Investigating the Psychological World

Scientific Method in the Behavioral Sciences

problem data phenomenon theory model explanation problem data phenomenon theory model explanation problem data phenomenon theory model explanation problem data phenomenon theory model explanation problem data phenomenon theory model explanation theory model explanation problem data phenomenon theory model

Brian D. Haig

Investigating the Psychological World

Life and Mind: Philosophical Issues in Biology and Psychology
Kim Sterelny and Robert A. Wilson, Series Editors

Investigating the Psychological World: Scientific Method in the Behavioral Sciences, Brian D. Haig, 2014

Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life, revised edition, Eva Jablonka and Marion J. Lamb, 2014

Cooperation and Its Evolution, Kim Sterelny, Richard Joyce, Brett Calcott, and Ben Fraser, editors, 2013

Ingenious Genes: How Gene Regulation Networks Evolve to Control Development, Roger Sansom, 2011

Yuck! The Nature and Moral Significance of Disgust, Daniel Kelly, 2011

Laws, Mind, and Free Will, Steven Horst, 2011

Perplexities of Consciousness, Eric Schwitzgebel, 2011

Humanity's End: Why We Should Reject Radical Enhancement, Nicholas Agar, 2010

Color Ontology and Color Science, Jonathan Cohen and Mohan Matthen, editors, 2010

The Extended Mind, Richard Menary, editor, 2010

The Native Mind and the Cultural Construction of Nature, Scott Atran and Douglas Medin, 2008

Describing Inner Experience? Proponent Meets Skeptic, Russell T. Hurlburt and Eric Schwitzgebel, 2007

Evolutionary Psychology as Maladapted Psychology, Robert C. Richardson, 2007

The Evolution of Morality, Richard Joyce, 2006

Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life, Eva Jablonka and Marion J. Lamb, 2005

Molecular Models of Life: Philosophical Papers on Molecular Biology, Sahotra Sarkar, 2005

The Mind Incarnate, Lawrence A. Shapiro, 2004

Organisms and Artifacts: Design in Nature and Elsewhere, Tim Lewens, 2004

Seeing and Visualizing: It's Not What You Think, Zenon W. Pylyshyn, 2003

Evolution and Learning: The Baldwin Effect Reconsidered, Bruce H. Weber and David J. Depew, editors, 2003

The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain, William R. Uttal, 2001

Cycles of Contingency: Developmental Systems and Evolution, Susan Oyama, Paul E. Griffiths, and Russell D. Gray, editors, 2001

Coherence in Thought and Action, Paul Thagard, 2000

Investigating the Psychological World

Scientific Method in the Behavioral Sciences

Brian D. Haig

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu.

This book was set in Sabon LT Std by Toppan Best-set Premedia Limited, Hong Kong. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Haig, Brian D., 1945–.
Investigating the psychological world : scientific method in the behavioral sciences / Brian D. Haig.
pages cm.—(Life and mind : philosophical issues in biology and psychology) (A Bradford book)
Includes bibliographical references and index.
ISBN 978-0-262-02736-6 (hardcover : alk. paper)
1. Psychology—Research—Methodology. I. Title.
BF76.5.H335 2014
150.72'1—dc23
2013032413

10 9 8 7 6 5 4 3 2 1

The attempt to understand and improve methods, and to do so via theorizing them, is at the center of an intelligently evolving cognition.

—Clifford Hooker (1987, 291)

Above all, if a raised standard of education in methods is to be achieved, it is necessary to engender, beyond any knowledge of particular skills and formulae as such, a *perspective* as to what methods are most appropriate to various areas and occasions.

—Raymond Cattell (1966, 5)

Contents

Preface

Although modern science is made up of many parts, scientific method is its centerpiece. The centrality of method to science stems from the fact that it provides scientists with the primary form of guidance in their quest to obtain knowledge about the world. As fallible inquirers, scientists face immense challenges in their efforts to learn about the complexities of nature. In good part, these challenges are met through the use of methods, which provide scientists with the cognitive assistance that they need to undertake successful inquiry.

However, despite its undoubted importance, scientific method receives less considered attention than it deserves, from both scientists and educators. Of course, scientists take method seriously, but I believe that they do not take it seriously enough. Scientists themselves, including psychologists, learn about research methods and how to use them to conduct their research. However, the nature of this learning, and of the instruction they receive about how to employ these methods, is better described as a mix of training and indoctrination than as a genuine education designed to provide a critical, in-depth understanding of the methods. Although professional science educators sometimes promote the importance of the epistemological foundations of scientific method, the influence of this source of learning on the regular teaching of research methods is minimal. Psychology, which provides extensively in its curriculum for teaching research methods, uses textbooks that make little or no effort to inform students in depth about the nature of scientific method. Nor does its curriculum foster a critical appreciation of the various research methods that its textbooks deal with. Consequently both psychological scientists and psychology students tend to have a limited understanding of scientific method, which in turn contributes to a misuse of research methods and a suboptimal level of scientific literacy.

I think that the missing key in this educational failure is scientific methodology. Methodology is the domain officially charged with fostering the evolution and understanding of scientific methods, and it is our official repository of knowledge about those methods. Scientific methodology is not the exclusive domain of any particular discipline. Rather, it is a central part of cognitive theory, which is itself regarded as an interdisciplinary endeavor. It spans the domains of statistics, the philosophy of science, the sociology of science, the various disciplines of cognitive science, and more; but it is reducible to none of them. As a practical endeavor, methodology is concerned with the mutual adjustment of means and ends. It judges whether methods are sufficiently effective for reaching certain goals. But methodology is also critically aim oriented and considers what goals the research enterprise should pursue. Clearly no single discipline can realistically aspire to cover all the tasks of methodology.

The methodological literature in psychology is dominated by the field of statistics. Quantitative methods receive the large majority of attention in both research methods textbooks and research practice. Qualitative research methods are regarded as a poor cousin and remain on the margins of methodology, although there are signs that they are gaining some acceptance. As important as statistical methods are to science, they cannot be all that there is to scientific method. Consequently the clarion call for statisticians to be the purveyors of scientific method (e.g., Marquardt, 1987) is inappropriate. The guiding assumption of this book is that treating scientific method with the seriousness it deserves requires taking scientific methodology seriously. I do this by giving special consideration to behavioral science methodology, the philosophy of science, and statistical theory. Thus the book is interdisciplinary in nature.

The philosophy of science figures more prominently in this book than is usual for methodology texts. The reason for this emphasis is that contemporary philosophy of science contains an array of important methodological insights that are impossible to ignore when coming to grips with scientific method. In recent years, philosophy of science has increasingly sought to understand science as it is practiced, and although it has much work to do in this regard, it now has important things to say about how science is, and should be, conducted. As part of this concern with scientific practice, philosophers of science have given increased attention to research methods in science. A positive development in this regard has been the focus on the methodology of experimentation over the last thirty years, although the methodology of theory construction remains the dominant focus in the philosophy of science.

Of late, philosophers of science have also shown a willingness to deal with methodological issues in sciences other than physics. Biology has been the major beneficiary, although psychology has received some philosophical attention. There is, then, a developing literature in contemporary philosophy of science that can aid both our understanding and our use of research methods and strategies in psychology (e.g., Trout, 1998). At the same time, a small number of theoretically oriented behavioral and social science methodologists have produced work on the conceptual foundations of research methods that helps illuminate those methods. Thus the work of both professional philosophers of science and theoretical scientists should be included in a philosophical examination of behavioral research methods.

Three major philosophies of science are of particular relevance to psychology: empiricism, social constructionism, and scientific realism (Greenwood, 1992; Manicas & Secord, 1983). Nineteenth-century British empiricism had a major influence on the development of British statistics in the first half of the twentieth century (Mulaik, 1985). The statistical methods developed in that intellectual milieu remain an important part of psychology's statistical research practice. For example, Karl Pearson's product moment correlation coefficient was taken by its founder to be the quantitative expression of a causal relation viewed in empiricist terms. Similarly, Ronald Fisher's endorsement of inductive method as the proper view of scientific method stemmed from a commitment to the empiricism of his day. Even in today's postpositivist philosophical climate, authors of research methods textbooks sometimes portray quantitative research as essentially positivist in its empiricist commitments (see Yu, 2006). The traditional empiricist outlook is much too limiting because it restricts its attention to what can be observed, and regards theories merely as instruments that organize claims about observables.

For their part, qualitative methodologists tend to bolster their preferred conception of qualitative research by comparing it with an unflattering positivist picture of quantitative research. At the same time, they frequently adopt a philosophy of social constructionism that is expressed in an implausibly strong form. This form is opposed to the traditional notions of truth, objectivity, and reason and maintains that our understanding of the world is determined by social negotiation. Such a view of social constructionism tends to be employed by those who are opposed or indifferent to quantitative methods. It is a view at odds with the philosophical outlook adopted in this book.

In what follows, I adopt a scientific realist perspective on research methods. Although the subject of considerable debate, and opposed by many antirealist positions, scientific realism is the dominant philosophy of science today. In addition, a commonsense version of realism seems to be the tacit philosophy of most working scientists. With its increasingly heavy emphasis on the nature of scientific practice, the philosophy of scientific realism is becoming a philosophy *for* science, not just a philosophy *of* science. Scientific realism is, in fact, a family of positions, and in chapter 1, I sketch a view of realism that I think is appropriate for psychology. Scientific realism boasts a rich conception of methodology, which can be of considerable help in understanding and guiding behavioral science research. It is a methodology that is at once naturalistic, problem focused, and aim oriented. It also promotes both generative and consequentialist reasoning, and the importance of justifying knowledge claims on both reliabilist and coherentist grounds. The influence of this conception of methodology occurs throughout the book.

In this book, I take psychology's commitment to scientific method very seriously. I do this principally by constructing a broad theory of scientific method, which is genuinely informed by insights in contemporary scientific methodology and speaks to the conduct of psychological research. This account of method I call the *abductive theory of method* (hereafter *ATOM*) in recognition of the importance it assigns to explanatory reasoning. In contrast to the popular hypothetico-deductive method, ATOM portrays research as a bottom-up process comprising two broad phases. The first phase involves the detection of phenomena, such as empirical generalizations. The second phase involves the construction of explanatory theories to explain claims about the phenomena. The book draws from the "new experimentalism" (Ackerman, 1989) in philosophy of science to help illuminate the process of phenomena detection. It also examines in detail different abductive methods of theory construction, drawing, where appropriate, from the varied philosophical literature on abductive reasoning: the widely used method of exploratory factor analysis is presented as an abductive method of theory generation; the strategy of analogical modeling is presented as an abductive approach to theory development; and the neglected method of inference to the best explanation, particularly the theory of explanatory coherence, is presented as an appropriate method of theory appraisal.

An important feature of ATOM is that it functions as a broad framework theory within which a variety of more specific research methods can be located and employed. A coherent treatment of those methods is enhanced by placing them within the framework of ATOM. In turn, the specific methods help give ATOM a good deal of its operational detail. A number of the specific methods I refer to are well known to behavioral scientists, but some are not. Psychology has tended to emphasize data analytic methods at the expense of methods of theory construction. However, ATOM assigns equal importance to the two classes of method.

A subsidiary focus of this book is a concern with science education in relation to behavioral research methods. It follows John Dewey's (1910) lead and suggests that we adopt an inquiry-oriented conception of education that accords an important place to scientific method. The narrow nature of, and uncritical approach to, the teaching and use of research methods in psychology are highlighted in some of the chapters. The need to teach for a more critical understanding of research methods is a natural consequence of acknowledging the importance of the domain of research methodology. In light of the requirements of a genuine liberal education, I make constructive proposals for reforming the methods curriculum. The nature of ATOM and its methodological foundations shape many of these curriculum proposals.

Chapter 1 introduces the topic of scientific method by providing some background material to better appreciate the more focused discussion of method in the ensuing chapters. I begin by briefly considering the idea of scientific method and different criticisms that have been leveled against it. Next I outline and provisionally assess four prominent theories of scientific method. I then move to a consideration of the nature of scientific methodology before providing a selective overview of the key elements of the philosophy of scientific realism. Finally, I present a brief overview of ATOM to provide a conceptual framework for locating and better understanding the various methods and strategies examined in the book.

Chapter 2 draws from the new experimentalism in the philosophy of science so as to reconstruct the important process of phenomena detection as it applies to psychology. In doing so, I propose a four-stage model of data analysis. The model begins with the initial examination of data, proceeds in turn through exploratory and confirmatory data analytic phases, and finishes with the stage of constructive replication. The threefold distinction between data, phenomena, and explanatory theory is drawn, and its implications for understanding the nature of psychological science are spelled out.

Chapter 3 considers the abductive nature of theory generation by examining the logic and purpose of the method of exploratory factor analysis. I argue that the common factors that result from using this method are not fictions but latent variables, which are best understood as genuine theoretical entities. I support this realist interpretation of factors by showing that exploratory factor analysis is an abductive generator of elementary theories that exploits an important heuristic of scientific methodology known as the *principle of the common cause*.

Science uses many different approaches to modeling. In chapter 4, I selectively examine one important approach to scientific modeling, analogical modeling. The strategy of analogical modeling is adopted by ATOM as its chief means of theory development. Accordingly, I spell out here the structure of analogical models and the use of analogical abductive reasoning both to expand and to evaluate the plausibility of models.

Chapter 5 recommends the use of inference to the best explanation for evaluating the worth of theories in psychology. I suggest that it is a more appropriate account of theory appraisal than both the popular hypothetico-deductive method and the widely heralded Bayesian approach. I discuss a number of different explications of inference to the best explanation, in particular the theory of explanatory coherence, which is the most detailed extant explication of inference to the best explanation.

The concluding chapter rounds out the extended characterization of ATOM. First I outline an account of the nature of research problems, and then I discuss the nature and limits of ATOM. This is followed by applications of ATOM to grounded theory method and to clinical reasoning. Toward the end of the chapter, I offer some thoughts about the importance of methodology for understanding research methods. The book concludes with some brief remarks about the future prospects for ATOM.

The methodology of the behavioral sciences is a subject of relative neglect in professional philosophy of science. Thus my hope is that this book will be welcomed by those in the philosophical community who want to learn about an important set of methodological practices in one of the interesting special sciences. Conversely, I would like to think that the book contains material that will enable psychological researchers to deepen their conceptual appreciation of a variety of research methods and associated methodological matters and thereby contribute to the conduct of sound psychological research. Although the book's primary focus is on psychology, I believe its contents are relevant to the behavioral sciences more generally.

Finally, I draw the reader's attention to two matters. First, it is sometimes important to distinguish between scientific method as a theoretical understanding of an inquiry procedure and scientific method as a material practice. Given the book's primary concern with ATOM, it mostly focuses on a theoretical understanding of method. Second, I have endeavored to keep abbreviations to a minimum. However, for convenience, I abbreviate the *abductive theory of method* as *ATOM* throughout the book. I also use abbreviations for *exploratory factor analysis* and *inference to the best explanation* in their respective chapters.

Acknowledgments

In the mid-1990s, Hillary Clinton argued that it takes a village to raise a child. The assertion occasioned a skeptical response from a number of quarters, but there can be no doubting the claim that it takes a village to raise a book. This book is no exception, for it has depended on the support of many people, and a number of institutions, over several decades.

Although my interest in scientific method spans more than forty years, the book project that became *Investigating the Psychological World* began in earnest in the second half of 2011 while I was on study leave with Denny Borsboom at the University of Amsterdam. I am grateful to Denny for his stimulating intellectual hospitality during that time, and for opportunities to collaborate with him in recent years and benefit from his versatile mind.

My interest in abductive reasoning began in the mid-1960s when I read the remarkable Charles Peirce's unpopular ideas on education, and was kindled when I started to acquaint myself with some of his work on science. That interest was nourished in the early 1970s by my doctoral supervisor, Bill Rozeboom, who was one of the first psychologists to appreciate the importance of abductive reasoning in science. Bill's critical acumen as a technically accomplished theoretical psychologist has helped shape my thinking on a number of methodological issues in psychology.

More recently, the work of a number of contemporary philosophers of science has been invaluable in helping me understand some of the complexities and subtleties of scientific methodology. In this regard, the following individuals deserve special mention: Tom Nickles for his richly suggestive writing on scientific method and scientific discovery; Jim Woodward and Jim Bogen for their instructive conceptualization of the process of phenomena detection; Rom Harré for his insightful depiction of the use of models and analogies in science; and Paul Thagard for his writing on inference to the best explanation, in particular his theory of explanatory coherence, which I regard as a major methodological accomplishment. I judge these philosophers as seminal thinkers on the topics just mentioned, and their influence on what I have to say about scientific method will be obvious to the reader.

Other people have also helped me develop ideas about research methods, including Adrienne Alton-Lee, Paul Barrett, Neville Blampied, Russil Durrant, Cameron Ellis, Colin Evers, Garth Fletcher, David Funder, James Grice, Deak Helton, Stephen Hill, Hugh Lauder, Tom Maguire, Dannette Marie, Keith Markus, Joel Michell, Katharina Näswall, Claire O'Loughlin, Denis Phillips, Robert Proctor, Vivianne Robinson, Bruce Ryan, Ken Strongman, Anton Turner, Tony Ward, Juliane Wilcke, and Brad Woods. Although I have profited from my interactions with all these people, two of them deserve special mention. David Funder has strongly supported my writing on method, and more besides. As a prominent empirical researcher in personality and social psychology, his belief that I am on the right track, methodologically speaking, has been reassuring. I have also benefited considerably from Keith Markus's intellectual generosity and his ability to conceptualize methodological issues in behavioral research methods in an insightful manner.

My greatest debt of gratitude is to my wife, Fran Vertue, who has nourished and sustained me for sixteen good years. Her constant intellectual companionship has contributed in innumerable ways to the development of this book and other writing projects.

I thank Kim Sterelny and Rob Wilson for deciding that a book on the methodology of the behavioral sciences belonged in their MIT Press series Life and Mind. I am grateful to Philip Laughlin, senior acquisitions editor in philosophy and cognitive science at MIT Press, for his expert guidance on this project. Judy Feldmann, senior editor at MIT Press, gave valuable help in the final preparation of the manuscript.

Fran Vertue, Ken Strongman, and Brad Miles read a complete draft of the book and gave me very helpful feedback. Brad also assisted in preparing the manuscript for submission.

The material in chapters 2, 3, and 5 draws heavily from previously published papers. Other chapters use short excerpts from additional published work. I am grateful to the publishers of the following articles for permission to use material contained in them:

Haig, B. D. (1987). Scientific problems and the conduct of research. *Educational Philosophy and Theory*, *19*, 22–32.

Haig, B. D. (1996). Grounded theory as scientific method. In A. Neiman (Ed.), *Philosophy of education 1995: Current issues* (pp. 281–290). Urbana: University of Illinois Press.

Haig, B. D. (2005a). An abductive theory of scientific method. *Psychological Methods*, *10*, 371–388.

Haig, B. D. (2005b). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, *40*, 303–329.

Haig, B. D. (2008). On the permissiveness of the abductive theory of method. *Journal of Clinical Psychology*, *64*, 1037–1045.

Haig, B. D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *American Journal of Psychology*, *122*, 219–234.

Haig, B. D. (2013). Detecting psychological phenomena: Taking bottom-up research seriously. *American Journal of Psychology*, *126*, 135–153.

Vertue, F. M., & Haig, B. D. (2008). An abductive perspective on clinical reasoning and case formulation. *Journal of Clinical Psychology*, *64*, 1046–1068.

1 Method, Methodology, and Realism

Epistemology without contact with science becomes an empty scheme. Science without epistemology is—insofar as it is thinkable at all—primitive and muddled.
—Albert Einstein (1949, 683–684)

1.1 Introduction

Modern science is a complex human endeavor comprising many parts. It articulates aims that it seeks to realize; it employs methods to facilitate its investigations; it produces facts and theories in its quest to obtain an understanding of the world; and it is shaped by the institutions within which it is embedded. Although all these dimensions are essential to a full-bodied characterization of science, method is arguably its most important feature. This is because everything we know in science is acquired in good part through the application of its methods, whether it be our knowledge of substantive matters, values, or the methods themselves. Method really matters to science.

Although method is vitally important to the conduct of science, discussion of the topic is not particularly fashionable. There are a number of possible reasons for this. One is that some people think there is no such thing as scientific method, or at most that there is very little to scientific method; others think it cannot be given an illuminating characterization; and still others think it is a complex investigative skill that is tacitly acquired by scientists in the course of learning their craft. Attitudes such as these have some currency because scientists themselves learn very little about scientific methodology in their formal science education. Instead they tend to acquire an operational facility with a small number of "tried and proven" methods that have been judged to work well in their own specialties. The result is that a number of mistaken ideas about method have gained a foothold in our common thinking about science.

A further reason for the devaluing of methodological knowledge is that it is often walled off within specific disciplines and so loses its interdisciplinary integrity. This devaluation seems to be exacerbated by a territoriality, where specialists in particular subjects, principally the philosophy of science and statistics, sometimes proclaim or assume guardianship of scientific method itself. This is not as it should be, because methodology properly understood and practiced is a strongly interdisciplinary undertaking. Important though the insights of philosophers of science and statisticians about scientific method are, to confine one's appreciation of the topic to what they say about it is to ignore important insights about method offered by other disciplines.

Given the complexity of scientific method implied in this chapter's epigraph, it is appropriate to present some relevant background material to assist us in articulating and understanding some of that complexity. As noted in the preface, I do this by considering a variety of ideas about method, methodology, and realist philosophy of science. An overview of ATOM sets the scene for its extended treatment in the following chapters.

1.2 Criticisms of the Idea of Scientific Method

Influenced by the founders of modern scientific method, Rene Descartes and Francis Bacon, seventeenth-century methodologists understood scientific method as a universally applicable logical procedure that was at once mechanical, rule based, ahistorical, content neutral, and *a priori* (Nickles, 2009). As such, it was simultaneously thought to be a method of discovery and justification that, upon its correct application, guaranteed the production of knowledge of both the surface features and deep structures of nature.

Not surprisingly, this fanciful conception of scientific method has been subjected to strong and prolonged attack by scientists, philosophers of science, and science studies specialists. Modern methodologists have strongly challenged the features of scientific method mentioned by Nickles (2009), and more, leaving us with diminished, and still disputed, conceptions of scientific method. Larry Laudan (1981) tells a suggestive story of how in the late eighteenth century and the early nineteenth century, both scientists and methodologists largely gave up on the

Baconian conception of inductive method in favor of the method of hypothesis, or the hypothetico-deductive method. Laudan gives two reasons for this general shift: the realizations that a fail-safe method that produced infallible knowledge could not be had, and that inductive method is unable to postulate hidden causes about material things.[1]

The idea that there is a scientific method characteristic of all scientific inquiry has been attractive to many scientists, and some methodologists still speak in favor of some or other version of *the* scientific method. Because of the historical importance of the inductive and hypothetico-deductive conceptions of inquiry, and their alleged powers to produce knowledge, it is not surprising that this idea has seemed plausible.

One prominent modern candidate for the title of *the* scientific method is Karl Popper's (1959) falsificationist construal of the hypothetico-deductive method, understood as a general strategy of conjecture and refutation. Presented as an all-purpose account of method, it promises to unify method within and across the natural and social sciences. Popper's method has the additional attraction of providing the demarcation criterion of falsifiability for distinguishing scientific practice from pseudoscientific, as well as nonscientific, practice. Despite its endorsement by a number of prominent scientists (some Nobel prize winners among them), Popper's account of method is less influential in science than is commonly believed. This is especially true of psychology (Uchino, Thoman, & Byerly, 2010).[2] Moreover, philosophers of science have largely rejected Popper's falsificationist theory of science and its depiction of scientific method (e.g., Nola & Sankey, 2007). This rejection includes the view that a single criterion, such as falsifiability, cannot effectively demarcate science from nonscience.

Despite the idea's popularity, there has been a growing realization that the existence of one true account of scientific method is untenable. The majority view today is that there can be no fixed, universal account of scientific method appropriate at all times for all sciences. A quick inspection of different disciplines such as physics, biology, and economics reveals a diverse array of methodological practices. This holds for psychology as well, although a good deal of its research practice has a disquieting sameness about it. The coexistence of the four major theories of scientific method to be canvassed shortly, and the broad spectrum of methodological concerns shown by ATOM, attest to the existence of numerous different scientific methods. In short, the claim that science employs various accounts of scientific method should be accepted immediately.

An arresting criticism of scientific method was put forward by Popper, who was fond of declaring that scientific method does not exist. By this he meant that "(1) There is no method of discovering a scientific theory. (2) There is no method of ascertaining the truth of a scientific hypothesis, i.e., no method of verification. (3) There is no method of ascertaining whether a hypothesis is 'probable,' or 'probably true'" (Popper, 1983, 6). However, these claims are really part of Popper's reasons for rejecting an inductive conception of scientific method and adopting a falsificationist construal of the hypothetico-deductive method in its place. The claims do not address other accounts of scientific method. Thus the three assertions that Popper thinks speak against the idea of scientific method would likely be accepted by many who adopted alternative conceptions of scientific method. For example, advocates of a modern inductive conception of scientific method do not regard it as a strong discovery method; most scientists take scientific method to be concerned with the justification of knowledge claims, and not with directly ascertaining their truth; and although Bayesian methodologists reject the third claim, most scientists do not assign probabilities to hypotheses and theories. In short, Popper was not really against the idea of scientific method, only one limited conception of scientific method.

In a book provocatively titled *Against Method* (1975), Paul Feyerabend presented a different criticism of scientific method. He railed against the idea that there is or can be one fixed method for all time, arguing that no methodological rules exist that have not been broken at some time or other in the interests of genuine scientific progress. Thus, for Feyerabend, the only rule that does not inhibit progress is the meta-rule "Anything goes." His argument has been endorsed by a number of commentators who want to de-emphasize the importance of scientific method. However, Feyerabend's criticism speaks against the fixity of methodological rules only. Nothing in his writing counsels against the flexible use of a variety of different methodological rules that are revisable in the light of experience and reason.

The criticisms of scientific method just considered are easily turned, largely because they present scientific method in an unflattering light. None of them consider conceptions of scientific method that are informed by the contemporary literature on scientific methodology. Much of this literature accepts the tenability—indeed, the importance—of the idea of scientific method, although it is replete with criticisms of the various major accounts of scientific method.

