control, 311–312; self-control requires signs, which thus become the stuff of consciousness, 312; *see also* inwardness; mind; self
continuity: concept of, 357n7; experience of combines sensation and thought, 80–82; experience of continuity is itself continuous, 82; thought's role sometimes analytic, sometimes synthetic, 85
convergence of opinion: Peirce's concept of, 339, distinguished from Putnam's, 341; Kuhn's argument against, 341; its continuance not guaranteed, 343–344; *see also* truth
critical common-sensism, 275–276; *see also* common sense
cosmology, Peirce's, 138n12
counterfactuals, *see* conditionals

degenerate, vs. genuine, 89–90
delome, *see* seme/pheme/delome; *see also* argument
determine: as meaning to limit (objects determine signs, and signs interpretants, in this sense), 167; Peirce's struggle to determine its meaning, 165–168
dicisign, *see* rheme/dicisign/argument

economics of research, 343n15
ellipticity, 72
entia rationis: introduced by hypostatic abstraction, 267; in empirical science, 268; sometimes real, 269
'exists', often used broadly, 87; for narrow use, *see* actuality; 2ndness
explanation: fundamental division is into mechanistic and anisotropic, 116; the principles of explanation form an emergent hierarchy, 144–145; always of aspects (Hempel), 125
explanation, forms of: anisotropic, 115; mechanistic always invokes laws relating particulars to particulars, 96–97; nomological, 97; probabilistic, 115; *see also* explanation, statistical; explanation, teleological; natural selection
explanation, statistical: standard models, 112–113; Salmon's models, 113–114; Railton's model, 114–115; mechanistic and anisotropic forms of distinguished, 115; anisotropic statistical explanation in statistical mechanics contrasted to teleological explanation, 123–124; *see also* statistical mechanics
explanation, teleological: explanation by final causes, a form of anisotropic statistical explanation, 138; Wright's analysis, 142–143; Brandon's analysis,

143–144; *see also* causation, final vs. efficient; cause, final; teleology
'external', Peirce's early uses of, 38n8

facts, concept of presupposes acquaintance with assertion, 247
fictions, not unactualized possibilities, 269–270
final cause, *see* cause, final
'finious', designates irreversibility less accurately than does 'anisotropic', 117–118
1stness: as monadicity and as quality of feeling, 75–76; of complexes, 76; 1sts are possibilities, 76, not reducible to their occurrences, 76, yet fully determinate, 78–79
foundationalism: 318; weak foundationalism attributed to Peirce wrongly, 337–338
foundherentism (Haack), 337n12
freedom: Enlightenment ideal of as autonomy, adopted by Peirce, 346; depends on possibility of objective inquiry, 347; requires faith, 347; always imperfect, 347
functionalism, *see* mind, contemporary theories of
fuzziness, *see* vagueness

generality: Peirce's use of 'general', xviii; positive (3rdness) vs. negative (1stness), 79; positive generality is the indeterminate, the continuous, and entails law, 79; in individual existence, 80
genuine vs. degenerate, 89–90
ground of significance (prior relation of a sign to its object): 162; distinguished from significance, 53; requires reality of potentiality, 53, and final causation, 54; fallibility of, 160–161

haecceity, 50, 77–78; *see also* 2ndness
hypoicons: a subset of iconic sinsigns, 216; images, diagrams, metaphors, 218
hyposemes (or subindices): indexical legisigns, 220; includes pronouns, 223, and proper names, 224–225
hypostatic abstraction: a tool of thought, 265; in second-order logic, 265–266; distinguished from prescisive abstraction, 266–267; in mathematics and empirical science contrasted, 267; does not always introduce *entia rationis*, 267–268; in empirical science, a limiting case of abduction, 268; varieties of, 269
hypostatic abstractions, scholastic (SHAs): 269; introduce actualities or fictions, never unactualized possibilities, 270; the warp of thought, 270; sometimes abstracted from other SHAs, 271; normally unformulated, 271–272; use of in intellectual history, 272; found in common sense, 272, abstracted from verbal practices, 273; inspecific, 274; designate rigidly, 279