A number of prominent scientists have also commented on scientific method in ways that devalue its very idea. Two well-known criticisms are those of the Nobel laureates Percy Bridgman and Richard Feynman. Bridgman, the father of the philosophy of operationism, forthrightly asserted that "the scientific method, as far as it is a method, is nothing more than doing one's damnedest with one's mind, no holds barred" (Bridgman, 1955, 535). This comment is cryptic in the extreme and does not differentiate scientific method from other types of method or from nonmethodic endeavors. Feynman also cryptically declares that scientific method "is based on the principle that observation is the judge of whether something is so or not. . . . Observation is the ultimate and final judge of the truth of an idea" (Feynman, 1998, 15). Feynman's comment, however, exaggerates the importance of observation in science and says nothing about the procedural dimension of scientific method. Although Bridgman and Feynman say different things about scientific method, neither of them characterizes it in an informative manner or acknowledges the sizable body of literature on scientific method that has accumulated since the time of Galileo (see, e.g., Gower, 1997). Although their pronouncements are often invoked by those who want to deflate the idea of scientific method, their remarks can hardly be taken as an informed guide to the topic. Despite their brevity, my remarks should suggest that these sorts of criticisms of the idea of scientific method carry little weight. Further, I think that the existence of major theories of scientific method attests to the notion that there is a great deal to the idea of scientific method.

1.3 Four Theories of Scientific Method

Modern scientific methodology has given considerable attention to a variety of different theories of scientific method. I now briefly review four of the most prominent theories: the inductive method, the hypothetico-deductive method, Bayesian hypothesis testing, and inference to the best explanation. Each has been endorsed by different methodologists as the best account of scientific method for scientists to adopt. However, I believe that none of them deserves a dominant position in the researcher's methodological armory. Rather, they should be thought of as local domain-specific methods.[3]

1.3.1 Inductive Method

The idea that scientific method involves inductive reasoning goes back at least to Aristotle and was given heavy emphasis, though in different ways, by Francis Bacon and John Stuart Mill. Inductive reasoning takes various forms. For example, it is found in the fashioning of statistical generalizations, in a form of reasoning by analogy, in the Bayesian assignment of probabilities to hypotheses, in the strategy of successively eliminating implausible hypotheses, and in the reasoning involved in moving from confirmed predictions to hypotheses in the standard formulation of the hypothetico-deductive method.

Historically speaking, the most popular inductive approach to scientific method is a simple form of inductivism (e.g., Chalmers, 2013). According to this account of method, science begins by securing observed facts, which are collected in a theory-free manner. These facts provide a firm base from which the scientist reasons "upward" to hypotheses, laws, or theories. The reasoning involved takes the form of enumerative induction and proceeds in accordance with some governing principle of inductive reasoning. As its name suggests, enumerative induction is a form of argument in which the premises count a number of observed cases from which a conclusion is drawn, typically in the form of an empirical generalization. However, enumerative induction can also take the form of a prediction about something in the future or a retrodiction about something in the past. The governing principle for an enumerative induction to a generalization can be stated informally as follows: "If a proportion of A's have been observed under appropriate conditions to possess property B, then infer the same proportion of all A's to have property B." This inductive principle can be taken to underwrite the establishment of statistical generalizations.

The simple account of inductive method has been criticized in various ways, although the criticisms are mostly directed at extreme versions of the method—versions claiming that observed facts can be known infallibly, that observations are made in an entirely theory-free manner, and that empirical generalizations can be secured through the use of a strongly justified principle of induction. However, this simple view of inductive method can be amended and defended in a moderate form as follows: observed facts can be established reliably, if fallibly; theory can be, because it has to be, used to guide observations; theoretical terms can be used to report observational statements without threatening the reliability of those statements; and principles of induction can be given an adequate justification on pragmatic grounds.

In psychology, the radical behaviorism of B. F. Skinner (1956, 1984) is a prominent example of a research tradition that uses an attractive nonstatistical inductive conception of scientific method. The major goals of radical behaviorist research are first to detect empirical generalizations about learning and then to systematize those empirical generalizations by assembling them into nonexplanatory theories. Murray Sidman's *Tactics of Scientific Research* (1960) is an instructive radical behaviorist account of the inductive methodology of this process. The Bayesian approach to hypothesis testing, which is slowly gaining some acceptance in psychology, can also be regarded as a sophisticated variant of inductive method.

1.3.2 Hypothetico-Deductive Method

Undoubtedly the most popular account of scientific method is the hypothetico-deductive method, which has been the method of choice in the natural sciences for more than 150 years (Laudan, 1981). This method has come to assume hegemonic status in the behavioral sciences, which have often placed a heavy emphasis on testing hypotheses in terms of their predictive success. In psychology, the pervasive use of traditional statistical significance test procedures is routinely embedded in a hypothetico-deductive structure.

The hypothetico-deductive method is characteristically described in one of two ways. According to the more popular account, the scientist takes a hypothesis or a theory and tests it indirectly by deriving from it one or more observational predictions, which are amenable to direct empirical testing. If the predictions are borne out by the data, then that result is taken as a confirming instance of the theory in question. If the predictions fail to square with the data, then that fact counts as a disconfirming instance of the theory. The other account comes from Karl Popper (1959). As noted earlier, he construes the hypothetico-deductive method in falsificationist terms. According to this rendition, hypotheses are viewed as bold conjectures, which the scientist submits to strong criticism with a view to overthrowing or refuting them. Hypotheses that successfully withstand such criticism are said to be corroborated, which is a noninductive notion of support.

Although the hypothetico-deductive method is used by many scientists and has been endorsed by prominent philosophers of science, it has received considerable criticism. Leaving aside Popper's less influential view, the major criticism of the hypothetico-deductive method is that it is confirmationally lax. This laxity arises from the fact that any positive confirming instance of a hypothesis obtained through its use can confirm any hypothesis that is conjoined with the test hypothesis, irrespective of the plausibility of that conjunct. Another criticism of the hypothetico-deductive method is that it standardly submits a single hypothesis to critical evaluation without regard for its performance in relation to possible competing hypotheses. Yet a further criticism of the method is that it mistakenly maintains that hypotheses and theories arise through free use of the imagination, and not by some rational, methodological, or logical means.[4]

Criticisms such as these have led some methodologists to recommend that the hypothetico-deductive method should be abandoned (e.g., Glymour, 1980; Rozeboom, 1997). Although this recommendation might be reasonable when applied to the method as it is standardly conceived, it is possible to correct its deficiencies and use the method to good effect in hypothesis testing research (e.g., Sprenger, 2011). For example, one might overcome the confirmational defects of the orthodox hypothetico-deductive method by employing a Bayesian approach to confirmation within a hypothetico-deductive framework. Further, with or without a commitment to the Bayesian approach, one could use the hypothetico-deductive method to deliberately test two or more competing hypotheses in relation to the evidence, rather than a single hypothesis in relation to the evidence. Further still, in testing two or more hypotheses, one might supplement the appeal to empirical adequacy by invoking criteria to do with explanatory goodness. This last correction might be considered to transform standard hypothetico-deductive method into a form of the method of inference to the best explanation, an idea that I will take up in chapter 5. Finally, typical formulations of the hypothetico-deductive method depict the empirical evidence as data, not phenomena. The contrast between data and phenomena will be laid out in the next chapter. Suffice it to say that specifying the evidence condition in terms of phenomena, rather than weaker data patterns, would provide the method with stronger empirical tests.

1.3.3 Bayesian Method

Although the Bayesian approach to evaluating scientific hypotheses and theories is looked on more favorably in philosophy of science than the hypothetico-deductive alternative, it remains a minority practice in psychology and the other behavioral sciences. However, it should be said that some methodologists in the behavioral sciences are now applying Bayesian ideas to a variety of methodological topics and problems.

With the Bayesian approach, probabilities are considered central to scientific hypothesis and theory choice (e.g., Howson & Urbach, 2006). In science, Bayesian hypothesis testing is a statistical affair, a practice that has been augmented by the allied philosophy of science known as *Bayesianism* (e.g., Earman, 1992). In using probability theory to characterize theory evaluation, Bayesians recommend the assignment of posterior probabilities to scientific hypotheses and theories in the light of relevant evidence. Bayesian hypothesis choice involves selecting from competing hypotheses the one with the highest posterior probability, given the evidence. The vehicle through which this process is conducted is Bayes's theorem. This theorem can be written in its simplest form as $Pr(H/D) = Pr(H) \times Pr(D/H) \div Pr(D)$. The theorem says that the posterior probability of the hypothesis is obtained by multiplying the prior probability of the hypothesis by the probability of the data, given the hypothesis (the likelihood), and dividing the product by the prior probability of the data.

Although Bayes's theorem is not controversial as a mathematical theorem, it is controversial as a guide to scientific inference. With respect to theory appraisal, one frequently mentioned problem for Bayesians is that the probabilistic information required for their calculations on many scientific hypotheses and theories cannot be obtained. It is difficult to know how one would obtain credible estimates of the prior probabilities of the various hypotheses and evidence statements that made up Charles Darwin's evolutionary theory, for instance, or a modern formulation of psychodynamic theory. Not only are the required probabilistic estimates for such theories hard to come by, they do not seem to be particularly relevant when appraising such explanatory theories.

The problem for Bayesianism presented by scientific theory evaluation is that scientists naturally appeal to qualitative theoretical criteria rather than probabilities. I note in the next section that scientific theories are often evaluated by employing explanatory reasoning rather than probabilistic reasoning.

1.3.4 Inference to the Best Explanation

Inference to the best explanation is founded on the belief that a good deal of what we know about the world is based on considerations of explanatory worth. This form of inference occurs informally in everyday life and professional affairs, and more systematically in science. Because a primary function of many theories in science is to explain, inference to the best explanation evaluates theories in terms of their explanatory merits. Theories that offer good explanations are deemed more likely to be correct than those that offer poor explanations.

Inference to the best explanation is quite different from the three preceding accounts of scientific method. Unlike inductive method, which generalizes in a descriptive manner, inference to the best explanation embodies a theoretical form of inference about explanations of facts that appeal to entities or processes that are different from those facts. In contrast to the hypothetico-deductive method, inference to the best explanation takes the relation between theory and evidence to be one of explanation, not logical entailment; and in contrast to the Bayesian approach, it takes theory evaluation to be a qualitative exercise that focuses explicitly on explanatory criteria, not a quantitative undertaking in which one assigns probabilities to theories.

A major attraction of inference to the best explanation is that it explicitly assesses explanatory theories in terms of the important scientific goal of explanatory power. However, a major challenge for proponents of inference to the best explanation has been to furnish an informative account of the criteria that should be used to determine explanatory power. The cognitive scientist Paul Thagard (1978) presented a historically informed, systematic account of three major criteria that have successfully been used in assessing the worth of scientific explanations: explanatory breadth, simplicity, and analogy. These criteria were subsequently incorporated into a fully fledged method of inference to the best explanation known as the *theory of explanatory coherence* (Thagard, 1992). The theory figures prominently in chapter 5, which canvasses the prospects of using inference to the best explanation as a worthwhile approach to appraising psychological theories.

Although a focus on theories that embrace unobserved theoretical entities is not an essential feature of inference to the best explanation, scientists are justified in believing in such entities because their existence is proposed by scientific theories that provide the best available explanation of a wide range of phenomena. For example, the existence of electrons and viruses was widely accepted because of the explanatory goodness of theories that posited them. It seems that psychologists have also sometimes tacitly accepted the existence of human abilities and personality traits essentially for the same reason.

Advocates of inference to the best explanation do not hold that a theory covering a wide range of empirical phenomena gives a better explanation than its rival because it is true. However, many of the proponents of inference to the best explanation do seem to accept the idea that a theory is more likely to be true because it provides a better explanation of the relevant phenomena than its rival does. In fact, some go so far as to claim that inference to the best explanation provides a reasonable guide to the truth, or at least the approximate truth, of theories. The extent to which methods are truth conducive is a challenging topic, and there is no settled opinion about how method and truth relate in this regard. I have something to say about this relationship in chapter 5, though I make no attempt to solve the problem.

The four theories just considered are commonly regarded by philosophers of science as the major theories of scientific method. Although each of the theories has sometimes been proposed as the principal claimant for the title of *the* scientific method, they are all better thought of as restrictive accounts of method that can be used to meet specific research goals, not broad accounts of method that capture what is essential for all scientific inquiry. Each of these methods covers only a part of the methodological activity of science. To take any one of them as *the* account of scientific method would be to unduly restrict the scope of scientific inquiry. Indeed, this would still be the case even if all four methods were somehow combined into one supermethod. In subsequent chapters, I will be at pains to suggest that inductive method is appropriate for phenomena detection, but not for theory construction. Similarly, I will insist that we should not regard inference to the best explanation as an all-purpose form of inference but instead think of it as a method particularly suited for evaluating the worth of competing explanatory theories. For its part, the hypothetico-deductive approach, appropriately modified, can productively be used to test for the empirical adequacy of local hypotheses, whereas the Bayesian approach can be used to assign probabilities to hypotheses for which we have the appropriate probabilistic information. As we will see, ATOM assigns no role to either hypothetico-deductive or Bayesian accounts of method.

1.4 The Nature of Methodology

The evolution and understanding of scientific methods are to be found in the domain of scientific methodology, a fact that makes this interdisciplinary sphere of learning one of major practical and educational importance. Yet we have few extended accounts of the nature of scientific methodology. Larry Laudan's (1996) normative naturalism is prominent in the philosophy of science, and more than forty years ago, Abraham Kaplan (1964) and Adriaan de Groot (1969) wrote book-length treatments of methodology for the behavioral sciences. Martin Hammersley (2011) recently provided a broad-ranging discussion of methodology for social scientists. None of these earlier works have had a palpable influence on psychologists' thinking about scientific method. In what follows, I sketch the broad contours of a modern conception of scientific realist methodology that is in broad agreement with Tom Nickles's (1987a, 1987b) insightful treatment of the topic. My formulation of ATOM is underwritten by this conception of methodology, along with a host of more specific methodological ideas.

1.4.1 The Tasks of Methodology

It is important to distinguish at the outset between method and methodology. The term *method* derives from a combination of the Greek words *meta*, meaning *following*, and *hodos*, meaning *the way*, to give *following the way*, suggesting the idea of order. Applied to science, method suggests the efficient, systematic ordering of inquiry. Scientific method, then, describes a sequence of actions that constitute a strategy to achieve one or more research goals that have to do with the construction and use of knowledge. Researchers sometimes use the term *methodology* as a learned synonym for *method* (and *technique*). However, the term is properly understood as denoting the general study of methods and is the domain that forms the basis for a genuine understanding of those methods. To repeat, methods themselves are purportedly useful means for helping us realize chosen ends, whereas methodology contains the resources for an informed understanding of our methods.

In its study of methods, methodology is at once descriptive, critical, and advisory (Nickles, 1987a; Reichenbach, 1938). It discharges these major tasks by describing relevant methods and explaining how they help researchers achieve their goals; it critically evaluates methods against their rivals; and it recommends what methods we should adopt to pursue our chosen goals. Thus a good methodology will offer researchers an informed description of methods, a judicious evaluation of them in relation to their rivals, and instructive advice on how to choose and use those methods. Methodology is important because the three major tasks it addresses are essential to the conduct of high-quality research.

Being a practical endeavor, methodology is concerned with the mutual adjustment of means and ends. As such, it judges whether methods are sufficiently effective for reaching chosen goals. But methodology is also critically aim-oriented and considers what research goals the research process should pursue. How, for example, are we to understand the related goals of truth, understanding, and control? If truth is taken as a major goal of science (as I believe it must be), and if truth is construed as correspondence with reality (as I think it should be) (see Haig & Borsboom, 2012), then philosophical semantics becomes a part of methodology. If understanding has an important psychological dimension, as it undoubtedly does, then psychology becomes a part of methodology. And if the exercise of control over science is regulated to an appreciable extent by institutions, then policy science enters into methodology. From a genuine concern with questions such as these, it follows that methodology must constantly attend to possibilities of fashioning and deploying methods in the face of varied and changing goal demands. In doing so, it becomes the management science of research (Simon, 1969; Nickles, 1997).

In reconciling the means and ends of inquiry in these ways, it is evident that methodology should not be identified with single disciplines such as applied statistics or philosophy of science, though these and other branches of learning are well positioned to make valuable contributions to methodology. Rather, as stated in the preface, methodology is a central part of the broad domain of cognitive theory and is therefore best understood as an interdisciplinary field.

1.4.2 Problem-Oriented Methodology

Although talk of research problems abounds in behavioral science inquiry, it largely serves a rhetorical purpose rather than doing useful methodological work. Behavioral scientists seldom attend to the nature of problems and their place in the research enterprise. The methodological treatment of research problems that does exist typically amounts to the recommendation that we cast our research hypotheses in the form of questions. However, such a suggestion has limited value, for it involves no attempt to understand problems by developing and using an informative theory of questions.[5] Demands that researchers formulate their research questions are frequently just requests for an operationalization of research hypotheses by empirically specifying the relevant independent and dependent variables (e.g., Johnston & Pennypacker, 2009). Relatedly, solutions to the original "problems" often involve answering the questions by conducting experimental tests of the research hypotheses. It is true that John Dewey's (1938) problem-solving account of inquiry has occasionally been taken as an appropriate model for behavioral science research (e.g., Kerlinger & Lee, 2000), but unfortunately Dewey's psychological construal of problems does not readily translate into a useful counterpart at the methodological level.

In good part, the neglect of research problems as a methodological idea has occurred because we have subscribed to theories of scientific method that do not systematically provide for the use of problems thinking. According to the standard account of inductive method outlined earlier, research begins with the scientist gathering and reporting observations in a theory-free manner. However, as Hempel (1966) noted, inquiry could never get under way in such a fashion. The first stage of gathering all the facts could never be completed, because they are enormous in number and variety. Collecting facts would be possible only if our methods could select those facts that are relevant to our purpose. However, Hempel maintained that one could not determine relevance by incorporating problems into a simple inductive model of inquiry. He believed that the idea of a problem is too vague to be an effective device for the selection of relevant facts, and nothing less than a hypothesis or theory is required to initiate and direct inquiry. Therefore Hempel rejected simple inductivism and opted for a hypothetico-deductive perspective on scientific method in which inquiry is viewed as a relation between a theory and its consequences. This line of thinking is consistent with the standard portrayal of the hypothetico-deductive method, which makes no serious appeal to problems.

However, a few philosophers of science have focused on the methodology of scientific problems and their importance for the conduct of science. Although endorsing a variant of the hypothetico-deductive method, Popper (1972) insisted that science *is* a problem-solving enterprise. However, he faltered with his account of problems by locating most of its resources in the theoretical background rather than in the immediate space of inquiry. In his well-known book *Progress and Its Problems* (1977), Larry Laudan presented a general theory of science as a thoroughgoing problem-solving endeavor. However, it is a conception of science that leaves no room for the idea that science also pursues truth. Finally, Nickles (1981) developed an instructive theory of problems as a general approach to scientific methodology. I believe that this theory, which views problems as sets of constraints on their own solutions, is the most methodologically resourceful account of problems available today. I use it in completing my articulation of ATOM at the beginning of chapter 6.

1.4.3 Generative and Consequentialist Methodology

I now identify and briefly discuss two important methodological ideas that have received limited attention in the literature. These ideas are presented in two contrasts: (a) generative and consequentialist methodology, and (b) reliabilist and coherentist justification.

Modern scientific methodology promotes two different research strategies that can lead to justified knowledge claims. These are known as *consequentialist* and *generative* strategies (Nickles, 1987b). Consequentialist strategies justify knowledge claims by focusing on their consequences. By contrast, generative strategies justify knowledge claims in terms of the processes that produce them. Although consequentialist strategies are used and promoted more widely than generative strategies in contemporary science, an adequate conception of research methodology requires both. Consequentialist reasoning receives a heavy emphasis in scientific research through use of the hypothetico-deductive method. Consequentialist methods reason from the knowledge claims in question to their testable consequences. As such, they confer a retrospective justification on the theories they seek to confirm.

In contrast to consequentialist strategies, generative strategies reason from warranted premises to an acceptance of the knowledge claims in question. The method of exploratory factor analysis, which is the focus of chapter 3, is a good example of a method of generative justification. It affords researchers generative justifications by helping them reason forward from statements about established correlational data patterns to the rudimentary explanatory theories that the method generates. Judgments of initial plausibility constitute the generative justifications provided by methods like exploratory factor analysis. Generative justifications are forward looking because they are concerned with heuristic appraisals of the prospective worth of theories. ATOM's account of theory generation is explicitly underwritten by a generative conception of methodology.

1.4.4 Reliabilist and Coherentist Justification

In addition to embracing both generative and consequentialist reasoning strategies, an adequate methodology will use two different theories of justification known as *reliabilism* and *coherentism*. Reliabilism asserts that a belief is justified to the extent that it is acquired by reliable processes or methods (e.g., Goldman, 1986). For example, belief in the accuracy of temperature readings by the appropriate use of a calibrated thermometer is justified by the reliable process of its production. By contrast with reliabilism, coherentism maintains that a belief is justified in virtue of its coherence with other accepted beliefs. One prominent version of coherentism, explanationism, asserts that coherence is determined by explanatory relations and that all justification aims at maximizing the explanatory coherence of belief systems (Lycan, 1988).

However, the claim that all justification is concerned with explanatory coherence is too extreme, as the existence of reliabilist justification makes clear. Although reliabilist and coherentist approaches to justification are distinct, they are complementary. The complementarity will be spelled out in my presentation of ATOM. This theory of method deems reliabilist justification appropriate for the discovery of empirical generalizations, whereas a particular form of coherentist justification is best employed in the appraisal of explanatory theories.

1.4.5 Methodology with a Knowing Subject

Underwriting the conception of methodology I am sketching here is the anti-Popperian view that epistemology must take "the knowing subject" seriously. Applied to methodology more specifically, this attitude leads to a rejection of the fanciful idea that the researcher is a "computationally omnipotent algorithmizer" in favor of a more realistic conception that accords with our actual epistemic makeup. Herbert Simon's (1977) view of the researcher as a "satisficer" is an influential part of this more realistic conception of ourselves as knowers. According to this view, our rationality is bounded by temporal, computational, memorial, and other constraints and thus proceeds in good part by using heuristic procedures.

William Wimsatt (1986) helpfully characterizes heuristic procedures as having at least the following four properties. First, the proper employment of heuristics does not ensure that a solution will be found, much less that a solution will be the correct one. Second, heuristics are cost-effective procedures in that they make considerably fewer demands on time, effort, and computational complexity than their algorithmic counterparts. Third, the errors that result from using heuristic procedures are biased in systematic ways, so that we can often predict the conditions under which they will fail, and make appropriate adjustments. Fourth, applying heuristics to a problem may produce a transformation of the problem into one of related and more useful form.[6] The notion of heuristic procedures is central to the liberalized conception of methodology being glossed here and encourages us to treat the domain of pragmatic reasoning as a crucially important part of the research endeavor.

I should point out that this overview of the nature of methodology is incomplete in a number of respects: it ignores the social dimension of research, including institutional and economic considerations, and it does not dwell on the fact that research is often a nonlinear, bootstrapping, multipass enterprise (see, e.g., Nickles, 1987a). Despite these omissions, I do express a running concern in this book with the institutional matter of the need to reform psychology's research methods curriculum.

I now consider the philosophy of scientific realism, which, despite vigorous debate, may fairly be taken as the majority position in the philosophy of science.[7] Most scientists seem to be scientific realists of one sort or another, though they subscribe to the philosophy in tacit fashion. The conception of methodology just sketched should be accepted as part of the realist philosophy that I now outline. Indeed, given the centrality of method to science, and a commitment to a method-centered conception of epistemology, methodological realism is a core commitment of the philosophy of scientific realism.

1.5 Scientific Realism

1.5.1 Varieties of Realism

Scientific realism (hereafter simply *realism*), like many "isms," comes in a variety of forms. Among the many contemporary versions of realism, we find Cliff Hooker's naturalistic realism, Mario Bunge's hylorealism, Roy Bhaskar's critical realism, Ilkka Niiniluoto's quite different form of critical realism, Richard Boyd's abductive realism, Ian Hacking's entity realism, John Worrall's structural realism, Ron Giere's perspectival realism, J. D. Trout's measured realism, and Anjan Chakravartty's semi-realism, to mention just some of the prominent available alternatives. Realism, then, cannot be given a straightforward characterization, and it will always be possible to take issue with one or other of its formulations. For example, the tension between formulating realist theses in global terms and local terms runs through the realist literature. Although global accounts of realism have dominated historically, realists are starting to see local realism as an attractive way to formulate their philosophy.

In this book, I adopt a realist perspective on science. Although the link between realism and method is not direct, what I have to say about method is better understood against a backdrop of realism than, say, antirealist options such as empiricism and strong forms of social constructivism. To repeat, although the subject of considerable debate, and opposed by many antirealists, realism is the dominant philosophy of science today. This fact, combined with an increasing willingness to focus on the nature of scientific practice, makes realism an appropriate philosophy for science.

Most versions of realism display a commitment to at least two doctrines. First, there is a real world of which we are part, and second, both the observable and unobservable features of that world can be known by the appropriate use of scientific methods. Some versions of realism incorporate additional theses (e.g., the claims that truth is the primary aim of science, and that successive theories more closely approximate the truth), and some also nominate optional doctrines that may, but need not, be used by realists (e.g., the claim that causal relations are relations of natural necessity; see, e.g., Hooker, 1987). Others who opt for an "industrial-strength" version of realism for the physical sciences (e.g., Boyd, 1984) are more cautious about its successful reach in the behavioral sciences. For example, Trout (1998) subscribes to a more modest brand of realism in psychology, owing to his skepticism about the discipline's ability to produce deeply informative theories like those in the physical sciences.

1.5.2 Naturalistic Realism

One particularly important feature of the realism that I subscribe to is its thoroughgoing commitment to naturalism. For this reason, it might be called *naturalistic realism*. A perspicacious form of this philosophy is offered by Hooker (1987). According to this brand of realism, scientific reasoning, including theorizing, is a natural phenomenon that takes its place in the world along with other natural phenomena. Further, philosophy and science make up a mutually interacting and interconnected whole. As a philosophical theory about science, naturalistic realism has no privileged status and may be revised in the light of scientific knowledge. Similarly, the naturalistic realist foresees that philosophical conclusions, tempered by scientific knowledge, may force changes in science itself.

According to one influential view of naturalism, philosophy and science are interdependent. This interdependence takes the form of mutual containment (Quine, 1969), though the containment is different for each. Philosophy is contained by science, being located within science as an abstract critical endeavor that is informed by science. Science is contained by philosophy because philosophy, among other things, provides a normative framework for the guidance of science.

Naturalistic realism maintains that philosophy of science is the part of science concerned with the in-depth critical examination of science with respect to its aims, methods, theories, and institutions. Philosophy of science naturalized is, in a sense, science applied to itself. It employs the methods of science to study science. It is, where appropriate, constrained by the findings of science. And it is itself a general theory of science. As such, naturalized philosophy of science is at once descriptive, explanatory, advisory, integrative, and reflective of science. Being positioned within science, naturalistic philosophy is well placed to study science, learn from science, and help instruct science.

Not all naturalists are scientific realists, and not all scientific realists are naturalists, thus raising the question: why is it advantageous to combine scientific realism and naturalism in a single philosophy? One reason is that naturalism is the best methodology we have available to us. It gives us our best methods from which to choose and encourages us to constrain our theorizing in light of reliable scientific knowledge. Another reason is that naturalism's principled commitments to both anti-anthropocentrism and fallibilism enable us to offer a tenable defense of realism, one that is true to our makeup as cognizers. Finally, by embracing naturalism, realism becomes an integrated whole that affords us the best current explanatory theory of the cognitive dynamics of science (Hooker, 1987).

1.5.3 Local Realism

As noted earlier, most formulations of realism are global in nature (e.g., Boyd, 1989; Kitcher, 1993; Psillos, 1999). They are presented as overarching general philosophies of science that are intended to apply to all sciences at all times. Largely focusing on the achievement of physics, these formulations of realism are intended to apply to mature science that is in a state of advanced theoretical development. An important consequence of this focus is that global realism has limited value as a philosophy for the behavioral and social sciences, which have generally been less successful in their theoretical achievements.

The prominent theoretical psychologist Paul Meehl (1993) correctly argued that the philosophy of science can genuinely help to improve the quality of scientific thinking in psychology. However, he suggested that the received view in philosophy of science, which he takes to be a modified form of logical empiricism, is the appropriate philosophy for psychology. I think that this suggestion is mistaken on two counts. First, despite its achievements, logical empiricism is largely an outdated philosophy of science, even in an amended form. Second, as a global philosophy of science fashioned in an image of physics, it speaks poorly to the concerns of psychological science and therefore has limited value as a philosophy for psychology. A worthwhile realism must be realistic about the sciences to which it speaks.

To take advantage of the understanding of science that realism is capable of providing, the behavioral sciences need local, fine-grained formulations of realism that are appropriate to their particular natures and achievements (e.g., Kincaid, 2000). One sensible way to proceed would be to replace the core theses of global realism with revised theses along the lines suggested by Uskali Mäki (2005). What follows is a brief treatment of five core realist theses, three of which are influenced by Mäki's formulations.

Possible Existence By focusing on the mature sciences, standard formulations of realism insist that the entities postulated by successful theories in the mature sciences do in fact exist, and that they are pretty much like the theories say they are. Thus our best theories in physics entitle us to believe that entities such as atoms, electrons, and quarks are part of the world's furniture, and they have the properties described by the relevant best theories. However, this formulation ignores two important facts of epistemic life: all sciences exhibit uneven rates of theoretical progress, and different degrees of epistemic confidence should attach to the different phases of the development and appraisal of scientific hypotheses and theories. For example, when a scientist postulates a new entity, it is often appropriate to think that it *might* exist, not that it *does* exist. Considerable progress is required before one can express confidence in a new entity's existence. Commitment to a thoroughgoing fallibilism, combined with a self-critical approach to scientific practice, suggests that this is the appropriate epistemic attitude to adopt (Mäki, 2005). It follows that we should be wary of strongly tying our ontological commitments to the latest and "best" theory (Burian & Trout, 1995). Many entities once thought to exist turned out not to exist. Other entities were shown to exist, but to be wrongly described in earlier attempts to understand them. Still others are characterized by competing theories, resulting in a high degree of uncertainty about them. Ontological progress in science is mostly piecemeal and characteristically occurs in fits and starts. To be a realist, it is enough that we hold to the view that an entity might exist, and that we give ourselves every chance of showing that it does exist. This will often require concerted work spanning several generations.

These comments about possible existence clearly apply to psychology, though it will sometimes be more difficult to gauge ontological progress there than in the natural sciences, given the special challenges psychology can face in accessing its hidden causal mechanisms. For example, we cannot say with full assurance that the credentials of the Spearman-Jensen theory of intelligence entitle us to think that general intelligence

(g) exists for sure. The reason for this is that both this theory and many others in psychology are not sufficiently well developed and justified to warrant drawing such a conclusion. Although Spearman and Jensen's theory is a respectable theory, it competes with similarly credentialed theories of intelligence, none of which is widely accepted as the best theory.

Mind Dependence Standard realism also subscribes to the ontological thesis that scientific entities exist apart from our mental representations of them. Although this commitment is appropriate for the physical sciences, whose subject matters exist whether or not they are investigated, it is inappropriate for the large tracts of nonneuroscientific behavioral and social science, for there is an important sense in which mental and social objects such as beliefs, desires, attitudes, marriage, money, and universities are not mind independent. Rather, they are mind dependent in that they are partly constituted by our conceptions or representations of them. Money is a familiar example of an ontologically subjective entity. Something is money only because we regard it as money (Searle, 1995). If humans did not exist in a modern economy, then there would be no such thing as money. More generally, if there were no minds, there would be no mental and social entities.

There is no good reason why a realist philosophy should insist on mind independence. In fact, as Mäki (2005) has noted, a realist philosophy adequate to the social and behavioral sciences can provide for mind dependence by thinking of mental and social objects as *science* or *inquiry* independent. Theories of mental and social objects typically do not have the power to create those objects. This is sufficient to satisfy the demand that mental and social objects be studied objectively.

Possible Truth The foregoing remarks about existence apply in analogous fashion to truth. Orthodox realism says that our best theories in the mature sciences are literally true, or approximately true, and the appropriate use of reliable methods enables us to say that this is the case. However, rather than take our best theories to be true, or approximately so, a realism that is sensitive to the growth of scientific knowledge should accept the view that our theories might well be true in the future, if not right now. This will certainly be the case when our theories are first conceived. Therefore it is more realistic to nominate our theories as *candidates* for truth. Consistent with this, truth should be understood as an orienting ideal, which we approximate by fashioning and justifying our theories. Because we cannot expect immediate truth in science, truth should be understood as a distal goal, not a proximal goal.

A suitable notion of correspondence truth comports well with realism, for there will be facts of the matter that make our truth-nominated theories true or false, though we might not have strong grounds for determining their truth value. What matters is that our theories be given a decent opportunity to be judged true. As it is for existence, so it is for truth: considerable resources of time, money, and other types of institutional support are needed for inquiry to be undertaken successfully.

Observables and Unobservables Standard formulations of realism explicitly embrace unobserved theoretical entities. Specifically, it is claimed that such entities exist, and science's best theories successfully refer to such entities. However, Mäki (2005) thinks that the social sciences, including folk psychology, mostly study observed or manifest entities, which he calls *commonsensibles*. Commonsensibles are the familiar objects that we deal with on a daily basis, such as money, stock markets, beliefs and attitudes, and social institutions. For Maki, these sorts of entities are part of our familiar observed ontology. They are not newly postulated theoretical entities that we add to our ontology by hypothesizing their existence. Rather, our folk understanding of them is refined and validated through social and behavioral science inquiry.

I agree with Mäki that some of our commonsensibles are observables. However, I think that many of them have the status of unobserved theoretical entities. The folk psychological entities such as beliefs and desires, for example, are dispositions, inferred on the basis of their presumed effects under specified stimulus conditions. In a realist interpretation, these are appropriately thought of as theoretical entities (e.g., Rozeboom, 1973, 1984). However, it is important to adopt an attitude of letting the ontological chips fall where they may. Whether entities and processes are observable or unobservable will make a difference for how we investigate them, but it will make no difference to whether we should adopt a realist attitude toward them.

Aims Some formulations of scientific realism depict science as an aim-oriented endeavor. In this regard, it is commonly said that the fundamental aim of science is to discover the truth about the world. This core thesis of realism is sometimes spelled out by making a number of related claims. For example, Sankey (2008) insists that scientific progress must be thought of as an advance toward truth, where truth is understood as valuable truth, and valuable truth is understood as explanatory truth. He also maintains that the claim "science involves the pursuit of truth" is an epistemological claim because science is a knowledge-seeking enterprise.

However, science is a complex and varied endeavor and for this reason is better thought of as pursuing multiple aims. In addition to pursuing truth, science is also concerned with achieving understanding through the establishment of facts and theories, as well as the attainment of control—broadly understood to include, for example, the experimental regulation of inquiry, and the application of knowledge to bring about desirable social outcomes. As we have seen earlier, science can also fruitfully be regarded as a problem-solving endeavor, as Popper, Laudan, and Nickles have emphasized in different ways.

These brief and selective remarks about realism might seem like an unnecessary excursion. However, I want to signal that the methodological matters I deal with when articulating ATOM in the following chapters are best understood against a backdrop of realist philosophy of science. Further, I will have occasion to explicitly note links between realist philosophy and behavioral science methodology as I proceed. Although my primary purpose is to articulate and promote a broad understanding of psychological inquiry, I intend my remarks about realism and methodology to allow for the possibility of fashioning a local realist philosophy that is appropriate for psychology.

Now that I have assembled a number of background ideas to do with method, methodology, and realism, it remains for me to provide a sketch of ATOM that will form the principal focus of the book. This account of method is broad in scope, and the overview will give the reader an overarching structure by which to understand better the different research methods discussed the following chapters.

1.6 An Overview of the Abductive Theory of Method

According to ATOM, scientific inquiry proceeds as follows. Guided by evolving research problems that comprise packages of empirical, theoretical, and methodological constraints, scientists analyze sets of data to detect robust empirical regularities, or phenomena. Once detected, these phenomena are explained by abductively inferring the existence of underlying causes that are thought to give rise to them. Here abductive inference involves reasoning from claims about phenomena, understood as presumed effects, to their theoretical explanation in terms of underlying causes. Upon positive judgments of the initial plausibility of these explanatory theories, researchers attempt to elaborate on the nature of the causal mechanisms in question.[8] They do so by constructing plausible models of those mechanisms by analogy to relevant ideas in domains that are well understood. When the theories are well developed, they are assessed against their rivals with respect to their explanatory goodness. This assessment involves making judgments of the best of competing explanations.

An important feature of ATOM is its ability to serve as a framework within which a variety of more specific research methods can be located, conjoined, and used. Operating in this way, these otherwise separate, specific research methods can be viewed as submethods of the overarching abductive method. In turn, the submethods provide ATOM with the operational bite that helps it make scientific inquiry possible. Comprehensive methods are often constituted by a number of submethods and strategies that are ordered according to an overarching structure (Ross, 1981). By incorporating a good number of submethods within its fold, ATOM is therefore intensely compositional. And although the structure of the theory is stable, its specific composites can vary markedly, depending on their suitability to the investigation at hand.

In characterizing ATOM in the following chapters, I show in some detail how it deploys a number of specific research methods within its compass. Table 1.1 contains a variety of research methods and strategies that can be placed within the structure of ATOM. I discuss a number of these in the exposition of the method that follows, but most of them are not required for its characterization.[9] The majority of methods selected for consideration in the book have been chosen primarily to facilitate the exposition of the processes of phenomena detection and theory construction without attempting to give an essential characterization of these processes. Some of the details of ATOM would have to be modified as a function of the nature of the methods chosen to operate within it.

Both inductive and abductive forms of reasoning play major roles in ATOM. However, because of the prominence of abductive reasoning in the theory construction phases of the method, I refer to it as an *abductive theory*. The exposition of the theory begins with an account of phenomena detection and then considers the process of constructing explanatory theories.

Table 1.1
Phases, strategies, methods, and inferences in the abductive theory of method

		Theory construction		
Phases	Phenomena detection	Generation	Development	Appraisal
Strategies	Controlling for confounds Calibrating instruments Analyzing data Constructively replicating findings	Generating rudimentary, plausible explanatory theories	Developing theories through analogical modeling	Evaluating the explanatory worth of developed theories in relation to rival theories
Methods	Initial data analysis Exploratory data analysis Computer intensive resampling methods Meta-analysis	Exploratory factor analysis Grounded theory Heuristics	Evaluating positive and negative analogies from source models	Theory of explanatory coherence Structural equation modeling
Inferences	Enumerative induction	Existential abduction	Analogical abduction	Inference to the best explanation

Note: For the most part, the methods, strategies, and inferences contained in ATOM are appropriate either for phenomena detection or for theory construction, but not for both. Exceptions include exploratory factor analysis and grounded theory method, both of which have data analytic components that can also contribute to phenomena detection.

1.6.1 Phenomena Detection

ATOM places great importance on the task of detecting empirical phenomena, and it views the completion of this task as a prerequisite for subsequent theory construction. In understanding the process of phenomena detection, phenomena must be distinguished from data (Woodward, 1989). Phenomena are relatively stable, recurrent, general features of the world that researchers seek to explain. The Flynn effect of intergenerational gains in IQ (Flynn, 2009) is a prominent example of a phenomenon in psychology. Although phenomena commonly take the form of empirical regularities, it is more useful to characterize them in terms of their role in relation to observation and prediction. Phenomena give scientific explanations their point. They also, on account of their generality and stability, become the appropriate focus of scientific explanation. Data, by contrast, are ephemeral and pliable.

The methodological importance of data lies in the fact that they serve as evidence for the phenomena under investigation. In extracting phenomena from data, scientists often engage in data exploration and reduction using graphical and statistical methods. Generally speaking, these data analytic methods help directly in the detection of phenomena, but not in the explanation of explanatory theories.

To establish that data are reliable evidence for the existence of phenomena, scientists use a variety of strategies. They include controlling for confounding factors, carrying out replications, calibrating instruments, and engaging in data analytic strategies of both statistical and nonstatistical kinds.

In the next chapter, I outline a statistically oriented, multistage account of data analysis to further characterize the phenomena detection phase of ATOM. The model proceeds through the four stages of initial data analysis, exploratory data analysis, close replication, and constructive replication. These four phases are concerned respectively with data quality, pattern suggestion, pattern confirmation, and generalization. The overall process of phenomena detection is one of enumerative induction in which one learns empirically, on a case-by-case basis, the conditions of applicability of the empirical generalizations that represent the phenomena.

1.6.2 Theory Construction

According to ATOM, phenomena serve the important function of prompting the search for their understanding in the form of relevant explanatory theories. For ATOM, theory construction comprises three methodological phases: theory generation, theory development, and theory appraisal. The first two phases are temporal in nature; theory appraisal begins with theory generation, continues with theory development, and is undertaken in concerted fashion in the so-called phase of *theory appraisal*. ATOM characterizes each phase of theory construction as abductive in nature, though the character of abductive inference is different in each phase.

Abductive reasoning is a form of inference that takes us from descriptions of data patterns, or phenomena, to one or more plausible explanations of those phenomena (e.g., Josephson & Josephson, 1994). A brief characterization of abductive inference can be given as follows: some phenomena are detected that are surprising because they do not follow from any accepted hypothesis or theory; we notice that the phenomena would follow as a matter of course from the truth of a new hypothesis or theory (in conjunction with accepted auxiliary claims); we conclude that the new hypothesis or theory has initial plausibility and therefore deserves to be seriously entertained and further investigated.

In chapter 3, I discuss exploratory factor analysis as an example of a method in psychology that facilitates the abductive generation of theories about latent factors (Haig, 2005b). With this method, theories are generated through a process of existential abduction in which the existence, but not the natures, of the causal mechanisms is hypothesized. The claim for the existence of general intelligence is psychology's best-known example of a hypothesis about latent factors arrived at by such means.

ATOM is also a method for theories-in-the-making. It encourages researchers to regard their theories as developing entities. Because we often do not have knowledge of the nature of the causal mechanisms we abductively probe, such nascent theories stand in clear need of development. ATOM urges us to construct models of those mechanisms by imagining something analogous to mechanisms whose nature we do know. In this regard, ATOM adopts the strategy of analogical modeling to help develop explanatory theories (Abrantes, 1999). Because analogical modeling increases the content of explanatory theories, I refer to the reasoning it embodies as *analogical abduction*. With analogical modeling, one builds an analogical model of the unknown subject or causal mechanism based on the known nature and behavior of the source from which the model is drawn (Harré, 1976). The computational model of the mind, based on an analogy with the computer, is a clear example of a model that has been developed by using this strategy.

ATOM takes the systematic evaluation of mature theories to be an abductive undertaking known as *inference to the best explanation*, whereby a theory is accepted when it is judged to provide a better explanation of the evidence than its rivals. ATOM takes inference to the best explanation to be centrally concerned with establishing explanatory coherence (Thagard, 1992). The theory of explanatory coherence maintains that the propositions of a theory hold together because of their explanatory relations. Relations of explanatory coherence are established through the operation of seven principles: symmetry, explanation, analogy, data priority, contradiction, competition, and acceptability. The explanatory coherence of a theory is determined by means of three criteria: explanatory breadth, simplicity, and analogy. Each criterion is embedded in one or more of the principles. Explanatory breadth, which is the most important criterion for judging the best explanation, captures the idea that a theory is more explanatorily coherent than its rivals if it explains a greater range of facts or phenomena than its rivals. The notion of simplicity deemed most appropriate for theory choice is captured by the idea that preference should be given to theories that make fewer special assumptions than their rivals. Finally, explanations are judged more coherent if they are supported by analogy to theories that scientists already find credible. Darwin's theory of evolution by natural selection has been shown, through use of the theory of explanatory coherence, to be a more explanatorily coherent theory than its creationist alternative (Thagard, 1992). The theory of explanatory coherence offers the researcher an integrated account of the criteria deemed important for the appraisal of explanatory theories. The theory of explanatory coherence is implemented through a computer program that enables the researcher to make systematic decisions about the best of competing explanatory theories.

ATOM aspires to be a coherent theory that brings together a number of different research methods and strategies that are normally considered separately. Although ATOM is a broad theory of scientific method, it is not a fully comprehensive account. Rather, it is a singular account of scientific method that is appropriate for the detection of empirical phenomena and the subsequent construction of explanatory theories.

As stated in the preface, I present ATOM chapter by chapter as follows: Chapter 2 provides a wide-ranging account of phenomena detection. Chapter 3 discusses the abductive nature of theory generation by focusing on the method of exploratory factor analysis. Chapter 4 considers the process of theory development as it is carried out via the strategy of analogical modeling. In chapter 5, inference to the best explanation, in the form of the theory of explanatory coherence, is presented and recommended as a fruitful approach to theory appraisal. In the last chapter, the exposition of ATOM is completed by presenting a theory of research problems.

1.7 Conclusion

Despite its centrality to science, scientific method is given less respectful attention in psychology than it deserves. The discipline's modal research practice, its uncritical science education, and the narrow interests of its professional methodologists contribute collectively to an attitude of disinterest in the topic.

In this chapter, I reaffirmed the importance of method in science. Not only is scientific method a centerpiece of science, but it is also unscathed by the superficial criticisms offered by commentators who do not bother to evaluate the extant theories of scientific method. I briefly considered four major theories of scientific method and concluded that each is appropriately thought of not as the best global account of scientific method but as a local method with domain specific application; there is no such thing as *the* scientific method.

I then sketched a heterodox account of the philosophy of scientific realism as a foundational backdrop to the book's ongoing concern with scientific method. I suggested that this brand of realism is apt for behavioral science disciplines such as psychology, whose theoretical achievements are more modest than those of the physical sciences. The main features of a realist conception of scientific methodology were given particular attention. This liberalized view of methodology underwrites much of the material presented in the book.

Finally, I presented a preview of ATOM to provide an orienting structure for its more extended treatment in the book. By providing explicitly for both inductive and abductive reasoning within its fold, the abductive theory of method supports the idea that there are several "logics" to scientific discovery.

2 Detecting Psychological Phenomena

Phenomena! Now there's a word to conjure with. It is what our theories try to explain, and what we use to justify those theories. It is what instrumentalists try to save, and what realists try to get beyond. It is what Ian Hacking thinks we create in the laboratory (in contrast to nature) and what Kant took to be partly the work of the mind (in contrast to noumena).

—James Brown (1994, 117)

2.1 Introduction

Since the 1950s, much psychological research has employed a top-down research strategy in which a minimalist account of the hypothetico-deductive method, in tandem with null hypothesis testing, is used to test hypotheses and theories (Rorer, 1991; Rozeboom, 1997). This practice has several weaknesses, one of which is a narrow view of data analysis in which the core information yield is a binary accept–reject statistical decision about the hypotheses and theories under test. As a consequence of this focus on top-down hypothesis and theory testing, psychology has failed to sufficiently recognize an important complementary, bottom-up research strategy that pursues data-to-theory research (Haig, 2013). This bottom-up strategy is captured by ATOM and has two primary aspects: the detection of phenomena, mostly in the form of empirical generalizations, and the subsequent explanation of those phenomena through the abductive construction of theory.

This chapter focuses on the important process of detecting empirical phenomena with reference to psychology. Although psychologists look to detect phenomena, they do so without a full appreciation of its methodological nature—a problem that is sometimes partially obscured by reconstructing phenomena detection in a hypothesis-testing guise. I draw here from contemporary philosophy of science to provide a methodologically informative account of phenomena detection. First, I present the important threefold distinction between data, phenomena, and explanatory theory that was introduced by Bogen and Woodward more than twenty years ago (Bogen & Woodward, 1988). However, my primary concern is to distinguish between data and phenomena, and I mention explanatory theory only insofar as it helps to elucidate the nature of the data-phenomena relation.[1] I then discuss a number of methodological strategies that are used to identify empirical phenomena. I propose, as one of these strategies, a multistage model of data analysis, which goes well beyond psychology's tendency to focus on traditional confirmatory data analysis. In the second part of the chapter, I consider aspects of the nature of science that are prompted by reflecting on the distinctions between data, phenomena, and explanatory theory. These include whether scientific facts are discovered or made, the distinction between empirical and theoretical progress in science, and the type of knowledge justification appropriate to phenomena detection. Taken together, these considerations press for significant changes in the way we think about and practice psychological research. Before concluding the chapter, I consider some of these changes and make several recommendations that would help psychology correct a number of its current research deficiencies.

2.2 The Nature of Phenomena

As James Brown's epigraph at the beginning of the chapter makes clear, we have always understood the nature and role of phenomena in science in various ways. Historically, scientists insisted on "saving the phenomena" in the instrumentalist sense of rendering an adequate description of the phenomena studied. In contrast to this narrow empiricist view, most scientists today are realists in their outlook, first because they are concerned to discover and properly describe phenomena, but also because they endeavor to construct explanatory theories to understand the underlying causal factors that are thought to produce them.

Scientists and philosophers frequently speak as though science is principally concerned with establishing direct relationships between observation and theory. Empirical evidence indicates that psychologists speak, and sometimes think, in this way (Clark & Paivio, 1989). Similarly, philosophers of science of quite different persuasions often say that scientific theories are evaluated in respect of statements about relevant data (Bogen & Woodward, 1988). Despite what they say, scientists frequently behave in accord with the view that theories relate directly to claims about phenomena, and claims about phenomena relate directly to claims about data. That is, talk of a direct relationship between data and theory tends to be at variance with empirical research practice, which often works with a threefold distinction between data, phenomena, and explanatory theory.

Science assigns major importance to the task of detecting empirical phenomena, and it often views the completion of this task as a requirement for subsequent meaningful theory construction. The next section discusses the nature of phenomena detection in science, with some reference to psychology. I begin by considering the basic distinction between data and phenomena.

2.2.1 The Distinction between Data and Phenomena

In a series of articles, Bogen and Woodward (Bogen, 2010, 2011; Bogen & Woodward, 1988, 1992; Woodward, 1989, 2000, 2010, 2011) argued in considerable detail that it is phenomena, not data, that scientific theories typically seek to predict and explain. In turn, it is the proper role of data to provide the observational evidence for phenomena, not theories.[2] Unlike data, phenomena are relatively stable, recurrent, general features of the world that we seek to explain. Hacking (1991) succinctly characterized the most popular class of phenomena as "noteworthy discernible regularities," which are often described in lawlike generalizations. The more striking regularities are often called *effects*, and they are sometimes named after the person considered to be their principal discoverer (e.g., the Compton effect in physics, the Baldwin effect in biology, the Flynn effect in psychology).[3] The so-called *phenomenal laws* of physics are paradigm cases of claims about phenomena. By contrast, the *fundamental laws* of physics explain the phenomenal laws. For example, the electron theory of Lorentz is a fundamental law that explains Airy's phenomenological law of Faraday's electro-optical effect (Cartwright, 1983). Examples of the innumerable phenomena claims in psychology include the matching law (the law of effect), the Flynn effect of intergenerational gains in IQ scores, and recency effects in human memory.

Although phenomena commonly take the form of empirical regularities, they make up a varied ontological bag that includes objects, states, processes, events, and other features that are hard to classify. For example, in psychology, the detected phenomena are often effects, which are empirical generalizations, but they also include the capacities of organisms as objects of explanation (e.g., the capacity to learn a language or to the capacity empathize with people).[4]

Because of this variety, it is generally more appropriate to characterize phenomena in terms of their *role* in relation to explanation and prediction, rather than in terms of their natures (Bogen & Woodward, 1988). For example, the relevant empirical generalizations in cognitive psychology might be the objects of explanations in evolutionary psychology that appeal to mechanisms of adaptation. Those mechanisms might in turn serve as phenomena to be explained by appealing to the mechanisms of natural selection in evolutionary biology.