icon/index/symbol, division based on grounds of significance, 214; *see also* icon; index; symbol
icon: sign that signifies by its own qualities, 215; pure cannot be likenesses, 215–216; impure are likenesses, samples, examples, 218; pure (e.g., music), a limiting case of sign, 205; icons as embodied 1sts vs. icons as the 1stnesses of 2nds and 3rds, 217–218; each 2nd or 3rd is many icons, 217; iconic legisigns (diagrams), 223; iconic qualisigns, 217; iconicity defended from Goodman's arguments, 215n3; some complex cases in the arts, 216–217
idealism, *see* realism
incommensurability: problem of, 322–323; its assumption of 'basic' theories undercut by recognition of SHAs, 333–334
index: introduction of, 47–49; a component of knowledge, 49; discovery of led to phaneroscopy and the category of 2ndness, 49–50; inadequate concept of in the 'New List', 48n13; not necessarily compulsive, but always in existential relation to another existent, 219; pure and impure, 219–220; role of in a symbol's growth in meaning, 286
indexicals, 222–223
individuals: Peirce's early theory of as general, 38–40; the 'absolute' individual is not general but also not

individuals (*cont.*)
real, 38–39; early theory criticized, 39; Peirce's later theory of, in which individuals are reconceived in terms of haecceity, 50–51, and as law-governed continua of 2nds, 87; includes genotypes (Ghiselin), 129n6
inspecificity: a variety of vagueness distinct from fuzziness, 274; inspecific concepts are often indefinitely applicable without being fuzzy, consistent with each of many mutually inconsistent theories of the same thing, and less open to doubt than more specific concepts of the same thing, 274–275
intentionality: 6–7; intentional inexistence, 7; Brentano's two theses, that intentionality is the mark of the mental, 7–8, and that the mental is inexplicable, 9–11; nonpsychical phenomena that seem to possess intentionality, 9–11; formal mode criteria of, 11–15, 16, 174; Chisholm's project analyzed, 12–13; Quine on intentional idioms, 13; intentional verbs, 14–15; formal mode criteria applied dialectically, 14; not dependent on thought, 175; explained by purposeful action, 175–177; Putnam's argument against evolutionary explanations of intentionality not germane to Peirce's view, 309–310
interpretant: distinguished from interpretation, 18; equivalence of distinct interpretants relative to interest, 18n5; Peirce's introduction of the term, 29–30; extended to include actions and feelings, 52; infinite progression of interpretants eliminated, 56; may be mistaken in either of two ways, 159–160; may be of any category except the impossible, 163–164; always has a purpose, which is the interpreter's, 171–172
interpretants, divisions of: reasons for, 178–179; two trichotomies of are frequently presented in the same passages as distinct, 180–181; modal argument for conflating the two trichotomies rebutted, 179
interpretants, emotional/energetic/logical
emotional: 204–206; idea of teleological and realistic, 205; thus, emotions are cognitive as such, and not as judgments, 206
energetic: 201–204; idea of teleological and realistic, 203–204
logical, 57
ultimate: a subdivision of logical, 57; as changes in habit, 58–59, 173; necessity for, 172–174; distinct from final interpretant in definition even when same in fact, 178–179; *see also* pragmatism
interpretants, immediate/dynamic/final
immediate: 54–56, 187–188; changes in conception of, 181–182; determined by grounds of significance, 189; distinguished from final by different types of mistake in interpretation, 183–184
dynamic: 183, differs from final in definition even when identical in fact, 188
final: 182–183, as that to which other signs are relevant, 190; changes in conception of, 182–183; distinguished from ultimate, 57–58; 'final' not always the best term for, 190n7; may be more than one per sign (contrary to Peirce), 190, not all of which cohere, 190; despite Peirce's tendency to identify with 'the final opinion', never the whole truth, 190; sometimes called 'normal', 183; relation to dynamic object varies, 202
interpretation: two senses of, 156–157; problem of arbitrariness of, 43; that problem solved, 56; as translation, error of exclusive attention to, 156; can be mistaken, 157; mistaken types distinguished, 189; entails intentional inexistence, 174; not always intellectual, 201; rules of are not interpretants, 221; variety of relations that justify, 161; relative to purpose yet can be shared by different, even antagonistic purposes, 189
interpreter, extended to include other animals as well as humans, 52–53
interprets, 'R interprets X as a sign of O' defined, 157–159

inwardness: as voluntary inhibition, observable in self-control, 314–315; *see also* consciousness; self

kinds: real vs. nominal, 87; artifactual (Wiggins), 278; natural are designated rigidly, 278, but contrary to Kripke et al., there is for each an SHA true of exactly it, 279

law: can be apprehended only in a symbol, 85; not reducible to regularity, 87; *see also* conditionals; continuity; reality; 3rdness
legisign: a law established to signify (Peirce's usual and our stipulated use; cf. legisign(G)), 210; defined by formation rules subordinately to rules of interpretation, 212–213; excludes causal laws, 211; and Augustine's *signa data*, 26, 211–212; many not conventional, 211; not all are symbols, 222–223; significance of consists in their existing for the purpose of signifying, 210; *see also* qualisign/sinsign/legisign; replicas
legisign(G): any law that is a sign, 210; term introduced to resolve ambiguity of Peirce's broader and narrower use of 'legisign', 210
lekton, Stoics' concept of, 23