As just indicated, phenomena, not data, are often taken as the proper objects of scientific explanation.[5] The two features of phenomena that make this appropriate are their stability and their generality. Typically, phenomena have to endure across a time interval long enough to allow theorists to construct explanatory theories about those phenomena, say, from three to thirty years.[6] In addition, science requires its *explananda* (the objects of explanation) to have a degree of generality that makes their explanation both tractable and economically viable. It would be ludicrous for science to try to explain individual data points one by one, and even impractical to explain local data patterns one at a time. Similarly, it would be practically unworkable for science to focus its major attention on highly local events that have little or no generality.[7] For good reason, psychology, as a basic science, is interested in why people generally behave the way they do, not why a particular person behaves in a particular way (D'Andrade, 1986).

To understand the process of phenomena detection, we must distinguish phenomena from data. Unlike phenomena, data are idiosyncratic to particular investigative contexts. Because data result from the interaction of a large number of causal factors, they are not as stable and general as phenomena, which are produced by a relatively small number of causal factors.[8] Data, then, are ephemeral and pliable, whereas phenomena are robust and stubborn. Phenomena have a stability and repeatability that are demonstrated through the use of different procedures, which often engage different kinds of data. Data are recordings or reports that are perceptually accessible; they are observable and open to public inspection. Despite the popular view to the contrary, phenomena are not, in general, observable; they are abstractions wrought from the relevant data, often as a result of a reductive process of data analysis. Indeed, as Cartwright (1983) remarked in her discussion of phenomenal and theoretical laws in physics, "The distinction between theoretical and phenomenological has nothing to do with what is observable and what is unobservable. Instead, the terms separate laws which are fundamental and explanatory from those that merely describe" (2). An important related point is that although data serve as evidence for phenomena, their perceptual qualities in this regard are of secondary importance. As Bogen and Woodward (1992) put it, "The epistemic significance of perception has to do with its reliability, not with its distinctively phenomenal or subjective experiential character. . . . Nonperceptual techniques of measurement and detection are just as epistemically valuable as perceptual techniques as long as they are reliable" (611). Methodologically speaking, what matters in science, then, is not the phenomenal or experiential qualities of perception but whether perception is a reliable process. For this reason, obtaining measurements using physical recording devices is just as important as using human perceptual techniques in detecting phenomena.

Data themselves are of scientific interest and importance only because they serve as evidence for the phenomena under investigation. Examples of data that serve as evidence for the psychological effects mentioned earlier are rates of operant responding (evidence for the matching law), IQ score gains (evidence for the Flynn effect), and error rates in psychological experiments (evidence for recency effects in short-term memory). Later I present a well-known example of the data-phenomena distinction that illustrates a number of the points just made.

Bogen and Woodward (Bogen & Woodward, 1988; Woodward, 1989) note that one can further distinguish between data and phenomena by appreciating the different kinds of error that are appropriate to each. Data-related errors arise from inaccuracies in their perception, and recording inaccuracies in their transcription. They also include deliberate efforts to manufacture data, as in the case of fraud. Errors of this kind are often simple in nature but can have far-reaching consequences because they threaten to undermine the adequacy of data as appropriate sources of evidence for claims about phenomena. Errors to do with phenomena detection are more complex and varied, reflecting, as they do, the complexity and variety of phenomena detection procedures. For example, in psychology they might include inappropriately using analysis of covariance to control statistically for nuisance variables, suboptimally using meta-analysis as a basis for claiming that an empirical generalization exists, and mistakenly believing in the robustness of a phenomenon claim by misusing Campbell and Fiske's (1959) multitrait-multimethod matrix.

As scientists scrutinize cases such as these, their overriding concern is to fathom whether they have detected genuine phenomena rather than pseudophenomena.

Bogen and Woodward (1988) helpfully illustrate the importance of distinguishing between data and phenomena by critically discussing an example of the melting point of lead.[9] The relevant fact, or phenomenon, here is that lead melts at 327.46 degrees Centigrade. How is knowledge of this obtained? Obviously, scientists do not determine the melting point of lead by liquefying one lead sample and observing a single thermometer reading of the melting temperature. Instead they carry out a series of relevant measurements using a reliable measuring instrument, such as a properly calibrated thermometer. Assuming the sources of systematic error have been eliminated or controlled for (e.g., the lead sample has been expunged of all relevant impurities, and the thermometer measurement is taken in the appropriate way), the scatter of recorded observations from the repeated measuring operations will be taken to include the true value of the melting point of lead. Furthermore, the determination of that true value will depend on a number of additional assumptions about the existence and independence of small and contingent unknown errors, the nature of the distribution of measurement, and the appropriateness of the sample estimate of the true value with its associated standard error.

As Bogen and Woodward (1988) remark, the lead example points up two important differences between data and phenomena. The first is that data are observed, either by human perception or with the aid of instruments, but the phenomenon of the true melting point of lead is not observed. Rather, the phenomenon statement about the true melting point of lead is inferred from claims about the observed data on the basis of classical sampling theory and its associated assumptions. The second point is that even the best theory of the molecular structure of lead could not explain why the array of data points occurred, because it depends not just on the melting point of lead but also on factors such as the purity of the lead, the working of the thermometer, background knowledge about measurement theoretic assumptions (in this case, true score theory), and how the readings should be taken. For these reasons, it is the phenomenon, not the set of data, that gets explained by the relevant theory of molecular structure.

The Flynn effect, mentioned earlier, provides a good example in psychology of an empirical phenomenon and as such helps one appreciate the difference between data and phenomena. Named after its principal discoverer, James Flynn, this effect is the striking fact that IQ scores have increased steadily across generations throughout the world. More precisely, Flynn documented the fact that, on average, IQ gains of about three points per decade occurred in some twenty nations from regions such as Europe, Asia, North America, and Australasia. IQ scores, obtained by using measuring instruments such as the Wechsler Scales and the Raven's Progressive Matrices, are data. These data provide empirical evidence for the Flynn effect. This effect is the stable generalization about the IQ score gains, which is abstracted from the data in light of relevant methodological criteria and represented statistically in terms of means and standard deviations for individual nations. Initially the Flynn effect was a baffling phenomenon for which we now have a variety of theoretical explanations, a fact made possible by the difference between, and relative autonomy of, claims about phenomena and explanatory theory.

In the various sciences, it is common to talk about the activity of extracting a signal from a sea of noise. Woodward (1989) observed that this model of signal and noise often usefully describes the challenge facing scientists when they seek to discover phenomena.[10] In detecting phenomena, we extract a signal (the phenomenon) from a sea of noise (the data). The data embody a great deal of noise because they result from a host of unknown causal factors, many of them local and idiosyncratic. For this reason, when extracting phenomena from the data, we often engage in data exploration and reduction by using graphical and statistical methods to manage the sea of noise. We enlist a variety of procedures to extract phenomena from the noise that masks them. Getting these procedures to work properly is essentially a problem of tuning.

I turn now to the process of phenomena detection. In doing so, I present a number of different procedures that scientists use to detect phenomena.

2.3 Procedures for Phenomena Detection

In establishing that data are reliable evidence for the existence of phenomena, scientists employ a variety of methodological strategies (Franklin, 1990; Woodward, 1989). Some, but not all, of these strategies are regularly used in psychology. They include controlling for confounding factors (both experimentally and nonexperimentally), empirically investigating equipment (including the calibration of instruments), engaging in data analytic strategies of both statistical and nonstatistical kinds, and undertaking the constructive replication of study results. Whereas these procedures are variously used in phenomena detection, they are not, in general, used to construct explanatory theories.[11] The later discussion of the importance of reliability assessments in phenomena detection helps indicate why this is so.

2.3.1 Controlling for Extraneous Factors

One basic requirement of sound experimental design involves the control of extraneous factors, which might otherwise confound the results by producing data mistakenly thought to be produced by the relevant phenomenon. Such control can be achieved by physically isolating the relevant potential confounds. In physics and chemistry, experimenters have been extraordinarily successful in controlling for extraneous influences. The same is true of experimental psychology. In one class of experiment, the Skinner box is used as an experimental chamber that isolates a number of influences extraneous to the investigation of operant conditioning phenomena by incorporating features such as light tightness, sound attenuation, and automated functions that prevent the subject from coming into direct contact with the experimenter. Alternatively, randomization procedures can be used in experimental contexts on the assumption that the influence of nuisance variables will be distributed uniformly over the various treatments in the long run.

Extraneous influences can also be controlled for in a statistical manner in research contexts where neither physical control nor randomization is possible or appropriate. Consider, for example, the common strategy of checking for what statisticians call *nonspuriousness*, where a variable, X, is established as a direct cause of another variable, Y. Such a relationship is judged nonspurious when we have grounds for thinking that no third variable, Z, confounds the X-Y relationship. In this regard, we often use partial correlation procedures to establish that the third variable, Z, is not a common cause of X and Y or a cause intervening between X and Y.[12]

2.3.2 Triangulation

As already mentioned, one of the distinctive features of claims about phenomena is their robustness. Robustness is a methodological notion that has long been considered important in the various sciences (Levins, 1968; Wimsatt, 1981). Robustness carries the idea that there have to be multiple means for establishing the nature and existence of phenomena, an idea that is based on the strong conviction that we are entitled to infer the existence of a phenomenon that stands up under a variety of different tests. As Wimsatt (1994) remarked: "Robustness has the right kind of properties as a criterion for the real, and has features which naturally generate plausible results. Furthermore, it works reliably as a criterion in the face of real world complexities" (11).

The notion of robustness is essentially the same idea as triangulation, which is more familiar to psychologists. Campbell and Fiske's (1959) classic multitrait-multimethod matrix is an important triangulation procedure for investigating the robustness of psychological constructs. With this procedure, validation involves obtaining convergent results through the use of independent measuring procedures, and the notion of discriminant validity serves to check that the invariance across tests, methods, and traits is not a result of their insensitivity to the variables under study. In experimental psychology, the idea of robustness is more commonly called *converging operations*, in accordance with Garner, Hake, and Eriksen's (1956) pioneering work.

2.3.3 Calibration

Another strategy that provides a justification for the confidence in experimental results is calibration. Calibration is the metrological process of determining the evidential reliability of an instrument by comparing it with a trustworthy standard. More particularly, it involves using a substitute signal to standardize a measuring instrument (Franklin, 1997), an operation that is achieved by a variety of complex procedures. This complexity is well exemplified by Chang's (2004) extensive study of the history of thermometry, which lies behind the routine use of mercury and other thermometers to measure temperature. Calibration is important in science because instruments must be calibrated before they can be used in a dependable manner. Although routinely carried out in the physical sciences and widely used in experimental psychology, calibration has received little systematic attention in other areas of psychology.

Also, because instruments tend to go out of calibration, they may need to be recalibrated. Of course, even when properly calibrated instruments are used in measurement, some random error is to be expected. To test whether measured values obtained from an instrument represent chance fluctuations or signal a loss of calibration, we can use a test of statistical significance (Baird, 1992). The normal curve is widely used as a model of chance fluctuations or errors of measurement. In this context, errors can be thought to result from numerous small, independent disturbances, such as slight variations in the mechanical or electrical components of the measuring instrument.

It is a truism that progress in science depends in part on the quality of its measurements. In psychology, we have a clear need to take calibration more seriously. In discussing the importance of calibrating psychological measures, Sechrest, McKnight, and McKnight (1996) were surely right to conclude that "knowledge, understanding, and progress in the science of psychology would be furthered greatly by concerted efforts to calibrate psychological measures in a variety of ways that are now available and that are sadly neglected" (1071).

2.3.4 A Model of Data Analysis

Given the importance of the detailed examination of data in the process of phenomena detection, it is natural that statistical analyses of data figure prominently in that exercise. Researchers in psychology often analyze rich data sets, and they are increasingly being called on to analyze massive sets of data. Thus data reduction often becomes the core feature of data analysis. With this in mind, I briefly outline the broad contours of a statistically oriented, multistage account of data analysis, which provides another way to characterize the process of phenomena detection. The exposition draws from Haig (2005a). The model comprises the four sequenced stages of initial data analysis, exploratory data analysis, close replication, and constructive replication. However, it should be noted that although psychology makes heavy use of statistical methods in data analysis, qualitative data analytic methods can also be used in phenomena detection (Strauss, 1987).

Initial Data Analysis The initial examination of data (Chatfield, 1985) is the first informal scrutiny and description of data undertaken before exploratory data analysis proper begins.[13] It involves screening the data for quality. Initial data analysis variously involves checking the accuracy of data entries, identifying and dealing with missing and outlying data, and examining the data for their fit to the assumptions of the data analytic methods used. Data screening thus enables one to assess the suitability of the data for the type of analyses intended.

This important, time-consuming, preparatory phase of data analysis has not received the amount of explicit attention that it deserves in psychological research practice and education. However, the American Psychological Association's Task Force on Statistical Inference (Wilkinson & the Task Force on Statistical Inference, 1999) recommended changes to current practices in data analysis that are broadly in keeping with the goals of initial data analysis. It is now a straightforward matter to use a computer to produce the graphical displays and descriptive tabulations that are used in the initial examination of data. Fidell and Tabachnick (2003) provided a useful overview of the importance of the work required to identify and correct problems in data.

It should be clear, even from these brief remarks, that the initial examination of data is necessary for successful data analysis in science because data that lack integrity can easily result in the subsequent misuse of data analytic methods and drawing erroneous conclusions.

Exploratory Data Analysis The last thirty years have witnessed the strong development of an empirical, data-oriented approach to statistics. One important part of this movement is exploratory data analysis, which contrasts with the more familiar traditional statistical methods and their characteristic emphasis on the confirmation of knowledge claims. Like initial data analysis, this newer movement places a heavy emphasis on the close examination of data. However, its basic purpose is to identify provisional patterns in the data.

Exploratory data analysis uses multiple forms of description and display and often involves quantitative detective work designed to reveal the structure or patterns in the data under scrutiny (Behrens & Yu, 2003; Tukey, 1977).[14] The exploratory data analyst is encouraged to undertake an unfettered investigation of the data and perform multiple analyses using a variety of intuitively appealing and easily used techniques.

The compendium of methods for data exploration is designed to facilitate both the discovery and communication of information about data. These methods are concerned with the effective organization of data, the construction of graphical displays, and the examination of distributional assumptions and functional dependencies. The stem-and-leaf display and the box-and-whisker plot are two well-known exploratory methods.

Two attractive features of exploratory methods are their robustness to changes in underlying distributions and their resistance to outliers in data sets. Exploratory methods with these two features are particularly suited to data analysis in psychology, where researchers are often confronted with *ad hoc* data sets on manifest variables that have been acquired in convenient ways.

Close Replication Successfully conducted exploratory analyses will suggest potentially interesting data patterns. However, it will normally be necessary to check on the stability of the emergent data patterns by using appropriate confirmatory data analysis procedures.

Computer-intensive resampling methods such as the bootstrap, the jackknife, and cross-validation (Efron & Tibshirani, 1993) make up an important set of confirmatory procedures that are often well suited to this role. By exploiting the modern computer's massive computational power, methods such as these free researchers from the assumptions of orthodox statistical theory (such as the belief that the data are normally distributed) and permit them to gauge the reliability of chosen statistics by making thousands, even millions, of calculations on many data points. Researchers use resampling methods to establish the consistency or reliability of sample results. They are particularly suited to ascertaining the validity of the data patterns initially suggested by the use of exploratory methods. In doing this, they provide us with the kind of validating strategy that is needed to achieve close replications.

Constructive Replication In establishing the existence of phenomena, it is often necessary for science to undertake both close and constructive replications. The statistical resampling methods just mentioned are concerned with the consistency of sample results that help researchers achieve close replications. By contrast, constructive replications are undertaken to check the validity of the results obtained by close replication. This is achieved by doing two things. First, a concerted effort is made to faithfully reproduce the conditions of the original study, often by an independent investigator or research group. This is sometimes called *direct replication*. Strictly speaking, this is a form of constructive replication because although it attempts to literally replicate the first study, it involves a change in geographic time, location, and researchers. Second, research is undertaken to demonstrate the extent to which results hold across different methods, treatments, and occasions. This form of constructive replication, in which researchers vary the salient study conditions, is a triangulation strategy designed to ascertain the generalizability of the results identified by direct replication (Lindsay & Ehrenberg, 1993). Both forms of constructive replication are time-honored strategies for justifying claims about phenomena.

The four-stage model of data analysis just outlined assists in phenomena detection by attending in turn to the different but related tasks of data quality, pattern suggestion, pattern confirmation, and generalization. In effect, the outcome of this sequenced process is a form of enumerative induction in which one learns empirically, on a case-by-case basis, the conditions of applicability of the empirical generalizations that represent the phenomena.

2.3.5 Meta-analysis

It is important to appreciate that this model of data analysis is clearly not the only statistical means by which we can detect phenomena. In addition to the several strategies mentioned earlier, meta-analysis is a prominent example of a distinctive use of statistical methods by behavioral scientists to aid in phenomena detection. As is well known, meta-analysis is widely used to conduct quantitative literature reviews. It is an approach to data analysis that involves quantitative analysis of the data analyses of primary empirical studies. By calculating effect sizes across primary studies in a common domain, meta-analysis helps us detect general positive effects (Schmidt, 1992). By using statistical methods to ascertain the existence of robust empirical regularities, meta-analysis can usefully be viewed as a statistical approach to constructive replication. Although meta-analysis is thought by some to do explanatory work, and is used widely in evaluation research, it is in the descriptive-cum-generalizing role just mentioned that it performs its most important work in science today. Contrary to the claims made by some of its critics in psychology (e.g., Sohn, 1996), meta-analysis can be regarded as a legitimate and important means of detecting empirical phenomena in the behavioral sciences (Gage, 1996). I briefly refer to the achievements of meta-analysis when considering the matter of scientific progress in psychology later in this chapter.

2.4 Reasoning from Data to Phenomena

Given that data serve as evidence for phenomena, the question naturally arises: how do scientists reason from claims about data to claims about phenomena? The first thing to note is that the inference involved is ampliative, or content increasing; it is not nonampliative or deductive. That is, the claims about the existence or the nature of the phenomena go beyond all information contained in assertions about the relevant data. Second, the ampliative inference cannot be hypothetico-deductive in nature, for the hypothetico-deductive method itself says nothing about how a hypothesis (the phenomenon claim) is formulated. Third, the inference from data claims to phenomena claims is in some sense inductive. In the description of the second strategy of constructive replication provided earlier, I noted that the reasoning process is a type of enumerative induction in which the generalization (the conclusion of the inductive reasoning process) is established on a serial basis as successive replications are undertaken to establish the scope of the generalization. Finally, being inductive in the sense just noted, the inferences from claims about data to claims about phenomena are not essentially abductive or explanatory in nature. As it is understood here, inductive inference is descriptive inference in that it reaches conclusions about "more of the same kind." By contrast, abductive inference is an explanatory mode of inference that scientists use when they reason from phenomena claims to theory claims that purport to explain why the phenomena occur.

This brief characterization of the inductive reasoning involved in moving from data claims to phenomena claims proceeds more by contrast than by direct analysis. Therefore it leaves a great deal unsaid about the details of the reasoning involved. In this regard, it is important to appreciate that illuminating accounts of inductive reasoning, as they are employed in actual cases of scientific research, will have to be cast as *material* inductions, not as formal inductions. Norton (2003) has characterized material induction as local rather than global reasoning in which contingent matters of fact pertinent to the domain in question are included in the formulation. For example, a material inductive characterization of the discovery of the melting point of lead would have to include reference to relevant contingent facts such as those mentioned in the earlier discussion of this phenomenon. By contrast, Bayesian accounts of inductive inference, which center on the probability calculus, are essentially formal and universal and make little or no reference to the welter of case-dependent detail required of good material inductions. Such formal accounts of induction are incapable of properly illuminating data-to-phenomena inferences.

Even a worked-out account of the material conditions involved in the inductive character of reasoning from data to phenomena will leave a great deal unsaid. In fact, a blow-by-blow account of the process of phenomena detection would have to focus on the procedures that are used in the chosen approach. For example, a reflective researcher who used the four-stage model of data analysis presented in this chapter would make innumerable judgment calls at each stage that were based on all kinds of specific considerations. They would involve posing and answering questions such as "Should I use log10 transformations to normalize my seriously skewed dependent variables?" "Will a back-to-back stem-and-leaf display give me sufficient comparative descriptive information about the two data sets?" "Can I use the jackknife as an adequate replacement for the more flexible bootstrap procedure?" "Is this new method sufficiently independent of the original method to enable me to move beyond the published generalization without fear of pseudo-robustness?" Clearly the relevance of these sorts of questions, and the appropriateness of the answers to them, will depend on detailed local knowledge of both methodological and substantive sorts.

2.5 Phenomena Detection and the Nature of Psychological Science

Accepting the distinction between data, phenomena, and explanatory theory has important consequences for our understanding of science, including psychology. Here I briefly comment on the matters of whether scientific facts are discovered or made, the division between theoretical and empirical research, and the different types of knowledge justification appropriate for phenomena claims and explanatory theory.

2.5.1 Are Phenomena Discovered or Constructed?

As noted at the outset, the account of phenomena detection adopted in this chapter is consistent with a commitment to a realist outlook on science. Among other things, this outlook commits one to the view that phenomena are ontological existents of various kinds, including empirical regularities. They occur in nature and are the sorts of things that can be discovered through scientific research, and about which we can have genuine knowledge. Many of these phenomena in the physical and biological sciences, and areas of psychology such as psychophysics and neuropsychology, are part of the world's furniture that exists independently of human interests, theoretical commitments, and sociocultural factors. Other phenomena to be found in areas such as social and economic psychology do not exist independently of these social factors. A question to be addressed here, then, is whether these social factors allow one to retain a realist outlook on phenomena that are influenced by them.

A number of sociologists of science adopt a strong social constructionist outlook on science and tend to deny that phenomena are real in the realist sense just noted. Latour and Woolgar's (1979) well-known ethnographic study of life in a scientific laboratory is a good case in point. The authors of this study pressed their viewpoint by noting that the word *fact* comes from the Latin noun *factum*, which derives from the past participle of *facere*, meaning "to do or make." For them, facts or phenomena are made, not discovered. So they do exist. However, their reality "is the consequence of scientific work rather than its cause." Latour and Woolgar went further by claiming that "phenomena are thoroughly constituted by the material setting of the laboratory." They are not real regularities in nature waiting to be discovered; rather, they are made by us.

However, Latour and Woolgar's empirical attempt to document the social construction of scientific facts was confined to an examination of laboratory inscriptions such as photographs, graphs, and written papers. They showed little interest in scientists' own understanding of their laboratory behavior, and for that reason their research was poor ethnography. Moreover, Latour and Woolgar failed to consider the process by which claims about scientific facts are socially constructed. That is, they failed to document the transition from transitory data claims to stable phenomena claims, and they chose to ignore the testing, modification, acceptance, and sedimentation of claims about such phenomena (Weinert, 1992). As a result, Latour and Woolgar mistook the data from a particular experiment for phenomena, which are stable and repeatable events.

Latour and Woolgar's (1979) strong social constructionist claim that scientific facts are solely manufactured is implausible. However, Hacking (1983) adopted a more moderate and subtle social constructionist view of phenomena. Although a realist of sorts, Hacking maintained that phenomena are typically created, not discovered. He believed that few phenomena exist in nature waiting to be discovered. Mostly there is "just complexity" in nature, and we mostly isolate phenomena by devising appropriate experimental arrangements that will produce them in a reliable manner.

In reply to Hacking's contention that most phenomena are created by experiments, Bogen and Woodward (1988) acknowledged that this is the case for some phenomena, such as those in physics created in very high-energy particle accelerators, but they maintained that this is not true for phenomena in general. They believe that Hacking ascribes to phenomena features that more appropriately characterize data.

Most empirical studies in psychology are not strictly experimental, and those that are do not create new phenomena in the manner of those forged by very high-energy particle accelerators. Think, for example, of the Flynn effect, which was detected nonexperimentally, and the law of effect, which was discovered and refined experimentally but operates in nonexperimental contexts. However, as just noted, some categories in psychology (e.g., "undergraduate student," "family," and "money") are socially constructed; they are social kinds rather than natural kinds (Hacking, 1999; Searle, 1995). However, the members of such socially constructed categories do exist and are therefore amenable to objective study. As discussed in chapter 1, the areas of psychology that study social kinds require a shift from realist orthodoxy with its commitment to the view that the objects of study are mind independent. We need to accept the fact that some subject matters are mind dependent (and partly representation dependent), in the sense that they would not exist without human minds. But given their social existence, they are inquiry independent and can therefore be studied objectively in a realist manner (Mäki, 2005).

Although a strong social constructionist perspective on detecting phenomena is unconvincing, the naive realist view that such facts are simply looked for and discovered is also unconvincing. This chapter adopts a moderate, flexible, realist position about scientific phenomena that acknowledges the role of social processes in the production of some phenomena while insisting that claims about phenomena are also significantly constrained by the world itself. Nor should it be thought that scientific phenomena exist only in the sense that they are created experimentally in the laboratory. Scientific phenomena exist in the world, typically masked by noise. We exhibit them in more or less pure form by forging them through both experimental and nonexperimental intervention. In speaking of phenomena detection in realist terms, I have in mind neither the observation nor the literal discovery nor the construction of inquiry-dependent facts.

2.5.2 Two Kinds of Scientific Progress

Psychologists have offered a number of different broad characterizations of their discipline's scientific progress. For example, they have appealed to Popper's (1959) notion of falsifiability in urging stronger tests of its theories (more honored in the breach than the observance), they have also followed Kuhn (2012) in judging its multiparadigmatic nature as a sign of disciplinary immaturity, and they have used Lakatos's (1970) methodology to evaluate the progressiveness of its research programs. However, with these portraits of scientific progress, they have focused more on the worth of psychology's theories and less on the nature of its empirical advances and the strength of its empirical claims.

Recognition of the fundamental importance of the distinction between phenomena claims and explanatory theories suggests the need to clearly differentiate between empirical progress and theoretical progress in science (Kaiser, 1996). The related aims of detecting empirical phenomena and constructing explanatory theories provide science with the two most fundamental goals in respect of which these different senses of scientific progress can be measured. That is, a discipline's empirical progress can be measured in terms of its success in detecting robust empirical phenomena, whereas its theoretical progress can be understood in terms of the goodness of its explanatory theories.

The question to be asked here is: has psychology made good progress in its quest to detect empirical phenomena? Some psychologists doubt that this is so. For example, Gergen (1973) maintained that the behavioral sciences deal with facts that are often nonrepeatable, and at best they produce generalizations that hold for a limited time only because they are invalidated by cultural and historical factors. Furthermore, he distrusted meta-analysis as a basis for claiming that empirical generalizations exist (Gergen, 1994). Relatedly, Cronbach (1975) believed that the interactive complexity of psychology's subject matter ensures that its generalizations have a short half-life. Furthermore, Lykken (1991) argued that psychology has made poor empirical and theoretical progress and, with respect to the empirical, contended that many of its findings fail to replicate.

In the face of negative assessments such as these, Gage (1996) countered that the results of meta-analysis include an array of stable and robust first-order and interaction effects that support the conclusion that the behavioral sciences have detected numerous empirical phenomena worthy of theoretical explanation. Furthermore, Hedges (1987) provided an example of one type of study that is needed to make informed judgments about empirical progress in psychology. He showed that a comparison of the empirical consistency of the results of replicated exemplary experiments in physics and psychology, which use the same numerical methods, reveals a similar degree of empirical cumulation. This is a piece of knowledge about empirical progress in psychology that challenges popular opinion.

Clearly we need more empirical work to further our knowledge of just how effective psychology has been in detecting empirical phenomena. Regarding the process of phenomena detection, the following catalog of questions deserves to be considered in future studies: What forms do the phenomena take (e.g., are they characterized as empirical generalizations or as capacities)? What is the means by which the phenomena have been detected (e.g., by constructive replication through use of experiments, or by the use of well-conducted meta-analytic studies)? Have the phenomena been detected by reliable means (e.g., are the measuring instruments properly calibrated, and do they retain their calibration)? To what extent are the phenomena generalizable, and does the scope of the generalizations change over time? What is the strength of the evidential support for the phenomena claims, and are they qualitative or quantitative in nature?

Good science strives to make both empirical and theoretical progress. Given that these are different types of progress and that psychology is a discipline of many parts, we should be sensitive to the likelihood that it has made uneven rates of progress of both sorts in its different areas. Overall, I think that psychology has done better at phenomena detection than theory construction. However, to substantiate this claim would require many detailed assessments of the quality of psychology's efforts and achievements with respect to these two fundamentally different processes.

2.5.3 Phenomena Detection and Reliabilism

Modern scientific methodology distinguishes between two important and different theories for justifying knowledge claims: reliabilism and coherentism. Reliabilism maintains that a belief is justified to the extent that it is acquired by reliable processes or methods (Goldman, 1986). The examples are numerous and varied. They include the use of calibrated thermometers to measure temperature, as in the case of determining the melting point of lead discussed earlier. Furthermore, under appropriate conditions, beliefs produced by perception, verbal reports of mental processes, and even sound argumentation can all be justified by the reliable processes of their production. The crucial point to make here is that reliability judgments are the appropriate type of justification for claims about empirical phenomena. As noted earlier, statistical resampling methods, such as the bootstrap, and the two strategies of constructive replication, provide different sorts of consistency tests that researchers can use to establish phenomenon claims by showing that data provide reliable evidence for the existence of phenomena. The use of consistency tests to validate knowledge claims on reliabilist grounds is widespread in science. It should be understood that this use of reliability as a mode of justification differs from the normal psychometric practice in which reliability and validity are presented as contrasts.

By contrast with reliabilism, coherentism asserts that a belief is justified in virtue of its coherence with other beliefs. One prominent version of coherentism, explanationism, maintains that coherence is determined by explanatory relations, and all justification aims at maximizing the explanatory coherence of belief systems (Lycan, 1988). However, the claim that all justification is concerned with explanatory coherence is too extreme, as the prominence of reliabilist justification in science makes clear. Rather, the role for explanatory coherence is to provide a justification for the acceptance of explanatory theories. For example, Thagard's (1992) theory of explanatory coherence articulates a method that enables researchers to decide whether one theory is superior to its rivals on the basis of criteria to do with explanatory breadth, simplicity, and analogy. To repeat, phenomena detection and explanatory theory construction are two fundamentally different research processes for which different approaches to knowledge justification are appropriate.

However, note that although reliabilism and explanationism are different and are often presented as rivals, they do not have to be seen as competing theories of justification. One can adopt a broadly coherentist perspective on justification that accommodates both reliabilism and explanationism and allows for their coexistence, complementarity, and interaction. This can be achieved by encouraging researchers first to seek and accept knowledge claims about empirical phenomena based solely on reliabilist grounds and then proceed to construct theories that will explain coherently those claims about phenomena. This is exactly what Flynn did, by first systematically documenting the effect that bears his name, and then endeavoring, with Dickens, to explain the effect by proposing their theory of environmental richness (Dickens & Flynn, 2001; see also Flynn, 2009).

One might add that the acceptability of the claims about phenomena will increase when they coherently enter into the explanatory relations that contain them. Alternatively, the explanatory breadth, and therefore the explanatory coherence, of a theory will decrease as a consequence of rejecting a claim about a relevant phenomenon that was initially accepted on insufficient reliabilist grounds.

I now turn to consider a number of important implications for psychological research that are occasioned by the account of phenomena detection that has been presented. These implications issue some clear recommendations that speak against current orthodoxy and are designed to improve the quality of research in the discipline.

2.6 Implications for Psychological Research

2.6.1 Adopting Bottom-Up Reasoning in Science

The hypothetico-deductive method continues to be the method of choice in the natural sciences and, as repeatedly stated, is also prominent in psychology. Partly for this reason, most scientists view scientific inference as a top-down affair in which the thrust of reasoning is from hypotheses and theories to their empirical test predictions. Unfortunately, hypothetico-deductive testing in psychology is often constrained by null hypothesis significance testing, and in combination, their empirical predictions are often much weaker than predictions about new empirical phenomena. Given that the hypothetico-deductive method allows the deduction of claims about empirical phenomena, psychology's standard hypothesis and theory testing practice would be improved if it strove for hypothetico-deductive tests of the existence of new phenomena.

In stark contrast to the hypothetico-deductive method, the character of reasoning from data to phenomena is clearly bottom-up, culminating as it often does in inductive inference to empirical generalizations. Given that science often looks to detect empirical phenomena before constructing explanatory theories, and the detection of new phenomena often gives theory construction its point, the methodology of bottom-up reasoning in science certainly deserves a prominent place alongside the more familiar top-down sequence. Moreover, the bottom-up character of scientific inference extends abductively from claims about phenomena to theories that plausibly explain those empirical claims. ATOM is a broad bottom-up theory of scientific method that endorses the inductive discovery of phenomena followed by the abductive construction of explanatory theory. Although some areas of psychology, such as human experimental psychology, engage in bottom-up research, this practice is far from universal. It is presumably for this reason that, in a recent issue of *Perspectives on Psychological Science* that looked at future directions psychology should take, the social and personality psychologists David Funder (2009) and Paul Rozin (2009) recommended the adoption of a bottom-up approach to psychological research. Their vision of a better psychology assigned central importance to descriptive, data-oriented research in which the discovery of important and interesting phenomena preceded the construction of explanatory theory.

2.6.2 In Praise of Inductive Method

Down through time, many students of science, including Peirce and Popper, have cast doubt on the importance of inductive reasoning in science, though they have often had simple views of induction in their sights. However, as noted earlier, phenomena detection that leads to empirical generalizations is, perforce, inductive in nature. I also noted that when inductive reasoning is used in the scientific context of phenomena detection, it takes on a material character, which makes it an empirical, not a logical, matter.

This empirical conception of inductive reasoning in science enables psychologists to endorse the inductivism of radical behaviorist methodology while eschewing its instrumentalist prescriptions for theorizing in favor of a realist outlook on explanatory theory. Because the establishment of empirical claims and the construction of theories (both empiricist and realist) are different sorts of undertakings, scientists should be able to decouple them with little difficulty. The inductive part of radical behaviorism is an account of phenomena detection that can also be found in the biological sciences, which Skinner endorsed as a model for psychological science (Sidman, 1960; Skinner, 1984). As such, it deserves a wider adoption in psychology than it currently receives.

2.6.3 The Threefold Importance of Replication

From time to time, psychologists rightly stress the importance of replication in science (e.g., Sidman, 1960; Thompson, 1994) and lament its lack of emphasis in their discipline. That phenomena detection accords replication pride of place among its research procedures is perhaps the strongest justification of the importance of replication in science.

This chapter has stressed the need to distinguish between and use both close and constructive replication. To repeat, these are different but related validating strategies. Close replication features as a "just checking" strategy to establish that data patterns are real. Constructive replication comes in two forms: direct replication, which endeavors to faithfully reproduce the original study in its entirety; and the more familiar form, which is a triangulation strategy designed to reveal the extent to which the results identified by successful close replication can be generalized. In the pursuit of phenomena, science must regularly practice all three forms of replication.

It follows that psychological science would benefit considerably from greater attempts to capitalize on the variety of replication strategies that are available (e.g., Muller, Otto, & Benignus, 1983; Lykken, 1991). In particular, it needs to place greater emphasis on direct replication, which is a form of research that is undervalued and difficult to publish in psychology. Another replication strategy with genuine payoff involves carrying out a true pilot study followed by a full replication. A true pilot study is itself a genuine research study in the small (Meehl, 1990). It is conducted not to see whether something works, or to gather a particular piece of information, but to gauge whether one can ascertain the existence of an appreciable effect. However, for this to happen, the pilot study must have the basic features of the main study, and they must be implemented with a high degree of rigor, or else it will not be usefully predictive of the main study outcomes. Save for some possible minor improvements suggested by the limitations of the pilot study, the main study will function in effect as a direct replication of the pilot precursor, thus providing further evidence that the effect holds for the study conditions in question.

Finally, in recognition of the need to use statistical methods that are in keeping with the practice of describing predictable phenomena, researchers in psychology should seek the generalizability of relationships rather than their statistical significance (Ehrenberg & Bound, 1993; Hubbard & Lindsay, 2013). Hence the need to use observational and experimental studies with multiple sets of data, observed under quite different sets of conditions. The appropriate task here is not to determine which model best fits a single set of data but to ascertain whether the model holds across different data sets. To repeat, seeking reproducible results through different forms of replication requires data analytic strategies that are designed to detect significant sameness rather than significant difference. The regular use of these strategies would help put statistical significance testing in its rightful place.

2.6.4 Guarding against Pseudophenomena

Given that a good deal of work in science is concerned with separating artifacts from real effects, it is important to distinguish between pseudophenomena and genuine phenomena. Claims about phenomena that are not true can be harmful to science. It is not just that their status as knowledge claims is misleading but, more importantly, that constructing theories to explain them lacks proper motivation and is largely a waste of research time and money.

A number of well-known empirical claims in psychology have masqueraded as justified claims about phenomena for a time because they were not subjected to sufficient peer scrutiny. For example, John Watson's famous Little Albert experiment, allegedly demonstrating the phenomenon of the conditioned reflex, was really based on unconfirmed pilot data and accepted uncritically into psychology's book of knowledge (Samelson, 1980). The so-called Hawthorne effect (the idea that the behavior during the course of an experiment can be altered by a subject's awareness of participating in the experiment) has been enshrined in many textbooks, although the Hawthorne studies yielded little support for this alleged empirical generalization (Jones, 1992). And Richard Hernstein's claim that the high heritability of intelligence was "psychology's best proved, socially significant empirical finding" was based on Burt's twin study data that were subsequently shown to have no scientific merit (Tucker, 1994).

Ongoing claims about the existence of parapsychological or psi phenomena, such as telekinesis and telepathy, present in bold relief the challenge of safeguarding against pseudophenomena. Critics of parapsychological experiments claim that reported psi effects result from either defects in experimental design, such as the use of inappropriate randomization procedures, or flaws in the use of statistical methods (e.g., Diaconis, 1978; Hyman, 1985). These critics also point out that when good experimental designs are used, parapsychologists cannot consistently replicate their results. Whereas a majority of academic psychologists today do not accept the existence of psi phenomena, a small minority of researchers continue to argue for their existence. Using customary *p* values as his source of evidence, Bem (2011) recently claimed to have demonstrated the existence of precognition whereby future events retroactively influence peoples' responses. Wagenmakers, Wetzels, Borsboom, and van der Maas (2011) reanalyzed Bem's data and concluded that their more stringent Bayesian analysis showed no evidence in favor of precognition. Given that psychologists often carry out empirical tests in the quasi-exploratory manner of Bem's study, Wagenmakers et al. concluded that psychologists in general should perform more conservative confirmatory tests of controversial claims. Their methodological recommendation is in keeping with the view adopted in this chapter that falsely claiming that phenomena exist has the potential to do serious harm to science. Some years ago, Bem and Honorton (1994) provided an example of this potential harm by claiming, on insufficient grounds (Milton & Wiseman, 1999), that psi exists, and then proceeded to idly speculate about the mechanism that might plausibly produce such an alleged phenomenon.

2.6.5 The Need to Reform Data Analytic Practice

The account of phenomena detection presented in this chapter lends weight to recommendations that have been made to change our data analytic practices in psychology (e.g., Kline, 2013; Wilkinson & the Task Force on Statistical Inference, 1999). Taking phenomena detection seriously requires researchers to be bullish, not bearish, about data analysis (Bogen & Woodward, 1988). For its part, psychology needs to be more bullish about data analysis. Although it has yet to properly embrace Tukey's (1980) two-stage model of exploratory data analysis followed by confirmatory data analysis, psychology should work explicitly with something like the four-stage model of data analysis outlined earlier.

This model underscores the need to give greater attention to a number of different types of method than is currently the case in psychological research. First, as noted earlier, the initial examination of data, which is undertaken to screen data for its quality, must be adopted on a more systematic basis than occurs at present (Fidell & Tabachnick, 2003; Wilkinson & the Task Force on Statistical Inference, 1999). Data must be worthy of reception as a source of potential evidence for phenomena claims. Emphasizing the concerted use of initial data analyses to check the quality of data should not prevent psychological researchers from appreciating that attending seriously to the performance characteristics of the instruments they use in data acquisition is an important additional means of exercising control over their quality. The sporadic attention given to calibration procedures in psychology is symptomatic of its need to give more systematic attention to the quality of its data gathering instruments.

Second, we must give exploratory data analysis a regular place in research and curriculum endeavors. More than fifty years have passed since Tukey (1969) made a compelling case to psychologists for the need to undertake exploratory data analysis as an essential part of modern data analytic practice. Psychology is slowly acknowledging the need to embrace data analysis in the exploratory mode for the purpose of pattern suggestion, but as a casual inspection of standard instructional textbooks makes clear, it still has some way to go.

Third, there is a related need to recognize that computer-intensive resampling methods, such as the bootstrap family, constitute an important set of statistical procedures that are well suited to the role of pattern confirmation. The American Psychological Association's Task Force on Statistical Inference (Wilkinson & the Task Force on Statistical Inference, 1999) was charged with looking at the newer computer-intensive statistical methods, but unfortunately it said nothing about computer-intensive resampling methods. More recently, however, a small group of methodologists and practicing researchers in psychology have begun to promote such methods (e.g., Kline, 2013; Sherman & Funder, 2009; Yu, 2008). One might hope that, with the increasing availability of suitable software, these statistical resampling methods might soon become a companion resource for exploratory data analytic methods in the psychological researcher's toolbox.

Finally, it is worth noting that the perspective on phenomena detection presented here militates against the continued heavy use of null hypothesis significance testing as psychology's mainstay in data analysis. As noted earlier in the discussion of replication, phenomena detection essentially involves the pursuit of significance sameness, not significant difference. More specifically, the widespread tandem use of exploratory data analytic and computer-intensive resampling methods would have the desirable effect of helping put classical significance testing in its proper place: performing the minor task of assessing sampling uncertainty. Seen in this light, the current expression of concern by some "*p*-minded" methodologists about overly liberal data analytic practices would become a minor worry.

2.6.6 A Division of Cognitive Labor

I have repeatedly emphasized in this chapter that phenomena detection is a very different enterprise from the construction of explanatory theory. Mindful of this difference, the physics community operates with a clear institutionalized division between experimental and theoretical research. The products of both forms of research are recognized as major achievements in their own right, and each type of research is undertaken by different sorts of people with different research skills. It is rare in physics to find people who can do both types of work well. Instead, specialized empirical and theoretical physicists characteristically work together in research teams.

By contrast, in psychology it is not unusual for empirical and theoretical work to be done by the same person or by groups of people with the same basic research training. Given the fundamental difference between the processes of phenomena detection and theory construction, and granting the complexity of studying the mind, the brain, and human behavior, it is pertinent to ask whether psychology might make better progress as a science by encouraging its researchers to adopt this division of cognitive labor. This is not to suggest that all psychologists should do so, or that psychology should institutionalize a division between empirical and theoretical research as strong as the one in physics. However, given that from a science policy perspective we do not really know in advance how best to proceed, it makes good sense to adopt a mixed strategy, with some researchers doing empirical work, others doing theoretical work, and a small minority with the requisite strengths doing both. I might add that although some evidence indicates that psychology is at last beginning to acknowledge the importance of theoretical research as legitimate work in its own right (Kukla, 2001; Slife & Williams, 1997; Wachtel, 1980), theory construction in psychology is a research task that receives little formal recognition and encouragement. The virtual neglect of theory construction in the guidelines of the *Publication Manual* (American Psychological Association, 2010) is a prominent case in point.

2.7 Conclusion

Methodologists, teachers, and researchers in psychology have seldom offered a full and accurate account of what good empirical research is all about. This is true of the important practice of phenomena detection, although most psychological researchers seem to spend a good deal of their time engaged in activities that are directly relevant to the detection of empirical regularities. However, it should be acknowledged that the process of phenomena detection is not always easy to understand. Evidence suggests that university students have difficulty distinguishing between phenomena claims and explanations (Norris, Phillips, & Korpan, 2003). Also, sophisticated scientists and philosophers can disagree on whether a piece of research explains data or phenomena (see, e.g., the recent exchange between Burnston, Sheredos, & Bechtel, 2011, and Kievit, Romeijn, Waldorp, Wicherts, Scholte, & Borsboom, 2011). Furthermore, it is not difficult to find examples of psychological writing in which the authors unwittingly conflate claims about phenomena and explanatory theories. Moreover, some psychologists deliberately run the two together. For example, Stam (2006) suggested that "the distinction between theory and fact is a rather dubious and unhelpful one" (30), and Schmidt (1993) opined that it is appropriate to take all research processes, including the formation of stable empirical relationships, to count as explanation. In my view, pronouncements such as these conflate the two fundamentally different core endeavors of basic psychological science and so misconstrue its nature. Robust empirical generalizations have a life of their own, and the regularities they describe are distinct from the causal factors that produce them. For this reason, they will often become the appropriate objects of scientific explanation, whereas theories about causal factors are the vehicles that provide the sought-after explanations.

The successful detection of a phenomenon is an important achievement in its own right, and a significant indicator of empirical progress in science. Bogen and Woodward's account of the scientific process of phenomena detection, and its attendant conception of the nature of science, is a systematic reconstruction of this part of science that is seldom presented as a whole in methodological writings. Although it is an outlook on empirical inquiry that psychologists have ignored, it provides an important means by which they can improve their understanding of this process. I hope that the perspective on phenomena detection presented in this chapter will help psychologists implement their bottom-up research strategies in a more informed and rigorous manner.

This concludes my exposition of phenomena detection, which is the first phase of ATOM. The following three chapters are concerned with the second phase of ATOM, theory construction. This phase comprises theory generation, theory development, and theory appraisal. The next chapter is concerned with theory generation. It focuses on the method of exploratory factor analysis and presents it as an abductive method for generating elementary plausible theories.

3 Theory Generation: Exploratory Factor Analysis

Exploratory factor analysis is an abductive method for formulating hypotheses using the common cause principle, but also to be used along with confirmatory factor analysis, which tests hypotheses.

—Stanley Mulaik (2010, 433)

3.1 Introduction

Exploratory factor analysis is a multivariate statistical method designed to facilitate the postulation of latent variables that are thought to underlie and give rise to patterns of correlations in new domains of observed or manifest variables. Intellectual abilities, personality traits, and social attitudes are well-known classes of latent variables that are the products of factor analytic research. Exploratory factor analysis (EFA) is often contrasted with confirmatory factor analysis, which is concerned with the testing of factor analytic hypotheses and models.

The first sixty years of the hundred-year history of factor analysis were largely devoted to developing exploratory factor analytic methods. However, despite the advanced statistical state and frequent use of EFA within psychology and other behavioral sciences, debate about its basic nature and worth continues. Most factor analytic methodologists take EFA to be a method for hypothesizing latent variables to explain patterns of correlations. Some, however, understand it as a method of data reduction that provides an economical description of correlational data.[1,2] Further, with the advent of confirmatory factor analysis and full structural equation modeling, the prominence of EFA in multivariate research has declined. Today methodologists and researchers often recommend and employ confirmatory factor analysis as the method of choice in factor analytic studies.

In this chapter, I examine the methodological foundations of EFA and argue for the view that it is properly construed as a method for generating rudimentary explanatory theories.[3] In the first half of the chapter, I contend that EFA is an abductive method of theory generation that exploits an important precept of scientific inference known as the *principle of the common cause*. It is surprising that this characterization of the inferential nature of EFA does not figure explicitly in the factor analytic literature, because it coheres well with the generally accepted view of EFA as a latent variable method. Since abduction and the principle of the common cause are seldom mentioned in the factor analytic literature, I describe each before showing how they are employed in EFA. In the second half of the chapter, I refer again to ATOM, which I outlined in chapter 1. I then discuss a number of methodological features of EFA in the light of that method. In particular, I argue that, despite a widespread belief to the contrary, factorial theories do have genuine explanatory merit; the methodological challenge of factor indeterminacy can be satisfactorily met by both EFA and confirmatory factor analysis; and EFA as a useful method of theory generation can profitably be employed in tandem with confirmatory factor analysis and other methods of theory evaluation. The epigraph by Stanley Mulaik (2010) at the beginning of the chapter is part of his summary statement of my position on factor analysis (Haig, 2005b).

3.2 The Inferential Nature of Exploratory Factor Analysis

3.2.1 The Nature of Abductive Inference

It is commonly thought that inductive and deductive reasoning are the only major types of inference employed in scientific research. It is well known that conclusions of valid deductive arguments preserve the information or knowledge contained in their premises, but they do not add new information or knowledge. By contrast, inductive arguments are ampliative in that they add new information or knowledge to existing information and knowledge. However, inductive arguments, though ampliative, are *descriptive* in character because they reach conclusions about the same types of manifest attributes mentioned in their premises. Importantly, though, science also adds to its store of knowledge by reasoning from factual premises to *explanatory* conclusions. This type of inference, which is widely ignored in scientific methodology, is known as *abduction*.

The basic idea of abductive inference can be traced back to Aristotle, but its modern formulation comes from the pioneering work of the American philosopher and scientist Charles Sanders Peirce (1931–1958). Peirce's writings on abduction are underdeveloped and open to interpretation, but they are richly suggestive. They were largely ignored in the first half of the twentieth century, but more recent developments in the fields of philosophy of science, artificial intelligence, and cognitive science more generally (e.g., Josephson & Josephson, 1994; Magnani, 2001; Thagard, 1988, 1992) have built on Peirce's ideas to significantly advance our understanding of abductive reasoning.

Abduction is a form of reasoning involved in generating and evaluating explanatory hypotheses and theories. For Peirce, "Abduction consists in studying facts and devising a theory to explain them" (1931–1958, Vol. 5, 90). It is "the first starting of an hypothesis and the entertaining of it, whether as a simple interrogation or with any degree of confidence" (1931–1958, Vol. 6, 358).

Peirce maintained that abduction had a definite logical form that he came to represent in the following general schema (1931–1958, Vol. 5, 117):

The surprising fact, C, is observed.

But if A were true, C would be a matter of course.

Hence, there is reason to suspect that A is true.

Although Peirce's schematic depiction of abductive inference is suggestive, it needs to be amplified and modified in various ways to qualify as an instructive account of explanatory inference in science. First, as emphasized in the previous chapter, the facts to be explained in science are not normally particular events but empirical generalizations or phenomena, and, strictly speaking, they are not typically observed. Rather, the data themselves are observed and serve as evidence for the phenomena. In turn, phenomena, not data, serve as evidence for the abduced theories.

Second, confirmation theory in the philosophy of science makes clear that the facts or phenomena follow as a matter of course, not just from the proposed theory but from that theory in conjunction with accepted relevant auxiliary claims taken from background knowledge.

Third, we should not take the antecedent of the conditional assertion in Peirce's schema to imply that abductive inferences produce truths as a matter of course. Although science aims to give us true, or approximately true, theories of the world, the supposition that the proposed theory is true is not a requirement for the derivation of the relevant facts. All that is required is that the theory be plausible enough to be provisionally accepted. It is important to distinguish between *truth*, understood as a guiding ideal for science (a goal that we strive for but never fully reach), and the *acceptance* of theories, which is based on evaluative criteria such as predictive success, simplicity, and explanatory breadth. As proxies for truth, justificatory criteria such as these indicate truth but do not constitute it.

Fourth, note that the conclusion of Peirce's argument schema does not assert that the hypothesis itself is true, only that there are grounds for thinking that the proposed hypothesis might be true. This is a weaker claim that allows one to think of a sound abductive argument as delivering a judgment that the hypothesis is initially plausible and worthy of further pursuit. As we shall see, assessments of initial plausibility constitute a form of generative justification that involves reasoning from warranted premises to an acceptance of the knowledge claims in question.

Fifth, Peirce's schematic depiction of abductive inference focuses on its logical form only. As such, it has limited value in understanding the theory construction process unless it is conjoined with a set of regulative constraints that enable us to view abduction as an inference, not just to any explanation but to plausible explanations. Constraints that regulate the abductive generation of scientific theories will comprise a host of heuristics, rules, and principles that govern what counts as good explanations. In the next section, I argue that the principle of the common cause is a key principle (more accurately, a heuristic) that regulates abductive reasoning within EFA.