'matter', changes in meaning of, 95–96
meaning: meanings of, 162, 263; in semeiotic, the meaning of a sign is its immediate interpretant, 263; contrary to many, the translation theory is not a theory of meaning, 44; Grice's division of meanings restated in Peirce's terms, 213–214; grows with use, as a function of knowledge, 264; 'the' meaning of a term – a dubious concept, 283; Putnam's argument that 'meanings are not in the head' examined, 279; *see also* pragmatism; pragmatic maxim
mechanicalism, 98; *see also* physicalism
mechanics: changes in conception of, 95; science of, 95–96
mechanistic: 96; excludes the teleological 95, 97
mind, contemporary theories of: as 'inner representation' – a lingering Cartesian error Peirce long ago exposed, 301–302; neural-identity (Place, Armstrong, et al.), 291–292; eliminative materialism (Churchland et al.), 292; Dennett's theory, 292–293; Kim on supervenience, 293; Searle's theory, 293; functionalism (Putnam, Fodor, et al.), 292, its computer analogy, 292, and its problem with content, 295–297; teleological functionalism (Papineau, Millikan, et al.), 297–298, attempts to square it with mechanicalism, 299–300, requires Peirce's concept of final causation,301; Fodor's critique of teleological functionalism criticized, 298–299, as it succeeds only against timid versions,300; Dennett's response to Fodor's critique fails, 300–301
mind, Peirce on: his broad usage of 'mind' and 'mental', 290; writings on distinguished from implications of his semeiotic, 290–291; his 1892–3 theory of, 291n1; mind as observable, 295; mind as semeiosis of a higher order, 302; 'inner representations' required only when not acting is an option, 303; *see also* consciousness; intentionality; inwardness; thought; self
modality, *see* categories; conditionals
morality: application of semeiotic to moral discourse runs counter to the subjectivism of modern thought, 206; extension of analysis of commands to analysis of moral duty, 204; moral realism a corollary of treating moral feelings as emotional interpretants, 205–206, 214; 'discourse ethics' (Habermas), 255n15; *see also* value
music: as iconic, 204; as feeling contemplated, 204; dynamic and final emotional interpretants of, 204–205; emotional interpretant of identical to the feeling embodied in the sound, 204; its immediate and dynamic objects are identical, 205; logical interpretants of are deficient and inessential but useful, 204; reality of what it represents, 205; *see also* art

natural kinds, *see* kinds
natural selection: Peirce on central idea of, 128; 'selection of' and 'selection for' (Sober), 130; anisotropic, not mechanistic, 130–132; fundamental theorem of (Fisher), 132; tautology in, 124; consists of mechanical events, 132; no mechanism of, 132–133; improbability of types selected-for, 133–134; without purpose or direction, 145; nor do species or their members have a purpose, 145–146; controversy over reveals misunderstanding of dynamic nature of science, 327n4
naturalized epistemology (Quine): scants normative questions, 320–321; anticipated by Peirce (without scanting normative questions), 320n2
nonbeing: problem of the nonexistent object in Peirce's early theory, 42, 46; how to speak of what is not, 175–176
nonequilibrium thermodynamics (Prigogine), 134–135

object, breadth of the conception, 162–163
objectivity: the word's change in meaning and its new application to inquiry, 324; defined, 324–325; evolution of methods and aim poses problem for the defense of, 344; *see also* theory evaluation; value, objectivity of
objects, division of: immediate vs. dynamic, 191; immediate corresponds to immediate interpretant, dynamic to final interpretant, 191; distinguished in terms of process of interpretation, 191–192; commands have both objects, 201–203
objects, dynamic: explains difference between success and failure of interpretants, 191; Pierce conceived of as the object of the final opinion (qua complete knowledge), 195; contrary to Peirce, defined as that which can be signified by diverse signs and which explains differences between final and immediate interpretants, 195–196; no sign lacks one, 196; cannot be misrepresented by pure icons and pure indices but cannot be represented by them completely, 196; doctrine of embodies Peirce's realism, 199
objects, immediate: the Stoics' *lekton*, 179–180, 191; the immediate object *is* the dynamic object *as* represented, 196; need not obtain, 159; specified but not perfectly specific, 159
observation, generality of, 124–125; *see also* perceptual judgments
observation, justification of: agreement of observations is a logical relation among judgments, 68n6; distinction between justifying observations and justifying relying on observations, 338; convergence of opinion certifies reliability of observation, 339; observations are relied on without justification, 339, and that reliance is then justified by theories that explain observations and their agreement, 339; sources of erroneous view that there are warrants for individual observations, 338n13