Peirce's understanding of abduction was somewhat protean in nature, although for him it tended to take its place at the inception of scientific hypotheses and often involved making inferences from puzzling facts to hypotheses that might well explain them. However, recent work on abduction reveals that explanatory hypotheses can be abductively obtained in a number of different ways. In focusing on the generation of hypotheses, Thagard (1988) helpfully distinguishes between different types of abduction. One of these, existential abduction, hypothesizes the existence of previously unknown objects or properties. Another, analogical abduction, employs successful past cases of hypothesis generation to form new hypotheses similar to relevant existing ones. In the next section, I suggest that existential abduction is the type of abduction involved in the factor analytic production of explanatory hypotheses, although analogical abduction too is sometimes employed in this regard.

It should be clear from this series of remarks about abduction that Peirce's schematic depiction of the logical form of abduction needs to be changed to something like the following:

The surprising empirical phenomenon, P, is detected.

But if hypothesis H were approximately true, and the relevant auxiliary knowledge, A, were invoked, then P would follow as a matter of course.

Hence there are grounds for judging H to be initially plausible and worthy of further pursuit.

This recasting of Peirce's characterization of an abductive argument accommodates the fact that in science, hypotheses are typically produced to explain empirical phenomena. Moreover, it acknowledges the role of background knowledge in the derivation of hypotheses, assigns a regulative role to truth, and signals the importance of initial plausibility assessments in generating and developing new knowledge.

3.2.2 Exploratory Factor Analysis and Abduction

I turn now to consider my initial claim that EFA is fundamentally an abductive method of theory generation. I begin by briefly acknowledging two earlier efforts to characterize EFA as an abductive method and then elaborate on the claim that EFA largely trades in existential abductions. In part, this exercise will involve indicating that the modified Peircean schema for abductive inference applies to EFA.

Sixty years ago, Raymond Hartley (1954) drew a distinction between descriptive and inferential factor analysis and defended the then unpopular view that inferential factor analysis could justifiably be used to hypothesize unobserved causal factors. Hartley argued his case by analogy to the logic involved in the study of unobserved physiological entities, but he realized that one could make a compelling case for the inferentialist reading of factor analysis only by appealing to an appropriate theory of inference. Hartley expressed surprise at the time that factor analysis stood without appeal to any theory of inference. It is remarkable, then, that expositions of EFA sixty years later still do not refer explicitly to a theory of inference to characterize the reasoning involved in moving from descriptions of manifest variables to statements about latent variables.

Although the mainstream psychometric literature does not attempt to characterize EFA as an abductive method, both William Stephenson and William Rozeboom began to address this matter over forty years ago. Stephenson's (1961) insightful scientific creed contains a brief attempt to explicitly characterize EFA as an abductive method, and Rozeboom's work (1961, 1972) provides more detailed evidence supporting the view that EFA is an abductive method. Rozeboom spoke of *ontological inductions* that extend our referential reach beyond covariational information to hypotheses about latent factors, which are new ontological postulations. He also described EFA as an *explanatory inductive* method because it helps generate conceptions of latent factors that explain why the covariational regularities of interest obtain. Here Rozeboom used the term *induction* in a broad sense, where it has the same general meaning as *abduction*.

As noted earlier, existential abduction often hypothesizes the existence of entities previously unknown to us. The innumerable examples of existential abduction in science include the initial postulation of entities such as atoms, phlogiston, genes, viruses, tectonic plates, planets, Spearman's *g*, habit strength, and extraversion.[4] We now know that some of these entities exist, that some of them do not exist, and we are unsure about the existence of others. In cases like these, the initial abductive inferences are made to claims primarily about the *existence* of theoretical entities to explain empirical facts or phenomena. Thus, in the first instance, the hypotheses given to us through the use of EFA do little more than postulate the existence of the latent variables in question. They say little about their nature and function, and it remains for further research to elaborate on the first rudimentary conception of these variables.

The factor analytic use of existential abduction to infer the existence of the theoretical entity *g* can be coarsely reconstructed in accord with the earlier modified Peircean schema for abductive inference along the following lines:

The surprising empirical phenomenon known as the *positive manifold* is identified.[5]

If *g* exists, and it is validly and reliably measured by a Wechsler intelligence scale (or some other objective test), then the positive manifold would follow as a matter of course.

Hence there are grounds for judging the hypothesis of *g* to be initially plausible and worthy of further pursuit.

I remarked earlier that our conceptions of the latent factors of EFA come to us through existential abductions. In fact, the factor analytic generation of hypotheses is sometimes a mixture of existential and analogical abduction where we simultaneously posit the existence of a latent variable and offer the beginnings of a characterization of that entity by brief analogy to something that we understand quite well. Recall that analogical abduction appeals to known instances of successful abductive hypothesis formation to generate new hypotheses like them. To accommodate the presence of analogical abduction, the abductive argument schema just given would need an additional premise that indicates there is reason to believe that a hypothesis of the appropriate kind would explain the positive manifold. When Charles Spearman first posited general intelligence to explain correlated performance indicators, he thought of it as mental energy, likening it to physical energy—a process well understood by the physics of the time. His initial inference to claims about *g*, then, was a blend of existential and analogical abduction.

This example serves to illustrate the point that methodologists should take the method of EFA proper to include the factor analyst's substantive interpretation of the statistical factors. In this regard, it is important to realize that the exploratory factor analyst has to resort to his or her own abductive powers when reasoning from correlational data patterns to underlying common causes. This point can be brought out by noting that the modified Peircean schema for abduction, and its application to the factor analytic generation of Spearman's hypothesis of *g*, are concerned with the form of the arguments involved, not with the actual generation of the explanatory hypotheses. In each case, the explanatory hypothesis is *given* in the second premise of the argument. An account of the genesis of the explanatory hypothesis must therefore be furnished by some other means. I think it is plausible to suggest that reasoning to explanatory hypotheses trades on our evolved cognitive ability to abductively generate such hypotheses. Peirce himself maintained that the human ability to engage readily in abductive reasoning was founded on a guessing instinct that has its origins in evolution. More suggestively, Carruthers (2002) claimed that our ability to engage in explanatory inference is almost certainly largely innate, and he speculated that it may be an adaptation selected for because of its crucial role in the fitness-enhancing activities of our ancestors such as hunting and tracking. Whatever its origin, an informative methodological characterization of the abductive nature of factor analytic inference must appeal to the scientist's own psychological resources, as well as those of logic. To recall a tenet of the realist methodology outlined in chapter 1, it must be a methodological characterization that includes the knowing subject.[6]

Before leaving consideration of the general abductive nature of EFA, let us briefly note that a number of special features of EFA play an important role in facilitating the abductive generation of hypotheses. For instance (as we will see in chapter 5), simplicity, or parsimony, is an important desideratum in fashioning scientific explanations, and Thurstone's (1947) criteria for simple structure combine in an explicit formulation of parsimony in EFA. Stated in the distinctive language of factor analysis, Thurstone's insight was to appreciate that rotation to the oblique simple structure solution provided an objective basis for acceptable terminal factor solutions that included reference to latent as well as manifest variables.

3.2.3 The Principle of the Common Cause

Having suggested that abduction, specifically existential abduction, largely characterizes the type of inference employed in the factor analytic generation of theories about latent variables, I now want to draw attention to a methodological principle that drives and shapes the nature of the existential abductive inference involved in EFA. It is well known that EFA is a common factor analytic model in which the latent factors it postulates are referred to as *common* factors. Not surprisingly, these factors are often understood, and sometimes referred to, as common *causes*. Yet seldom have factor analytic methodologists attempted to formulate a principle or maxim of inference that guides the reasoning to common causes. There is, however, an important principle of scientific inference, known in philosophy of science as the *principle of the common cause*, that we can apply to good effect here. In what follows, I discuss the principle of the common cause before spelling out its central role in EFA.

In *The Direction of Time*, Hans Reichenbach (1956) maintained that, in both scientific and everyday reasoning, we often explain a coincidence by postulating a common cause. In recognition of this fact, he explicitly formulated a maxim that he called the *principle of the common cause*. Reichenbach stated the principle cryptically, and informally, thus: "If an improbable coincidence has occurred, there must exist a common cause" (157). For Reichenbach, this principle enjoins us to postulate a single common cause whenever there are events, or classes of events, that are statistically significantly correlated. To take one of Reichenbach's original examples, if two lights in a room go out suddenly, the principle of the common cause says we should look for an interruption in their common power supply, such as a blown fuse.

Although Reichenbach's formulation of the principle will not do as it stands, the principle can be formulated as an important precept of human reasoning that governs a good deal of inference in science. The principle of the common cause has received some consideration in the philosophical literature and sometimes appears to be tacitly employed in behavioral research, but it has been widely ignored in general scientific methodology.

In explicitly introducing the principle of the common cause, Reichenbach was concerned to capture the idea that if two events, A and B, are correlated, then one might be the cause of the other. Alternatively, they might have a common cause, C, where this cause always occurs before the correlated events. Reichenbach was the first to make this idea precise, and he did so by formulating it as a statistical problem. He suggested that when for simultaneous events A and B, $Pr(A \& B) > Pr(A) \times Pr(B)$, there exists an earlier common cause, C, of A and B, such that $Pr(A/C) > Pr(A/{\sim}C)$, $Pr(B/C) > Pr(B/{\sim}C)$, $Pr(A \& B/C) = Pr(A/C) \times Pr(B/C)$ and $Pr(A \& B/{\sim}C) = Pr(A/{\sim}C) \times Pr(B/{\sim}C)$ (Reichenbach, 1956, 158–159). The common cause C is said to "screen off" the correlation between A and B, when A and B are uncorrelated, conditional on C. A common cause screens off each effect from the other by rendering its correlated effects (conditionally) probabilistically independent of each other. For example, given the occurrence of a flash of lightning in the sky, a correlation between two people apparently observing that flash is not just a coincidence, but is due to the flash of lightning being a common cause. Further, the probability of one person seeing the flash of lightning, given that it does occur, is not affected by whether or not the other person observes the lightning flash. Reichenbach's principle of the common cause can thus be formulated succinctly as follows: "Simultaneous correlated events have a prior common cause that screens off the correlation."

Although Reichenbach's initial characterization of the principle of the common cause has some intuitive appeal and precision, more recent philosophical work (Arntzenius, 1993; Salmon, 1984; Sober, 1988) has suggested that the principle needs to be amended in a number of ways. First, not every improbable coincidence or significant correlation has to be explained through a common cause. For this reason, the principle is sometimes taken to say, "If an improbable coincidence has occurred, and there is no direct causal connection between the coincident variables, then one should infer a common cause." However, this amendment does not go far enough, for there are a number of other possible alternative causal interpretations of correlations. For example, two correlated variables might be mediated by an intervening cause in a developmental sequence, or they might be the result of separate direct causes, and so on. Responsible inference to a common cause must rule out alternative causal interpretations like these. We may therefore further amend Reichenbach's formulation of the principle to the following: "Whenever two events are improbably, or significantly, correlated, we should infer a common cause, unless we have good reason not to." Clearly the principle should not be taken as a hard-and-fast rule, for in many cases, proper inferences about correlated events will not be of the common causal kind. The qualifier "unless we have a good reason not to" should be understood as an injunction to consider causal interpretations of the correlated events other than the common causal kind. Also, occasions will arise when it is incorrect to draw any sort of causal conclusion. Some correlations are accidental correlations that are not brought about by causes.

The existence of different attempts to improve on Reichenbach's (1956) initial formulation of the principle of the common cause leads to the idea that more than one acceptable version of the principle might exist. We might expect this to be the case not just because Reichenbach's formulation of the principle needs improving but also because of the important point that different subject matters in different domains might well require different formulations of the principle. For example, Reichenbach, a philosopher of physics, took the principle to apply to correlated events that are spatially separated. However, behavioral and social scientists regularly infer common causes for events that are not spatially separated. This is clearly the case in psychology, where the correlated variables can be performance measures on tests of intelligence and personality. Further, Sober (1988) has argued that in evolutionary theory, phylogenetic inference to common ancestry involves postulating a common cause, but this will be legitimate only if certain assumptions about the process of evolution are true. Thus, in formulating a principle of the common cause in a way that can be used effectively in a given domain, relevant contingent knowledge about that domain will shape the formulation of the principle and moderate its use. As noted in my earlier characterization of abduction, the production of scientific knowledge is a three-termed relation between evidence, theory, and background knowledge. Routine use of a fixed, general formulation of the principle of the common cause that reasons from correlational data alone is unlikely to lead consistently to appropriate conclusions.

Two related features of the principle of the common cause should also be acknowledged: as Salmon (1984) has observed, the principle is sometimes used as a principle of explanation (we appeal to common causes to *explain* their correlated effects), and it is sometimes used as a principle of inference (we use the principle to *reason* to common causes from their correlated effects). The principle of the common cause is a form of abductive inference where one reasons from correlated events to common causes thought to explain those correlations. Thus we can go further than Salmon and claim that the principle of the common cause simultaneously combines these explanatory and inferential features to yield explanatory inferences.

The suggestion that there might be different versions of the principle of the common cause prompts mention of a closely related principle that Spirtes, Glymour, and Scheines (2000) call the *Markov condition*. This principle has recently been employed in Bayesian network modeling of causal relations. Roughly stated, the Markov condition says that, conditional on its direct causes, a variable is probabilistically independent of everything except its effects. The Markov condition is in effect a generalized screening-off condition from which one can derive a version of the principle of the common cause as a special case. As a generalized screening-off condition, the Markov condition applies to both common and intervening causes. By contrast, the principle of the common cause only screens off common causes from their correlated effects. Because of this restriction, the principle of the common cause can be taken as the appropriate screening-off requirement for EFA.

I turn now to the application of the principle of the common cause to EFA.

3.2.4 Exploratory Factor Analysis and the Principle of the Common Cause

The Need for the Principle of the Common Cause It is sometimes said that the central idea in factor analysis is that the relations between a large number of observed variables are the direct result of a smaller number of latent variables. McArdle (1996) maintains that this is a theoretical principle that empirical researchers employ to identify a set of underlying factors. However, while true of EFA, this principle does not constrain factor analysts to infer the *common* latent factors that are the appropriate outcome of using common factor analysis. For this to happen, the principle has to be linked to the principle of the common cause or recast in more specific methodological terms in accordance with that principle. Not only does the principle of the common cause enjoin one to infer common causes, but it also assumes that that those inferences will be to relatively few common causes. Reichenbach's (1956) original formulation of the principle, which allows inference to just one common cause, is obviously too restrictive for use in multiple factor analysis. However, amending the principle to allow for more than one common cause, combined with the restraint imposed by following Ockham's razor (do not multiply entities beyond necessity), will enable one to infer multiple common causes without excess.

Although EFA is used to infer common causes, expositions of common factor analysis that explicitly acknowledge the importance of the principle of the common cause are difficult to find. Kim and Mueller's (1978) basic exposition of factor analysis is a noteworthy exception. In discussing the conceptual foundations of factor analysis, these authors evince the need to rely on what they call the *postulate of factorial causation*. They characterize the postulate of factorial causation as "the assumption that the observed variables are linear combinations of underlying factors, and that the covariation between observed variables is solely due to their common sharing of one or more of the common factors" (78). The authors make clear that the common factors mentioned in the assumption are to be regarded as underlying causal variables. Taken as a methodological injunction, this postulate functions as a variant of the principle of the common cause. Without appeal to this principle, factor analysts could not identify the underlying factor pattern from the observed covariance structure.

Two features of the principle of the common cause that make it suitable for EFA are that it can be applied in situations where we do not know how *likely* it is that the correlated effects are due to a common cause (this feature is consistent with the views of Reichenbach [1956], Salmon [1984], and Sober [1988] on common causal reasoning), and also in situations where we are essentially ignorant of the *nature* of the common cause. The abductive inference to common causes is a basic explanatory move that is nonprobabilistic and qualitative in nature. It is judgments about the soundness of the abductive inferences, not the assignment of probabilities, that confer initial plausibility on the factorial hypotheses spawned by EFA.

It is important to appreciate that the principle of the common cause does not function in isolation from other methodological constraints. Embedded in EFA, the principle helps to limit existential abductive inference to situations where we reason back from *correlated* effects to one or more *common* causes. Although covariation is an important basic datum in science, not all effects are expressed as correlations, and, as noted earlier, not all causes are of the common causal variety. It follows from this that researchers should not always look for common causal interpretations of multivariate data, for there are numerous alternative latent variable models. The simplex model of latent variables is a case in point (e.g., Mulaik & Millsap, 2000). Further, the frequency of proper use of EFA should be much less than the frequency of proper use of the principle of the common cause, because the principle can be employed by non-factor-analytic means, as will be indicated later.

In this first half of the chapter, I have argued that an appeal to abductive inference, linked to the principle of the common cause, leads naturally to the view that EFA is an abductive method of theory generation that enables researchers to theorize the existence of latent variables. Although this method uses the statistical ideas of multiple regression and partial correlation, it does so to facilitate inferences to the latent variables. In the view presented here, EFA is glossed as a set of multivariate procedures that help us reason in an existentially abductive manner from robust correlational data patterns to plausible explanatory prototheories via the principle of the common cause.

3.3 Common Factor Analysis and Scientific Method

In the chapter's second half, I propose to speak about the place of common factor analysis in scientific inquiry broadly understood. To this end, I briefly discuss the restrictions of two well-known theories of scientific method, before adopting ATOM. This broader theory will serve to provide a useful methodological framework within which one can locate, further explicate, and evaluate the nature and role of EFA in scientific research. In this regard, my primary concern will be to argue that EFA helps researchers generate theories with genuine explanatory merit; that factor indeterminacy is a methodological challenge for both EFA and confirmatory factor analysis but is a challenge that can nevertheless be satisfactorily met; and that, as a valuable method of theory generation, EFA can be employed profitably in tandem with its confirmatory namesake and other theory evaluation methods.

3.3.1 Exploratory Factor Analysis and Scientific Method

Much of the history of the development of general theories of scientific method has discussed the relative merits of inductive and hypothetico-deductive theories (Laudan, 1981). Mulaik (1987) locates EFA historically within eighteenth- and nineteenth-century empiricist philosophy of science and its restrictive inductivist conception of scientific inquiry. The inductive view of scientific method was said to obtain knowledge from experience by establishing generalizations based on theory-free observations. The scientific ideal of that time held inductive method to be an organon for the discovery of secure knowledge that is devoid of explanatory hypotheses. Today, of course, it is a methodological truism to claim that such a method cannot exist, and Mulaik is clearly right to point out that we cannot expect EFA to deliver such knowledge. However, even a modern view of inductive method, understood as a fallible generator of empirical generalizations, cannot properly accommodate EFA as a latent variable method. As noted at the beginning of the chapter, generalizing inductive inference is descriptive inference, in the sense that it licenses inferences to more of the manifest attributes that are sampled; it does not have the conceptual resources to reach latent source variables that are understood as causal entities. For this to be possible, an explanatory form of ampliative inference is needed, as my earlier remarks on abduction and its relevance to EFA have sought to make clear.

As already noted, the hypothetico-deductive account of scientific method has assumed hegemonic status in twentieth-century psychology. As such, it continues to sustain the popular view that scientific research is essentially a matter of testing hypotheses and theories, as well as the corollary that there are no scientific methods for formulating hypotheses and theories (Hempel, 1966). Although confirmatory factor analysis finds a natural home within the confines of hypothetico-deductive method (more of which later), EFA stands outside that method, offering an abductive logic of theory generation that the hypothetico-deductive method implies is possible.

As sketched in its précis in chapter 1, ATOM attempts to bring together an array of ideas on important aspects of the research process, many of which fall outside the province of the standard inductive and hypothetico-deductive accounts of scientific method. Of particular relevance to this chapter is that theory generation is depicted as an abductive process, a fact that enables the abductive theory of method to incorporate EFA within its fold. When this happens, EFA functions as a submethod of ATOM and serves to provide a detailed methodological account of how theories about common causes can be abductively generated from correlational evidence. ATOM is also able to subsume the inductive account of method. With its emphasis on generalization, the inductive method can be seen at work in the process of phenomena detection.

Before turning to EFA again, let us note three points about the relation between EFA and ATOM. First, the justification for adopting ATOM is confined to the fact that it facilitates the examination of EFA in a suitably broad methodological perspective. Second, the justification for the abductive depiction of EFA, given in the chapter's first half, has been developed independently of the acceptance of ATOM and as such can be used outside its ambit. Third, the abductive employment of EFA within the theory generation phase of ATOM begs no important question about the abductive nature of that phase. Rather, it lends credibility to ATOM's outlook on theory generation by offering just one specific account of that process.

3.4 Exploratory Factor Analysis, Phenomena Detection, and Explanatory Theories

3.4.1 Exploratory Factor Analysis and Phenomena Detection

As just noted, ATOM contends that scientific research often involves the initial detection of empirical phenomena, followed by the construction of explanatory theories to understand those phenomena. Here I want to emphasize an important feature of EFA by suggesting that, strictly speaking, it contributes to phenomena detection as well as theory construction. As such, it is a "mixed method," having both data analytic and theory generation roles.[7]

Otherwise distinct accounts of scientific inquiry tend to share the view that scientific theories explain and predict facts about observed data. However, as noted earlier in the discussion of Peirce's (1931–1958) original characterization of abductive inference, this widely held view fails both to distinguish between data and phenomena and, in consequence, to appreciate that typically it is phenomena, not data, that our theories are constructed to explain and predict. Recall that phenomena, unlike data, are relatively stable, recurrent features of the world that we seek to explain, and it is their generality and stability that make them, not data, the appropriate objects of explanation. In extracting phenomena from the data, we often use statistical methods. EFA is a case in point. Its name notwithstanding, EFA is not a particularly exploratory method, but it is nevertheless used to seek replicable data patterns, which are a standard requirement for making claims about phenomena. We can see this in the methodological requirement, stated initially by Thurstone (1947) and endorsed by Cattell (1978), that the obtained factor pattern should be repetitive, or invariant, across different data sets in distinct populations. Both of these pioneers of factor analysis realized that an interpretation of extracted and rotated factor patterns made little scientific sense if they were specific to a particular covariance matrix and did not, or were unlikely to, generalize to other covariance matrices.

3.4.2 Exploratory Factor Analysis and Explanatory Theories

One challenge to the interpretation of EFA as an abductive method of theory generation is the claim that the theories it produces have little explanatory worth. In countering this criticism, I suggest that factorial theories spawned by EFA are essentially dispositional in nature, and dispositional theories do have genuine, though limited, explanatory import (Rozeboom, 1984; Sober, 1982). Recall that existential abduction postulates the existence of new entities without being able to characterize their nature. Thus, in exploiting this form of abduction, EFA provides us with an essentially dispositional characterization of the latent entities it postulates.

Dispositional theories provide oblique characterizations of the properties we attribute to things by way of their presumed effects under specified conditions (e.g., Mumford, 1998; Tuomela, 1978). For example, the brittleness of glass is a dispositional property causally responsible for the breaking of glass objects when they are struck with sufficient force. Our indirect characterization of this latent property, brittleness, is in terms of the relevant striking and breaking events. Similarly, Spearman's original theory of *g* was essentially dispositional in nature, for *g* was characterized obliquely in terms of children's school performance under the appropriate test conditions.

As I have just noted, dispositional theories have often been regarded as explanatorily suspect. Perhaps the best-known, and most frequently cited, example of this is Molière's scoff at explaining the soporific effects of opium by appeal to its dormitive power. However, as Rozeboom (1973) maintains, "the *virtus dormitiva* of opium *is* why people who partake of this particular substance become drowsy. Of course, that by itself leaves a great deal unknown about this power's nature, but learning of its existence and how to diagnose its presence/absence in particular cases is a necessary preliminary to pursuit of that knowledge" (67).

Similarly, with EFA, the existential abductions to latent factors postulate the existence of these factors without being able to say much, if anything, about their actual nature. It is the job of EFA to help us bring our factorial hypotheses and theories into existence, not to develop them and specify their nature. According to ATOM, the latter task is undertaken through the use of analogical modeling strategies. To expect EFA to develop theories, as well as generate them, is to fail to understand its proper role as a generator of dispositional theories.

An answer to the question of whether dispositional theories possess genuine explanatory worth requires us to focus on whether such theories have explanatory power. Two aspects of explanatory power that are relevant here are explanatory depth and explanatory breadth. For factorial theories, explanatory depth is naturally understood as existential depth. Existential depth is accorded those explanatory theories in science that are deep-structural in nature. Theories of this sort postulate theoretical entities that are different in kind, and hidden, from the empirical regularities they are invoked to explain. In postulating theoretical entities, deep-structural theories extend our referential reach to new entities and thereby increase the potential scope of our knowledge. The factorial theories afforded us by EFA have existential depth because the typical products of factor analytic abductions are new claims about hidden causal entities that are thought to exist distinct from their manifest effects. Existential depth deserves to be considered as an explanatory virtue of EFA's postulational theories.

The other feature of explanatory power, explanatory breadth, is a long-standing criterion of a theory's worth. Sometimes explanatory breadth is understood as *consilience*, which is often portrayed as the idea that a theory explains more of the evidence (a greater number of facts) than its competitors. The rudimentary theories of EFA do not have consilience in this sense, for they typically do not explain a range of facts. Nor are they immediately placed in competition with rival theories. However, factorial theories of this kind are consilient in the sense that they explain the *concurrences* embodied in the relevant patterns of correlations. By appealing to common causes, these factorial theories unify their concurrences and thereby provide us with the beginnings of an understanding of why they concur.

The two criteria that make up explanatory power are not the only dimensions of theory appraisal that we should consider when submitting a factorial theory to preliminary evaluation. The fertility of a theory is also an important evaluative consideration. In general terms, this dimension focuses on the extent to which a theory stimulates further positive research. It should be noted here that although our initial dispositional descriptions of latent factors are low in informational content, they do not, or need not, act as a heuristic block to further inquiry, as some commentators on factor analysis suggest. Lykken (1971), for example, judges latent variable explanations from factor analysis to be "stillborn," whereas Skinner (1953) declares that they give us false assurances about the state of our knowledge. However, given that EFA trades in existential abductions, the dispositional ascription of latent factors should serve a positive heuristic function. Considered as a preliminary to what we hope will eventually be full-blooded explanations, dispositional ascriptions serve to define the scope of, and mark a point of departure for, appropriate research programs. Viewed in this developmental light, dispositional explanations promote inquiry rather than block it.

3.4.3 Exploratory Factor Analysis and the Specter of Underdetermination

The methodological literature on factor analysis has given considerable attention to the indeterminacy of factors in the common factor model. Factor indeterminacy arises because the common factors are not uniquely determined by their related manifest variables. As a consequence, a number of different common factors can be produced to fit the same pattern of correlations in the manifest variables.

Although typically ignored by factor analytic researchers, factor indeterminacy is an epistemic fact of life that continues to challenge factor analytic methodologists. Some methodologists regard factor indeterminacy as a serious problem for common factor analysis and recommend using alternative methods such as principal components analysis because they are considered to be determinate in nature. Others have countered variously that component analysis models are not causal models (and therefore are not proper alternatives to common factor models), that they do not typically remain invariant under the addition of new variables, and that the indeterminacy of factor scores is seldom a problem in interpreting common factor analytic results because factor scores do not have to be computed.

One constructive perspective on the issue of factor indeterminacy has been offered by Mulaik and McDonald (McDonald & Mulaik, 1979; Mulaik, 1987; Mulaik & McDonald, 1978). Their position is that the indeterminacy involved in interpreting the common factors in EFA is just a special case of the general indeterminacy of theory by empirical evidence widely encountered in science, and it should therefore not be seen as a debilitating feature that forces us to give up on common factor analysis. Essentially, I agree with this outlook on the factor indeterminacy issue and will discuss it in this light. I argue that EFA helps us produce theories that are underdetermined by the relevant evidence, and the methodological challenge that this presents can be met in an acceptable way. I conduct my discussion against the backdrop of the sketch of ATOM provided in chapter 1.

Indeterminacy is pervasive in science. It occurs in semantic, metaphysical, and epistemological forms (McMullin, 1995). Factor indeterminacy is essentially epistemological in nature. The basic idea of epistemological, or more precisely methodological, indeterminacy is that the truth or falsity (better, acceptance or rejection) of a hypothesis or theory is not determined by the relevant evidence (Duhem, 1954). In effect, methodological indeterminacy arises from our inability to justify accepting one theory among alternatives on the basis of empirical evidence alone. This problem is sometimes referred to as the *underdetermination of theory by data*, and sometimes as the *underdetermination of theory by evidence*. However, because theories are often underdetermined by evidential statements about phenomena, rather than data, and because evidence in theory appraisal will often be superempirical as well as empirical in nature, I will refer to the indeterminacy here as the underdetermination of theory by *empirical evidence* (UTEE).

To construe factor indeterminacy as a variant of UTEE is to regard it as a serious problem, for UTEE is a strong form of underdetermination that needs to be reckoned with in science. Indeed, as an unavoidable fact of scientific life, UTEE presents a major challenge for scientific methodology.

Concerning scientific method, UTEE occurs in a number of places. The two that are relevant to common factor analysis are (a) ATOM's context of theory generation, where EFA can be employed as an abductive generator of rudimentary explanatory theories; and (b) the context of theory evaluation, where confirmatory factor analysis can be used to test factorial theories in an essentially hypothetico-deductive manner. Here I discuss factor indeterminacy as UTEE for EFA. I briefly address the issue of factor indeterminacy as it affects confirmatory factor analysis in the penultimate section of this chapter.

Mulaik (1987) sees UTEE in EFA as involving inductive generalizations that go beyond the data. I believe that the *inductive* UTEE should be seen as applying specifically to the task of establishing factorial invariance where one seeks constructive or external replication of factor patterns. However, for EFA we also need to acknowledge and deal with the *abductive* UTEE involved in the generation of explanatory factorial theories. The sound abductive generation of hypotheses is essentially educated guesswork. Thus, drawing from background knowledge and constrained by correlational empirical evidence, the use of EFA can at best only be expected to yield a plurality of factorial hypotheses or theories that are thought to be in competition. This contrasts strongly with the unrealistic expectation held by many earlier users of EFA that the method would deliver them strongly justified claims about the one best factorial hypothesis or theory.

How, then, can EFA deal with the specter of UTEE in the context of theory generation? The answer, I think, is that EFA narrows down the space of a potential infinity of candidate theories to a manageable subset by facilitating judgments of initial plausibility. It seems clear enough that scientists often make judgments about the initial plausibility of the explanatory hypotheses and theories that they generate. It is less clear just what this evaluative criterion amounts to (see Whitt, 1992). With ATOM, judgments of the initial plausibility of theories are judgments about the soundness of the abductive arguments employed in generating those theories. I suspect that those who employ EFA as an abductive method of theory generation often make compressed judgments of initial plausibility. Consistent with the view of research problems adopted by ATOM, initial plausibility may be viewed as a constraint-satisfaction problem. Multiple constraints from background knowledge (e.g., the coherence of the proposed theory with relevant and reliable background knowledge), methodology (centrally, the employment of EFA on appropriate methodological grounds; see Fabrigar, Wegener, MacCallum, & Strahan, 1999), and explanatory demands (e.g., the ability of factorial theories to explain the relevant facts in an appropriate manner) combine to provide a composite judgment of a theory's initial plausibility.

By conferring judgments of initial plausibility on the theories it spawns, EFA deems them worthy of further pursuit, whereupon it remains for the factorial theories to be further developed and evaluated, perhaps through the use of confirmatory factor analysis. I should emphasize here that using EFA to facilitate judgments about the initial plausibility of hypotheses will still leave the domains being investigated in a state of considerable theoretical underdetermination. I will also stress that the resulting plurality of competing theories is entirely to be expected and should not be thought of as an undesirable consequence of employing EFA. To the contrary, it is essential for the growth of scientific knowledge that we promote theoretical pluralism. The reason for this rests with our makeup as cognizers: we begin in ignorance, so to speak, and have at our disposal limited sensory equipment. However, we are able to develop a rich imagination and considerable powers of criticism.

These four features operate such that the only means available to us for advancing knowledge is to construct and evaluate theories through their constant critical interplay. In this way, the strategy of theoretical pluralism is forced on us (Hooker, 1987). Thus it is through the simultaneous pursuit of multiple theories with the intent of eventually adjudicating between a reduced subset of them that one arrives at judgments of best theory.

I have suggested that factor indeterminacy is a special case of the pervasive problem of UTEE. I have also argued that if we adopt realistic expectations about what EFA can deliver as a method of theory generation, and also grant that the method contributes to the needed strategy of theoretical pluralism, then we may reasonably conclude that EFA satisfactorily meets this particular challenge of indeterminacy.

3.5 Exploratory Factor Analysis and Confirmatory Factor Analysis

Now that I have argued that EFA is a method that facilitates the abductive generation of rudimentary explanatory theories, it remains to consider what implications this view of EFA has for the conduct of EFA research, including its relation to the more frequently used confirmatory factor analysis (CFA).

The abductive view of EFA does highlight and stress the importance of some features of its best use, and I will mention four of these. First, it should now be clear that an abductive interpretation of EFA reinforces the view that it is best regarded as a latent variable method, thus distancing it from the data reduction method of principal components analysis. From this, it obviously follows that EFA should always be used in preference to principal components analysis when the underlying common causal structure of a domain is being investigated.

Second, strictly speaking, the abductive interpretation of EFA also acknowledges the twin roles of the method of searching for inductive generalizations, and their explanations. As ATOM emphasizes, these research goals are different, but they are both important. To repeat, it is because the detection of phenomena requires the researcher to reason inductively to empirical regularities that the abductive use of EFA insists on initially securing the invariance of factors across different populations. And it is because the inductive regularities require explanation that one then abductively postulates factorial hypotheses about common causes.

Third, as noted earlier, the abductive view of EFA emphasizes the importance of background knowledge in EFA research. In this regard, the initial variable selection process, so rightly emphasized by Thurstone (1947) and Cattell (1978), is sufficiently important that it should be considered as part of the first step in carrying out an EFA study. For instance, in selecting the variables for his factor analytic studies of personality, Cattell was at pains to formulate and follow principles of representative sampling from a broad formulation of the domain in question. Further, the importance of background knowledge in making abductive inferences to underlying factors should not be overlooked. In this regard, the modified Peircean depiction of abductive inference presented earlier explicitly acknowledged some of the manifold ways in which such inference depends on background knowledge. It is an important truism that the factorial hypotheses generated through abductive inference are not created *ex nihilo* but come from the extant theoretical framework and knowledge of the factor analytic researcher. For most of our EFA theorizing, this source is a mix of our common sense and scientific psychological knowledge.

Finally, and relatedly, it should be made clear that acknowledging the importance of background knowledge in abductive EFA does not provide good grounds for adopting a general strategy where one discards EFA, formulates theories *a priori*, and uses factor analysis only in its confirmatory mode. This holds even though when using EFA one anticipates possible common factors to select sufficient indicator variables to allow one to overdetermine those factors. EFA has a legitimate place in factor analytic research because it helpfully contributes to theory generation in at least three ways: it contributes to detection of the empirical phenomena that motivate the need for generating factorial hypotheses; it serves to winnow out a lot of theoretically possible hypotheses at the hypothesis generation stage of inquiry; and it helps to present factorial hypotheses in a form suitable for subsequent testing by CFA.

This last remark, which supports the idea that abductive EFA plays a useful role in factor analytic research, raises the question of how EFA relates to CFA. In contrast to popular versions of the classical inductivist view of science that inductive method can generate secure knowledge claims, using EFA as an abductive method of theory generation can only furnish researchers with a weak logic of discovery that gives them educated guesses about underlying causal factors. For this reason, researchers who use EFA to generate theories need to supplement their generative assessments of the initial plausibility of those theories with additional consequentialist justification in the form of CFA testing or some alternative approach to theory appraisal.

In stressing the need for the additional evaluation of theories that are obtained through EFA, I am not implying that researchers should always or even standardly employ classical EFA and follow it with CFA. CFA is just one of a number of options with which researchers might provide a justification of factorial hypotheses. As an alternative, one might, for example, adopt Rozeboom's nonclassical form of EFA as a method to generate a number of models that are equivalent with respect to their simple structure by using his versatile Hyball program (Rozeboom, 1991a, 1991b) before going on to adjudicate between these models by employing CFA. Another legitimate strategy might involve formulating a causal model using EFA and following it with a procedure like the one defended by Mulaik and Millsap (2000), which undertakes a nested sequence of steps designed to test various aspects of a structural equation model.

A further possibility, which I do not think has been explored in the factor analytic literature, would be to follow up on the preliminary acceptance of rudimentary theories spawned by EFA by developing a number of factorial theories through whatever modeling procedures seem appropriate, and then submitting those theories to a non-factor-analytic form of theory appraisal. For example, it would be quite possible for competing research programs to develop theories given to them through EFA and then submit those theories to comparative appraisal in respect of their explanatory coherence. Thagard's (1992) theory of explanatory coherence, which I consider in chapter 5, is an integrated multicriterial method of theory appraisal that accepts as better those explanatory theories that have greater explanatory breadth, are simpler than their rivals, and are analogous to theories that have themselves been successful. This strategy of using EFA to abductively generate explanatory theories, and then employing the theory of explanatory coherence in subsequent appraisals of these explanatory theories, is abductive both fore and aft. As such, it fits nicely within the framework of ATOM.

Finally, I should say that there are a number of methods for abductively generating hypotheses and theories in psychology, EFA being but one. Grounded theory method (Strauss, 1987), for example, can generate theories that explain the qualitative data patterns from which they are derived (see chapter 6). Also, Howard Gardner's (1983) theory of multiple intelligences was generated using a "subjective," nonstatistical factor analysis. Furthermore, it is plausible to suggest that structural equation modelers sometimes abductively generate theories by non-factor-analytic means before submitting them to CFA scrutiny. As with factor analytic abduction, this could only be done by exploiting our naturally given cognitive abilities to abductively generate explanatory hypotheses and theories.

In this chapter, I have been concerned to argue that EFA has a legitimate and important role as a method of theory generation, and EFA and CFA should be viewed as complementary, not competing, methods of common factor analysis. However, a number of factor analytic methodologists have expressed views that discourage such an outlook. For example, Gorsuch (1983), in his well-known book on factor analysis, expresses a view about the relative importance of exploratory and confirmatory factor analysis that seems to be quite widely held today: "The space and time given to [EFA] is a function of the complexity of resolving its problems, not of its theoretical importance. On the contrary, confirmatory factor analysis is the more theoretically important—and should be the much more widely used—of the two major factor analytic approaches" (134).

Although Gorsuch (1983) makes his claim in emphatic terms, he provides no justification for it. There are, I think, at least two reasons that can be given for his conclusion. However, I do not think they add up to a convincing justification. First, there is a widespread belief that the essence of scientific research is to be found in the prevailing hypothetico-deductive conception of scientific method with its emphasis on theory testing for predictive success. However, this belief is difficult to defend, given that there are many other important phases of scientific inquiry that together demand most of the researcher's methodological time. As ATOM makes clear, these additional phases embrace the detection of empirical phenomena, and the generation, development, and full comparative appraisal of theories. Viewed in this light, theory testing is just one, albeit important, part of scientific method. Given that science is as much concerned with theory generation as it is with theory testing, and acknowledging that EFA is a useful abductive method of theory generation, EFA deserves to be regarded as one important method in the theory constructor's tool kit.

Moreover, both hypothetico-deductive orthodoxy and a good deal of CFA practice today need confirmational rehabilitation. Both suffer from the tendency to take theory evaluation as a noncomparative undertaking in which theories are assessed with respect to the empirical evidence, but not in relation to alternative theories. I suggested earlier that the hypothetico-deductive method can be repaired in this respect. Additionally, some CFA methodologists (e.g., Kaplan, 2000) have sensibly expressed the need to compare theories or models when assessing them with respect to their goodness-of-fit to the empirical evidence. It is here that the problem of UTEE arises for CFA, because associated goodness-of-fit indices sometimes fail to adjudicate between two or more competing factor analytic models. In these cases, CFA has to broaden its announced goal of testing for empirical adequacy through goodness-of-fit tests. This can be achieved in part by obtaining fit statistics weighted by parsimony indices, and more fully by invoking a number of additional superempirical criteria of theory goodness to supplement goodness-of-fit judgments.

I should emphasize that using goodness-of-fit is a minimum criterion of empirical adequacy (Rodgers & Rowe, 2002) and alone provides insufficient grounds for assessing the credibility of competing theories. The goodness-of-fit empirical adequacy of theories can be strengthened by also ascertaining their predictive worth. Hypothetico-deductive testing is often assumed, or recommended, in this regard, but this confirmational strategy faces a number of difficulties well known to philosophers of science. Of particular relevance here is that standard hypothetico-deductive confirmation founders on the problem of UTEE. This shortcoming brings us back to the recommendation advanced earlier that criteria of empirical adequacy need to be supplemented by the so-called superempirical or complementary virtues of explanatory power, fertility, and simplicity (McMullin, 1983). Virtues such as these reduce the gap between theory and empirical evidence, but they do not close it. This is because scientists do not strongly agree on the criteria that should be employed in theory evaluation. Moreover, even when scientists do agree on the evaluative criteria to be used, they will sometimes differ in the relative weight they assign to them. Nevertheless, if we use a composite of empirical and theoretical criteria, the problem of UTEE becomes manageable, though theory evaluation will seldom be a determinate exercise. To meet the challenge of UTEE, CFA, along with EFA, needs to supplement its judgments of empirical adequacy by appealing to the theoretical virtues.

A second reason for downplaying the importance of EFA is the supposition that although EFA has a role in generating knowledge claims, it does not have a role in evaluating them. Rather, full evaluative responsibility is assigned to CFA embedded within a hypothetico-deductive framework. However, as claimed earlier, the use of EFA as an abductive method of theory generation enables us to judge the initial plausibility of the hypotheses it spawns. Positive judgments of initial plausibility are stamps of epistemic approval that signal that factorial hypotheses have sufficient merit to warrant further investigation. Researchers assess initial plausibility to gauge whether hypotheses are worth pursuing, but such assessments do not provide sufficient warrant for treating hypotheses as credentialed knowledge claims. Those who recommend that the hypotheses thrown up by EFA should be tested subsequently with confirmatory factor analysis are right to stress the need for their subsequent justification. However, it is important to appreciate that EFA provides a provisional generative justification for the hypotheses it produces.

3.6 Summary and Conclusion

In examining the methodological foundations of EFA, I have said many things about the nature of this method. It will therefore be useful to bring together the main points in the form of an extended summary and a brief conclusion.

In summary:

1. The main goal of EFA is to generate rudimentary explanatory theories to explain robust covariational data patterns. As a preliminary to this goal, EFA functions as a data analytic method that contributes to the detection of empirical regularities.
2. The inferential move from manifest to latent variables in EFA is abductive in nature. The particular form of abductive inference typically involved is existential abduction. Existential abductions postulate the existence of objects or attributes, but they do not specify their natures.
3. EFA's use of abductive reasoning is facilitated by its employment of the principle of the common cause, which restricts factor analytic inferences to correlated effects and their common causes. This principle lies at the inferential heart of EFA.
4. EFA has a modest, albeit important, role in theory generation. It is a serviceable generator of elementary plausible theory about the common causes of correlated variables.
5. The abductive logic of EFA enables the method to confer a generative justification on the theories it produces. This form of justification involves judgments that the theories are the result of sound abductive reasoning and have sufficient initial plausibility to warrant further investigation.
6. Theories generated by EFA have the status of dispositional theories. The latent variables postulated by such theories can be genuine existents, though these theories say little, if anything, about their nature.
7. Despite their elementary nature, dispositional theories afforded by EFA do have genuine, although modest, explanatory power. This power resides in both their existential or explanatory depth and their consilience or explanatory breadth.

8. EFA is able to satisfactorily confront the problem of factor indeterminacy in theory generation by screening candidate factorial theories for their initial plausibility in an environment where theoretical pluralism is to be expected.

9. To satisfactorily meet the problem of factor indeterminacy, CFA research should embrace superempirical criteria in addition to both the goodness-of-fit and predictive criteria of empirical adequacy.

10. Because EFA and CFA tend to serve different methodological functions in multivariate research—theory generation for the one, theory testing for the other—they are best viewed as complementary rather than competing methods. It will sometimes be advantageous to employ the two common factor analytic methods in tandem.

11. Nevertheless, theories about common causes can be generated abductively without appeal to EFA, whereas theories generated by EFA may be tested by using methods other than CFA.

12. ATOM provides a useful framework within which to locate EFA. There EFA can function as a method of theory generation in domains with a common causal structure.

13. CFA can contribute to the goal of empirical adequacy in the subsequent hypothetico-deductive appraisal of common causal theories.

Although EFA has frequently been employed in psychological research, the extant methodological literature on factor analysis insufficiently acknowledges the explanatory and ontological import of the method's inferential nature. Arguably, abduction is science's chief form of creative reasoning, and the principle of the common cause is a maxim of scientific inference with important application in research. By bringing these two related elements into its fold, EFA is ensured an important, albeit circumscribed, role in constructing explanatory theories in psychology and other sciences. In this role, EFA can serve as a valuable precursor to CFA. I believe that factor analytic research would benefit considerably by returning to its methodological origins and embracing EFA as an important method for generating structural models about common causes.

As noted in the outline of ATOM provided in chapter 1, the rudimentary theories given to us by existential abduction by methods such as EFA need concerted development. In ATOM, this is undertaken by employing a strategy of analogical modeling. This strategy is the main focus of the next chapter.

4 Theory Development: Analogical Modeling

The process by which the nature [of the causal mechanism] is first ascribed in developing an explanation is psychologically an exercise of the imagination and philosophically an analogy. . . . The creative task is to present a plausible analogue of the mechanism which is really producing the phenomenon.

—Rom Harré (1976, 21)

4.1 Introduction

This chapter focuses primarily on the development of scientific theories. In particular, I aim to show that ATOM develops its theories by adopting a strategy of analogical modeling. However, before considering this strategy, I will provide a brief and selective overview of the nature and place of models in science. This overview should form a useful backdrop to the subsequent discussion of analogical modeling.

For the last hundred or so years, the role of models in science has been controversial. One view, held by prominent students of science before the twenty-first century, was that models were dispensable heuristic aids to formulating and understanding scientific theories—perhaps even props for poor thinkers. For example, the French physicist and philosopher Pierre Duhem (1954) was strongly skeptical of the value of building mechanical models to understand physical processes, and he famously derided the English scientists of his time for engaging in this practice.

According to the two major early twentieth-century philosophies of science, logical positivism and logical empiricism (Feigl, 1956), models played no important role in the conduct of scientific research. In the 1950s and 1960s, critics of logical empiricism pointed out that its view of models did not provide for the role that models play in the development of theories. Although not much influenced by logical empiricism and its critics, psychology itself has historically given limited explicit attention to models, although it has given increased attention to mathematical and statistical modeling in the last few decades (Rodgers, 2010).

This negative view of the cognitive value of models in science contrasts with the view held by many methodologists today that models are an essential part of the development of theories and are important elsewhere in science as well. Contemporary studies of scientific practice make clear that models play a genuine and indispensable cognitive role in science. Many scientists and philosophers subscribe to the view that reasoning in science is to a large extent model-based reasoning. Ronald Giere (1999), for example, goes so far as to say that science "is models almost all the way up and models almost all the way down" (56). Although a number of different sorts of model play important roles in scientific research, I think that Giere overstates the influence of models in science. For good reason, science draws on many disparate investigative strategies that have little or nothing to do with models. For example, some of the strategies for detecting empirical phenomena dealt with in chapter 2 do not use models.

Psychology's commitment to the hypothetico-deductive method, and to a lesser extent the inductive method, has helped discourage psychologists from using models for the purpose of theory development. The orthodox account of the hypothetico-deductive method assumes that hypotheses and theories emerge fully formed and ready for immediate testing.[1] For its part, traditional inductive method focuses first on the discovery of empirical generalizations, and then on fashioning theories that are organized summaries of their constituent empirical generalizations. Such an instrumentalist conception of theories discourages the development of deep explanations, and with it a need for modeling latent causes. This is the perspective on theory and method adopted by radical behaviorists.

In contrast to these two theories of scientific method, ATOM provides explicitly for the development of explanatory theories. The theories it generates through existential abduction are only dispositional in nature and require considerable elaboration before they are systematically evaluated against rival theories with respect to their explanatory goodness. As noted in chapter 1, ATOM recommends that this be done by building analogical models of the causes posited by existential abduction to obtain knowledge of the mechanisms that comprise those causes.

4.2 Types of Models

Given that just about anything can be a model of something for someone, we have an enormous diversity of models in science. This diversity includes, but is not limited to, scale models, data models, phenomenological models, theoretical models, analog models, iconic models, and mathematical models. Science uses these different types of model for different purposes. For example, so-called *iconic models* are constructed to provide a good resemblance to the object or property being modeled, whereas mathematical models offer an abstract symbolic representation of the object or property of interest.

Max Wartofsky (1979) has referred to the many senses of the word *model* that stem from this bewildering variety as the "model muddle." Philosophers such as Max Black (1962), Peter Achinstein (1968), and Rom Harré (1970) have provided different taxonomies that impose some order on the variety of available model types. However, it seems unlikely that the diversity of models in science will be captured by a unified taxonomy. Moreover, given that different types of models serve different research ends, we should refrain from thinking that one approach to modeling is inherently superior to another.

I confine my initial discussion to four types of models that are used in science: scale models, theoretical models, mathematical models, and data models. The fifth type of model (the analogical model, in which an unfamiliar domain is modeled by analogy to a familiar source) is discussed at length in the second half of the chapter.

4.2.1 Scale Models

Some models are physical structures that can represent or potentially represent things in the world. Physically constructed scale models are a good example. Scale models belong to a class of iconic models because they literally depict the features of interest in the original. As their name suggests, scale models involve a change of scale. They are always models of something, and they typically scale down selected properties of the objects they represent. Thus a model airplane stands as a miniaturized representation of a real airplane. However, a scale model can also be a magnified representation of an object, such as a small insect.

Although scale models are constructed to provide a good resemblance to the object or property being modeled, they represent only selected relevant features of the object. Thus a model airplane will almost always represent the fuselage and wings of the real airplane being modeled, but it will seldom represent the interior of the aircraft. Scale models are usually built to represent the properties of interest in the original object in an accessible and manipulable form. By scaling and idealizing a source that is complex in its natural form, scientists can study processes in a manageable way. A scale model of an aircraft prototype, for example, may be built to test its basic aerodynamic features in a wind tunnel.

However, not all iconic models are scale models, as James Watson and Francis Crick's physical model of the helical structure of the DNA molecule demonstrates. By idealizing and scaling data to some manageable form, graphs too can be considered scale models of the processes and distributions that they represent.

4.2.2 Theoretical Models

The important class of models known as theoretical models abounds in science. Unlike scale models, theoretical models are constructed and described by the scientist's imagination in that they are not constructed as physical objects. Further, unlike mathematical and analogical models, the properties of theoretical models are often better known than the subject matter that is being modeled. This is clearly the case when scientists attempt to model unknown theoretical entities. For example, the properties of latent variable models, such as the common factor model referred to in chapter 3, are better known to the investigator than the latent attributes represented by those models.

A theoretical model of an object, real or imagined, comprises a set of hypotheses about that object. The Watson-Crick model of the DNA molecule and Markov models of human and animal learning are two examples of the innumerable theoretical models to be found in science. Theoretical models typically describe an object by ascribing to it an inner mechanism or structure. This mechanism is frequently invoked to explain the behavior of the object. Theoretical models are acknowledged for their simplifying approximation to the object being modeled, and they are often small-scale theories with a limited scope of application. However, they can often be combined with other theoretical models to provide a more comprehensive understanding of the object of study. For example, the Rutherford-Bohr model of the atom is a modification of the earlier Rutherford model from the perspective of quantum physics.