percept and percipuum, 319n1
perceptual judgments: as 'first judgments' of what is before one's senses, 51; fallible, 51–52; contains general ideas and introduces assumptions, 318–319; an extreme case of abduction, 319–320; occurs uncontrollably but can be corrected later, 320; occurs without warrant, 337–338; semeiotic analysis of, 339–340
phaneron, the: vs. Lockean ideas, 66–68; does not exclude thinking, 81; its observation, 68, and description, 70
phaneroscopy: and phenomenology, 60–61; depends on algebra of relations, 64–66; presuppositions of, 70; not a form of intuitive knowledge, 61n3; to understand, reader must repeat observations himself, 70–71, 76; *see also* phaneron; phenomenology
pheme, *see* seme/pheme/delome
phenomenology: Continental, 8–9; problem of phenomenological description in Husserl and Peirce, 61; *see also* phaneron; phaneroscopy

physicalism: ambiguity of, 294; the physical not necessarily mechanical, 294
pragmatic maxim: 263; correctly entails inexhaustiblility of meaning, 58, that the list of verification conditions continues to grow, 288
pragmaticism, *see* pragmatism
pragmatism (pragmaticism): not a general theory of meaning, 57; contrasted to logical positivism, 84; not a theory of how reference is fixed, nor a verifiability theory of meaning, 287–288; makes practice to be for the sake of theory, 173; reformulated in 1907 as the doctrine of ultimate interpretants, 56; *see also* pragmatic maxim
prescinding, 71–72
prior relation of sign to its object, *see* ground of significance
probabilistic causation, *see* causation
propositions: Peirce and Austin on, 243–244; Peirce's inconsistent use of the term, 244–245; as distinct from its expression, not a sign, 245–246; reality of questioned, 246n8; *see also* assertion; statements
'purpose', ordinary usage of: 108–110; survives Darwin's theory, 109; 'having a purpose', 110; 'acting purposefully'/'acting for a purpose'/'purposeful action', 111; 'used for a purpose', 112; 'serving a purpose', 111, 'existing for a purpose', 111
purpose: as type, 92–93; not psychological, but objective, 93; defined, as type of outcome for which an agent acts or for which something is selected as a means, 110; not every final cause is a purpose, 135; does not have to be conscious, 110; can be irrational, 149, or absurd, 164; cannot be described without use of intentional idioms, 174–175; *see also* cause, final
purposeful actions, bases of, 155
purposefulness, evolution of: 146–150; emancipation of purpose from biology, 148–150; how some purposes come to be valued over others and endure or prevail, 148–150
qualisign/sinsign/legisign: division based on what a sign is in itself, which may be of any category, 209; qualisigns are mere possibilities, 209; sinsigns are single things or events, 209; legisigns, *see* legisign; legisign (G)
realism, and idealism: thing-in-itself vs. reality in Peirce's early, idealistic theory, 37, 38; Peirce relinquished his early form of idealism, 46–47; his subsequent use of the terms 'conditional idealism' and 'objective idealism', 47n12; idealism a root of his theory of inquiry, 324; typical arguments for subjective idealism, 196–197; Goodman's argument for semiotic idealism, 197; limitations of Putnam's defense of realism against Goodman's argument, 307n6; central argument for subjective idealism refuted, 198–199
realism, internal vs. metaphysical (Putnam): 199–200; Kuhn's 'post-DarwinianKantianism' an example of metaphysical realism, 342
realism, Peirce's: defined, 199; embraces but is distinguished from scientific realism, 199; Fisch on, 199; required both by purpose and by semeiosis, 200; neither internal nor metaphysical, 342–343; *see also* law; reality; 3rdness
realism, scientific: 197–198; argument for not internal to science, 340
realism vs. nominalism: xvii; Peirce's early discussions of, 40–42; his idiosyncratic definitions of these two doctrines, 40; his characterization of their 'theories of reality', 40–41; his failed attempts to reconcile those theories, 41–42
reality: 'real' defined, 41; as 3rdness, 86; no reality without actuality, 87; of processes and events, 87; *see also* law; 3rdness
reference: 'reference' refers to different things, 263; in semeiotic, the referent is the dynamic object, 263; fixed independently of meaning, 264, which depends on indexical signification, 266; 'traditional' theory of, 267; 'new'

reference (*cont.*)
or 'causal' theory of anticipated by Peirce, 267; causal theory clarified (reference fixed existentially, not always causally), 277–278; irreducible to psychology, 282–283
relations: analysis of, 72–73; reduction of, 73; some are irreducible (indecomposable), 73; external structure of, 73–74
replicas (instances of legisigns): can be produced only for purpose of replicating, 212; signify legisigns iconically and indexically, 212; indexical legisigns signify through their replicas, 213, yet the immediate interpretants of the two differ, 223–224; legisigns and their replicas do not entail two layers of significance, 223–224, posing a problem for Peirce's semeiotic taxonomy, 225
representamen, 55n18
reverse causation, *see* causation, final vs. efficient
rheme/dicisign/argument: a generalization of term/proposition/argument, 231–232; variant designations, 232; distinguished not by differences in compositional complexity but by mode of influence on interpreters, 233–234; *see also* argument; seme/pheme/delome
rigid designation (Kripke): 276–277; untoward consequences of its divorce from conception, 283–285

science as inquiry: developed in modern period, 326–327; in modern science, the purpose of theory is to advance inquiry, 327; Peirce defined science by its 'spirit' rather than by its method, 328; breadth of Peirce's conception, 328; scientific methods vary in objectivity, 328–329; aim of modern science remains intellectual, despite claims of many, 329; dynamic nature of modern science, 329–330; its evolution in methods and social forms, 329–330
scientism: Putnam's critique of and Peirce's avoidance of compared, 306–309
2ndness: as dyadicity and as two-sided experience of effort and resistance, 76–78; irreducibly dyadic, it is a fact of complexity, not a complex of facts, 77; contrasted to Locke's idea of solidity, 77; pervasive in experience, 77, contrary to what Hume implied, 77; additional to 1stness in the occurrence of a 1st, 78; as actuality, 78
self, the: a hypothesis introduced to explain ignorance and error, 312; self-control an observable process of which the self is not the agent but a product, 312–313; as 'teleological harmony of ideas' and semeiotic process, 313–314; as an *ens rationis* abstracted hypostatically from facts about control, on which higher grades of control depend, 314–316; an *ens rationis* consequential and therefore real 316; *see also* consciousness; inwardness; mind, Peirce on
seme/pheme/delome: 232; reason preferable to rheme/dicisign/argument, 248
semeiosis: as sign's 'action', 172; purpose a fourth element in, 158; the purpose essential to it is that of an interpreter, 171–172
semeiotic (Peirce's theory of signs): a science, 151; overextension limits its explanatory power, xvi, 177; development toward naturalism, 53; provides a naturalistic theory of mind, 290; not behavioristic, 289, nor reductive,290; does not make signs in general to have a function, 309–310
semeiotic taxonomy, principles of: 235; can only be justified a posteriori, 207–208; each sign must be of one division of each trichotomy, 232; no sign may be of more than one division of any trichotomy, 231; other forbidden combinations, 235–236; need to explain the striking pattern of adicities, 237–242; order of components in the sign relation, 238–242; relative simplicity, 239; determination, 240; the ten classes of 1903, 236–237; the ten trichotomies of 1908 criticized, 259–260

semiology (Saussure): 16; assumption that a sign is a two-part entity 16–17; assumption of arbitrariness, 17; relation of thought to language, 17–18
semiology vs. Peirce's semeiotic: danger of their conflation, xv–xvi; dyadic vs. triadic, 18; compositional vs. relational, 18–19; difference in breadth, 19–20; each aims to be a science, 20; absurd consequences of failure to see their fundamental opposition, 20–21; Eco's conflation of Saussure's semiology with Peirce's semeiotic criticized, 228–230
semiotic: non-Peircean sign theories derived from Saussure's semiology, xin1
SHA, *see* hypostatic abstraction, scholastic
sign: a technical term, 20, 151; Peirce's unchanging conception of, as one part of a triad, 30; extended to include natural effects and resemblances, 52; defined, 160; Peirce's definitions of, 164–165; as defined here agrees with Peirce's tendencies, 168; breadth of Peirce's conception, 185–186; signs need only be interpreted potentially, 161; not always produced purposefully, 186; how counted, 161, 188–189; may be of any category, 163, 209; reference to can be either opaque or transparent, 188; ambiguity of 'false sign', 160n4; false or misleading signs correspond to a type of mistaken interpretation, 189; genuine vs. degenerate, 230–231; compound, 161–162; 'sign to' a deceptive locution, 227; suppositions that icons and indices require a symbolic component (Eco) or that signs 'blend' (Jakobson) or that there is a 'perfect' or 'complete' sign (Peirce) refuted, 225–227
sign, other theories of: ancient Greeks, 21; Plato, 21–22; Aristotle, 22; Stoics, 23 (*see also lekton*); Epicureans, 23; Augustine, 23–26 (*see also signa data*); Locke, 2–4; Millikan, 303–306; Morris, xv, 19n6, 306; for Saussure, *see* semiology
signa data (Augustine): vs. *signa naturalia*, 24–26; term usually mistranslated, 24–25
significance: Peirce's early failure to explain, 43–44; not dependent on actual interpretation, 53; defined as grounded interpretability, 53, 162; still triadic because distinct from its ground, 214–215; a form of intentionality, 174; derivative from interpretation's purpose, 172–174; not dependent on 'inner representation', 301
space and time: perception of, 81n12; not presupposed by indices but apprehended by coordinating indices successfully, 192n9; spatio-temporal location, 97n2
statements, distinguished from sentences, 242–244; *see also* assertion
statistical inference, 103n8
statistical mechanics: reasoning in, 113, 118–123; ergodicity and ensembles, 121; quantum mechanics in, 121–122; explanation in is not by forces, and therefore not mechanistic, 122–123; introduces a new way of looking, 124; not *faute de mieux*, 125–127; Second Law, though explained statistically, is mechanistic, 127–128; *see also* explanation, statistical
statistical phenomena: reality of, 128; potency of, 133
symbols: legisigns whose objects are assigned by rules of interpretation, 220–221; signify types, 222; replicas of are not symbols, 221, yet may signify types through symbols replicated, 224; their replicas, being sinsigns, may signify individuals, 222; growth of, 285–288; as purpose, 287–288; their 'essences' are SHAs, 287

teleology: etymology of, 98; Aristotle's philosophy its *locus classicus*, 98; a sophisticated doctrine, 98–100; introduced to explain the emergence of order from chaos, 99; misunderstood because of mechanist assumptions, 100; Plato's teleology, 99–100, 101–102, 103, 103n8; theistic versions, 103–104; Peirce's sheriff

teleology (*cont.*)
analogy not theistic, 139n; Aquinas' versus Aristotle's, 105; contrasted to vitalism, 104; Kant's, 107; attempted mechanistic reductions of, 104; influence of the ideal on the actual, 94; *see also* cause, final; explanaton, teleological
teleonomy (Mayr et al.), 141
theorematic vs. corollarial reasoning, 265–266
theory evaluation: growth more important than surviving tests, 336; role of research programs (Lakatos) and relevance of inspecificity to, 336–337; tests by observations presupposing SHAs, 275, 334; crucial experiments, 334–335; Kuhn's view of theory choice, 342
3rdness: triadic, as in combination, 84; continuity 'the perfection of', 84; the whole of a 3rd is irreducible to its parts, 84–85; *see also* causation, mechanical; conditionals; continuity, experience of; law; reality
thought: as internalized discourse, 4–5; the 1868–9 doctrine of thought signs, 34, and its problem accounting for individual thoughts, 36
transcendental argumentation, Peirce's rejection of, 66
truth: its definition seen as evolving, 331–332; its definition as impersonal is a cultural development, 333; defined as that toward which objective inquiry progresses, 325–326, which definition is not a theory of truth in the contemporary sense, 332, but is rather an ideal that Peirce recommended,333; does not depend on actual convergence, 326; no guarantee a priori that truth in this sense exists, 331–332; Peirce anticipated the deflationary theory, 332–333; 'truth in' a treacherous locution,88n16; *see also* convergence of opinion
type/token distinction, same as legisign/replica, 209

universals and particulars, xvi–xviii

vagueness, as fuzziness, 274; as lacking specificity, *see* inspecificity
value: some dimensions of implicit in significance, 154–156; objectivity of, 154, 344–346; *see also* art; morality; music; objectivity; teleology; theory evaluation; truth