4.2.3 Mathematical Models

In the behavioral sciences, models are sometimes expressed in terms of mathematical equations. For example, factor analysis is commonly understood as a mathematical model of the relations between manifest and latent variables, where each manifest variable is regarded as a linear function of a common set of latent variables along with a latent variable that is unique to the manifest variable. It is important to emphasize that a statistical model and its interpretation are distinct entities. The basic equation for linear factor analysis, for example, is to be distinguished from the various substantive factorial theories that its use has helped bring about.[2] Sometimes in the physical sciences, a theory formulated in mathematical terms at the outset cannot subsequently be interpreted as a substantive and comprehensible source model. Many physicists have understood the so-called Copenhagen formulation of quantum mechanics to be this sort of model because its content comprises mathematical probabilities that do not describe an objective reality.

Mathematical models offer an abstract symbolic representation of their domains of interest. These models are often regarded as formalized theories in which the system modeled is projected onto the abstract domain of sets and functions, which can be manipulated in terms of numerical reasoning, typically with the help of a computer.

In psychology, the large majority of theories are constructed in a qualitative manner, and most of them remain so thereafter. To a limited extent, psychologists strive to formalize these theories in mathematical terms to provide them with a more rigorous formulation. For example, a number of theories in psychology characterize relationships between psychological constructs in terms of multiplicative functions.

4.3 Data, Models, and Theories

4.3.1 Data Models

In the early 1960s, Patrick Suppes suggested that science employs a hierarchy of models that range from experimental experience to theory (Suppes, 1962). He claimed that theoretical models, which are high on the hierarchy, are not compared directly with empirical data, which are low on the hierarchy. Rather, they are compared with models of the data, which are higher than data on the hierarchy. This insight anticipated a central idea of chapter 2, that phenomena, not data, should be taken to be the proper objects of typical scientific explanations.

The process of phenomena detection arises because scientific data on their own are intractable. Data are often rich, complex, and messy, and because of these characteristics, they cannot be explained. Their intractability is overcome by reducing them to simpler and more manageable forms. In this way, scientists rework data into models of data. As shown in chapter 2, statistical methods play a prominent role in this regard, facilitating operations having to do with assessing the quality of the data, the patterns they contain, and the generalizations to which they give rise. Because of their tractability, models of the data can be explained and used as evidence for or against theoretical models. For this reason, they are of considerable importance to science.

It is fair to say that in both their science education and research practices, psychological researchers have been more concerned with data models than other kinds of models in science.

4.3.2 Models and Theories

The relationship between models and theories is difficult to draw, particularly given that they can both be conceptualized in various ways. Some methodologists have suggested that theories are intended as true descriptions of the real world, whereas models need not be about the world and therefore need not be true. Others have drawn the distinction by claiming that theories are more abstract and general than models. For example, evolutionary psychological theory can be taken as a prototype of the more specific models it engenders, such as those of differential parental investment and the evolution of brain size. Relatedly, Giere (1988) has argued that a scientific theory is best understood as comprising a family of models, along with a number of theoretical hypotheses that link the models with things in the world.

Yet another characterization of models takes them to be largely independent of theories. In arguing that models are "autonomous agents" that mediate between theories and phenomena, Margaret Morrison and Mary Morgan (1999) contend that they are not fully derived from either theory or data. Instead models are technologies that allow one to connect abstract theories with empirical phenomena. Some have suggested that the idea of models as mediators does not apply to the behavioral and biological sciences because these sciences exhibit no appreciable gap between fundamental theory and phenomena in which models can mediate. However, this is an empirical claim, and the extent to which it holds is yet to be determined.

The position I adopt in this book is that modeling in science is basically a strategy of indirectly representing the world. Models are a type of theory that indirectly represents the world, whereas many other types of theory represent the world more directly (see Weisberg, 2007). To understand the world, the modeler first constructs a model as a hypothetical system. He or she then endeavors to determine the resemblance relations between the hypothetical system and the part of the world he or she is trying to understand. This general description of model-based science as a two-phase process employs the strategy of analogical modeling, to be discussed shortly.

4.4 The Functions of Models

4.4.1 Representation

Today most scientists and philosophers of science take a model to be a representational device that represents the target system that it models. A model can usefully be seen to represent its target in two ways (e.g., Mäki, 2011): first, in terms of its resemblance to the target in certain respects; and second, as a representative of a target system in the sense that it is a surrogate system that stands for, and is examined in place of, its target. When we evaluate the worth of a model, we need to consider both of these aspects of representation together. It is in good part as a consequence of being able to represent the world that models can be employed for a variety of purposes, including systematization, explanation, prediction, control, calculation, and derivation.

However, unlike model-free or "plain" scientific theories, models are generally not thought to be the sort of representational devices that can be true or false. Instead it is suggested that we think of models as having a kind of similarity relationship with the object that is being modeled, where the similarity can take different forms. With analogical models, for example, the similarity relationship is one of analogy, a relationship to be described shortly. In addition, it is sometimes said that models themselves are not linguistic entities and therefore cannot be the bearers of truth (e.g., Giere, 1988). Against this claim, many truth theorists maintain that language is not the only type of truth bearer, and as a consequence, models as nonlinguistic entities can also perform this role. It is also said that because models idealize and abstract away from reality, they do not tell the whole story and so must be false. However, models can be true in two senses (Mäki, 2011): they can be approximately true, depending on their degree of similarity to the target; and they can be partially true in virtue of one or more of their parts being true.

Regarding the falsity of models, note that science often adopts a deliberate strategy of adopting false models as a means by which we can obtain truer models. William Wimsatt (2007) has argued that this is done by localizing errors in models to identify and modify their problematic parts. One might add that scientists can also localize truths in models to help identify and correct errors in other parts of the models.

4.4.2 Abstraction and Idealization

Scientists often study systems that are highly complex. This complexity, combined with limited knowledge about the domains under study, as well as scientists' cognitive limitations, regularly forces them to adopt simplifying strategies to make their research problems tractable. Modeling is one way of simplifying the depiction of complex domains. The simplification is usually achieved through two related processes: abstraction and idealization. Abstraction involves deliberately eliminating properties of the target that are not considered essential to understanding the target. This can be achieved in various ways. For example, one can ignore the properties, though they continue to exist, by eliminating them in controlled experiments or by setting the values of unwanted variables to zero in simulations.

By contrast, idealization involves transforming a property in a system into a related property that possesses desirable features introduced by the modeler. Taking a spheroid object to be spherical, representing a curvilinear relation in linear form, and assuming that a human agent is perfectly rational are all examples of idealization in model building. Although no strong consensus exists in the philosophy of science about how the processes of abstraction and idealization should be understood, and although the terms *abstraction* and *idealization* are sometimes used interchangeably, they clearly refer to different processes. Each can take place without the other, and in particular cases, idealization can in fact involve complexification rather than simplification, for example, when one extends a model to another domain. Jones (2005) provides a helpful systematic treatment of the two processes in which idealizations are construed as deliberate misrepresentations and abstractions as mere omissions. Models, then, are almost always simplified representations of their objects of study in virtue of often having one or both of these features.

In broad terms, the foregoing remarks about models should be seen as consistent with the realist view of science sketched in chapter 1. Scale models have an obvious realist ring to them, because they are clearly direct representations of things that exist in the world. As noted in chapter 3, the latent variables of mathematical models (such as factor analysis) are best understood as genuine theoretical existents. Theoretical models are surrogate systems that refer to theoretical entities. Data models provide tractable empirical evidence for or against theories and theoretical models. And models that are understood as representational devices will have truth values; they realistically aspire to present approximate and partial truths understood as truth nominations in the sense presented in chapter 1. This is the case for analogical models in ATOM.

4.5 Modeling in ATOM

Theorizing about hidden causal entities, properties, and processes is undoubtedly the most frequent type of theorizing in science. We saw in chapter 3 that the nascent theories bequeathed to us by using the method of exploratory factor analysis refer to the existence of hidden causes. Conclusions about such causes are obtained by using an existential abductive reasoning process. However, existential abduction is unable to provide us with an informative characterization of the nature of those causes. Instead, theories given to us by existential abduction have the status of dispositional theories that provide us with oblique characterizations of the causes in terms of their presumed effects under specified conditions. To recall the example from chapter 3, the latent property of the brittleness of glass is described in terms of the relevant events of striking and breaking. Of course, this says nothing about the nature of brittleness, but diagnosing its presence in particular cases is often an important first step in obtaining that knowledge.

Sometimes psychologists are prepared to accept a dispositional construal of the hidden causes that interest them and concentrate their efforts on figuring out how those causes relate to one another and to more empirical matters of fact. For example, structural equation modeling, now a popular research practice in psychology, focuses on providing knowledge of variables assembled in causal networks. As such, it does not so much encourage the development of detailed knowledge of the nature of the latent variables it deals with as specify the range and order of causal relations into which such variables enter.

Although it is acknowledged that science needs to employ a variety of different modeling strategies, ATOM adopts the strategy of using analogical models to help develop its explanatory theories. Often psychologists want to move beyond the rudimentary nature of the dispositional characterization of causes and elaborate on their nature. ATOM's strategy of analogical modeling enables them to do so because it provides more detailed knowledge of causes by enumerating the components and operations of their mechanisms.

Recently, philosophers of science have given considerable attention to the role played by explanations in the life sciences that appeal to causal mechanisms (e.g., Bechtel & Abrahamsen, 2005; Machamer, Darden, & Craver, 2000). Mechanistic explanations, which explain empirical phenomena in terms of the operation of causal mechanisms, are fashioned in psychology with varying degrees of success. They vary from speculative conjectures, through plausible models that are consistent with known constraints, to quite good descriptions of how mechanisms work in reality. I think that analogical modeling is best suited to giving psychologists plausible models of mechanisms. To forestall a possible objection, I should point out that mechanistic explanations do not have to be cashed out in mechanical terms.

Table 4.1 depicts analogical modeling in relation to other parts of ATOM. It shows the objects of investigation of ATOM; the methodological phases of ATOM, with their associated reasoning processes; and the different types of knowledge claim that ATOM helps produce. The content of the table can be assembled in the form of an anticipatory summary as follows. The causes that produce the phenomena are

Table 4.1
The place of analogical modeling in the abductive theory of method

Objects of nature	Phases of ATOM	Products of ATOM
Phenomena (produced by) ↓	Phenomena detection (via enumerative induction)	Phenomena claims (explained by) ↓
Causal entities (represented by) ↓	Theory generation (via existential abduction)	Rudimentary explanatory theories (developed by) ↓
Analogical models of causal mechanisms (leading to) ↓	Theory development (via analogical abduction)	Analogical models (resulting in) ↓
Developed analogical models	Theory appraisal (via inference to the best explanation)	Developed explanatory theories

diagnosed by way of existential abduction from claims about the phenomena. The mechanisms of the causes are specified by building analogical models. As intimated in chapter 1, these are suggested by the antecedently known sources of the models. The specification is achieved by reasoning by analogy from the sources to their targets. Because the source models are human artifacts, they are classified as objects of nature. They are subject to investigation as surrogate systems that represent the causal mechanisms in nature. Analogical modeling in ATOM is a strategy that increases the content of theories that are explanatory in nature. Being explanatory in nature, the analogical reasoning takes the form of analogical abduction. Judgments of the initial plausibility of the causal entities in the phase of theory generation are strengthened by further judgments of plausibility of the analogical models. When the model theories are well developed, they are appraised further by a process of inference to the best explanation.

4.6 Analogical Modeling

The use of analogies to explain events in science is somewhat controversial. For example, the logical empiricist Carl Hempel (1965) maintained that although analogical models may have heuristic value in suggesting explanations, they do no epistemic work in furnishing genuine explanations and can therefore be dispensed with. However, given the weight of many historical case studies, this view has fallen into disfavor.

The idea that analogical models are important in the development of scientific theories can be traced back to the physicist and philosopher of science N. R. Campbell (1920), who insisted that analogies are not mere aids but an essential part of theories. Since that time, a number of philosophers of science have endorsed the value of analogical modeling in scientific theory construction (e.g., Abrantes, 1999; Harré, 1988; Hesse, 1966). The epigraph from Harré at the beginning of this chapter clearly emphasizes the importance of creatively developing explanatory theories through analogical reasoning about the nature of the causal mechanisms to which they refer.

Despite Campbell's claim for the ubiquity of models in theories, scientific explanations do not always use analogies. However, their role in theory development within ATOM is of central importance. The need for analogical modeling within ATOM stems from two features of its characterization of theory generation. First, as with exploratory factor analysis, the abductive generation of theories initially takes the form of existential abduction, through which the existence of theoretical entities is postulated. Therefore an appropriate research strategy is required to learn about the nature of these hidden entities. Analogical modeling is an appropriate strategy for doing the required elaborative work. Second, recall that the postulation of theoretical entities through existential abduction confers an assessment of initial plausibility on those postulations. For claims about those latent entities to have the status of genuine knowledge, further evaluative work has to be done. The construction of appropriate analogical models serves to assess the plausibility of the expanded understanding they afford, as well as to expand our understanding of those entities.

For ATOM, increasing the knowledge of the nature of its theories' causal mechanisms by analogical modeling is achieved by using the pragmatic strategy of conceiving of these unknown mechanisms in terms of what is already familiar and well understood. Well-known examples of models that have resulted from using this strategy are the model of chromosomal inheritance, based on an analogy with a string of beads; the model of natural selection, based on an analogy with artificial selection; and computational models of the mind, based on analogies with the computer.

Although I have used the term *model*, nothing is a model as such. A model is a relational complex. Thus, to understand the nature of analogical modeling, it is necessary to distinguish between a model, the source of the model, and the subject of the model (Harré, 1976; Hesse, 1966). A model is modeled on a source, and it is a model of, or for, a subject. From the known nature and behavior of the source, one builds an analogical model of the unknown subject or causal mechanism. In the biological example just mentioned, Darwin fashioned his model of the subject of natural selection by reasoning by analogy from the source of the known nature and behavior of the process of artificial selection. Used in this way, analogical models play an important creative role in theory development.

However, this creative role requires the source from which the model is drawn to be different from the subject that is modeled. For example, the modern computer is a well-known source for modeling human cognition, but the two are different. Because the brain is made of protoplasm, and the computer is made of silicon, our cognitive apparatus is not generally thought to be a real computer. Models in which the source and the subject differ are sometimes called *paramorphs*. This is a requirement for the analogical modeling of real and imagined processes, which is a focus of ATOM. By contrast, models in which the source and the subject are the same are sometimes called *homeomorphs* (Harré, 1970). For example, a toy airplane can be a homeomorphic model of a real aircraft.

The paramorph can be an iconic representation of real or imagined things. Iconic representation combines elements of visualizable and propositional information in a picture-statement complex that can ultimately be expressed in sentences. The idea of the field of potential in physics is a good example. It can be represented graphically to show how the ideas of field and potential are combined. At the same time, the graphical information, and information not contained in the graph, can be represented in sentential form.

Iconic paramorphs feature centrally in the creative process of developing theories through analogical modeling. These models are constructed as representations of reality, real or imagined. In ATOM, they stand in for the hypothesized causal mechanisms. Although they are representations, iconic models are themselves things, structures, or processes that correspond in some way to things, structures, or processes that are the subjects of modeling. They are therefore the sorts of things that sentences can be about (Harré, 1976). Here we are reminded that scientific theories that are models represent the world less directly than theories that are not models.[3]

In addition to developing nascent theories, the strategy of analogical modeling also serves to assess their plausibility. In evaluating the aptness of an analogical model, one must assess the analogy between its source and subject, and for this one needs to consider the analogy's structure. The structure of an analogy comprises a positive analogy in which the source and subject are alike in some respects, a negative analogy in which the source and subject are unlike in some respects, and a neutral analogy in which the source and subject are alike and unlike in ways that are as yet unknown. The neutral analogy is irrelevant for purposes of analogical modeling. Because we are essentially ignorant of the nature of the hypothetical mechanism of the subject apart from our knowledge of the source of the model, we are unable to specify any neutral analogy between the model and the mechanism being modeled. Thus, in considering the plausibility of an analogical model, one considers the balance of the positive and negative analogies (Hesse, 1966). This is where the relevance of the source for the model is spelled out. As we will see shortly, the analogical reasoning that scientists employ is informal and based on plausibility arguments.

In the next section, I discuss Darwin's use of analogical modeling in developing his theory of natural selection. In section 4.8, I present the dramaturgical model of human social interaction as an example of analogical modeling in psychology.

4.7 Analogical Abduction

Reasoning by analogy is an important form of inference, but it is difficult to characterize precisely. Historically, philosophers have often reconstructed analogical arguments as enumerative or simple inductions of a special form (e.g., Copi & Cohen, 1990; Hesse, 1966). Because analogical reasoning results in new knowledge claims, it is ampliative, a feature it shares with inductive reasoning. However, unlike arguments based on inductive inference, arguments based on analogy can produce knowledge claims about new kinds of things. Briefly, we may say that an analogy is an argument based on assumed or known parallels or similarities between two or more objects, properties, or events. What is known about one class of entities (the source) is employed to learn more about the other class of entities (the subject). A good analogical argument provides an understanding of the less familiar in terms of the more familiar by discerning that the two are alike in relevant respects, but not in other respects. As already mentioned, for example, psychological research frequently reasons by analogy from the known functioning of computers to the less well-known character of human cognitive processes.

Analogical reasoning is important in science and obviously lies at the inferential heart of analogical modeling. I emphasized in chapter 3 that abduction is a form of scientific reasoning in its own right. As intimated in chapter 1, because the theories fashioned by ATOM are explanatory theories, the use of analogical modeling to develop those theories will necessarily involve combining analogical and abductive forms of reasoning to produce a creative form of reasoning known as *analogical abduction*. Science often seeks to improve the quality of an explanatory theory by appealing to a similar type of explanation that is known and accepted by the scientific community. It is in this way that we can employ analogical reasoning of an abductive kind.

Note, however, that, unlike existential abduction, analogical abduction does not produce a hypothesis about an entirely new entity, property, or process. It is only concerned with the partly new, because it is driven by analogy to concepts that are well understood in the source model. The importance of analogical abduction as a form of creative reasoning in ATOM stems from the fact that it is the means by which knowledge about a theory's causal mechanisms is developed.

The basic structure of the reasoning involved in analogical abduction can be stated in the form of a general argument schema as follows:

Hypothesis H* about property Q was correct in situation S1.

Situation S1 is like situation S2 in relevant respects.

Therefore an analogue of H* might be appropriate in situation S2.

To take a prominent example, Darwin's theory of natural selection made essential use of analogical abduction. The general argument for analogical abduction just given can be rewritten in simplified form for Darwin's case as follows:

The hypothesis of evolution by artificial selection was correct in cases of selective domestic breeding.

Cases of selective domestic breeding are like cases of the natural evolution of species with respect to the selection process.

Therefore, by analogy with the hypothesis of artificial selection, the hypothesis of natural selection might be appropriate in situations where variants are not deliberately selected for.

In formulating his theory of natural selection, Darwin took advantage of the two most important features of analogical abduction: its ability to create, and its ability to justify. In reasoning by analogy, using known facts about artificial selection, Darwin was able to hypothesize the parallel mechanism of natural selection that explained diversity among natural species. At the same time, he was able to appeal to the epistemic worth of his carefully crafted analogy and proclaim the initial plausibility of his hypothesis of natural selection. Numerous creative scientists have been able to exploit the resources of analogical abduction in this manner.

Three things should be said about the structure of analogical reasoning as it is outlined in the argument schema. The first premise of the argument claims factual status for the relevant part of the source model. However, this is not always easy to ascertain and requires close knowledge of the source model. In Darwin's case, nineteenth-century breeding practices were rather controversial, and Darwin had to work hard to forge his analogy (Theunissen, 2012). For example, he had to downplay the importance of the two breeding techniques of crossing of varieties and inbreeding that many breeders thought were essential to obtain new varieties. The second premise of the argument asserts that relevant similarities that enable the transfer of explanations from source to subject have been identified. But this transfer clearly requires some knowledge of the subject of the model and need not be completely unidirectional. For example, evidence suggests that Darwin's developing knowledge of natural selection in nature helped him better understand his knowledge of artificial selection of domestic varieties (Herbert, 1971). The conclusion stated in the argument's third premise is appropriately tempered. To say that the analogy "might be appropriate" is in keeping with the plausibility assessments that the process of analogical modeling gives us. Just as good existential abductions confer a warrant of initial plausibility on the hypotheses they produce, so sound analogical arguments provide the grounds for judging the hypotheses about the mechanisms in question to be initially plausible. It is clear from Darwin's writings that he took the analogy between artificial and natural selection to lend some credence to his theory of natural selection. However, as we will see in the next chapter, Darwin sought further assessments of his theory by employing inference to the best explanation.

4.8 The Dramaturgical Model

An instructive example of an analogical model in psychology is Rom Harré's role-rule model of microsocial interaction, which he developed by explicitly using his own methodology of analogical modeling. As with the Darwin example of analogical modeling just discussed, Harré used the strategy of analogical modeling both to create and to justify his model of microsocial interaction. With the role-rule model, Erving Goffman's (1959) dramaturgical perspective on human action provides the source model for understanding the underlying causal mechanisms involved in the production of ceremonial, argumentative, and other forms of social interaction (Harré, 1979; Harré & Secord, 1972).

The role-rule model can also be presented in accordance with the simple argument schema used in the previous section to display the basic structure of its analogical abductive reasoning:

The theory of dramaturgy provides a correct account of behavior on the theatrical stage.

Behavior on the theatrical stage is like a good deal of human behavior in social life.

Therefore, by analogy with the theory of dramaturgy, much human social behavior might be understood and monitored as acting on life's stage.

Of course, this schema is a bare-bones characterization of the analogical abductive reasoning used in constructing the dramaturgical model. Neither the nature of the analogical reasoning employed nor its justification is properly captured by its schematic representation. As with the inductive reasoning employed in detecting phenomena, a fine-grained depiction of the analogical reasoning involved in constructing the dramaturgical model must be material in nature. That is to say, the relevant limits of the similarity relation between the source and subject of the model are decided with reference to contingent matters of fact that are specific to the case.

The basic idea of the dramaturgical perspective is that we observe and hear a simulacrum of life on the stage, and our knowledge of how this is produced provides us with a guide to the creation of real life. Goffman's dramaturgical perspective provides a detailed analytical account of the roles and rules that human agents follow on life's stage combined with a "watchful consciousness" of the actor, producer, audience, and critic.

As a source model, the dramaturgical model has both positive and negative analogies, for clear similarities and differences exist between the subject domain of real life and the source domain of dramatically staged acts. Regarding similarities, Goffman noted that to be understood as the person he or she is portraying, the actor has to act in a manner that parallels what the audience would expect of that kind of person. Clearly there are differences between stage drama and real life. The differences involve sequences of acts and actions that are at once selective, simplified, and heightened. For example, in comparison with real life, only a limited number of life sequences are followed on the stage, time is compressed, and resolutions are effectively reached (Harré, 1979). The reduction in the number of life sequences and the compression of time are abstractive processes. The use of successful resolutions is an idealized move. In these ways, the modeling strategies of abstraction and idealization are employed to simplify the complex domain of microsocial interaction.

Despite these sorts of differences, there are sufficient likenesses to make the dramaturgical model well worth exploring. Harré has exploited the dramaturgical model to provide a role-rule perspective on social psychological performance that uses a reticulated analytical scheme to further our understanding of microsocial accounts of social interaction in everyday life. As such, it stands as an important and explicit example of analogical modeling in psychology.

4.9 Conclusion

The strategy of analogical modeling is sometimes used in the behavioral sciences to develop theories. This is not surprising, given that many of the hypothesized causes in these sciences are theoretical entities whose natures can be grasped only indirectly using such a modeling strategy.

The methodology of analogical modeling is well developed and provides a useful source of guidance for scientists intent on expanding their knowledge of latent causal mechanisms. Rom Harré's various works on analogy and modeling in science constitute a useful source in this regard. Methodological work that focuses specifically on analogical abduction is less well developed, although it contains broad guidelines for the aspiring analogical modeler. Paul Bartha's (2010) wide-ranging book *By Parallel Reasoning* is a detailed, instructive examination of how to construct and evaluate analogical arguments.[4]

There is little evidence to suggest that the behavioral sciences explicitly incorporate a strategy of analogical modeling into their methodological deliberations and science education practices. The limited methodological attention given to modeling in psychology is largely confined to statistical modeling, broadly construed (e.g., Jaccard, 2013; MacCallum, 2003; Rodgers, 2010). However, given the importance of analogical modeling as a strategy for the expansion of explanatory theories, methodologists in the behavioral sciences should promote it as vigorously as they have promoted structural equation modeling.

Thus far, I have suggested that, for ATOM, the epistemic worth of hypotheses and theories generated by existential abduction is evaluated in terms of their initial plausibility, and these assessments are subsequently augmented by judgments of the appropriateness of the analogies that function as source models for their development. However, with ATOM, well-developed theories are appraised further with respect to a number of additional criteria that are used when making judgments about the best of competing explanatory theories. This is the focus of the next chapter, where we will see that the criterion of analogy, in combination with additional criteria, figures in the further assessment of the plausibility of analogical models.

5 Theory Appraisal: Inference to the Best Explanation

If the fact that a theory provides the best available explanation for some important phenomenon is not a justification for believing that the theory is at least approximately true, then it is hard to see how intellectual inquiry could proceed.
—Richard Boyd (1984, 67)

5.1 Introduction

Contemporary scientific methodology boasts a number of general approaches for evaluating scientific theories. Prominent among these are the hypothetico-deductive method, which evaluates theories in terms of predictive success; Bayesian accounts of confirmation, which assign probabilities to hypotheses using Bayes's theorem; and inference to the best explanation, which accepts theories when they are judged to provide better explanations of the evidence than their rivals do. These are three of the four major theories of scientific method canvassed in chapter 1. Because of its focus on procuring descriptive generalizations, the simple inductive account of scientific method does not seriously address the matter of theory appraisal.

It has been stated repeatedly that the hypothetico-deductive method is by far the most widely used approach to theory appraisal in psychology (see, e.g., Rorer, 1991; Rozeboom, 1997). Despite some urgings (e.g., Edwards, Lindman, & Savage, 1963; Lee & Wagenmakers, 2005; Dienes, 2011), psychologists have been reluctant to use Bayesian statistical methods to test their research hypotheses and theories. They have mostly preferred to use classical statistical significance testing within a hypothetico-deductive framework. Unfortunately, inference to the best explanation has received almost no attention by psychological researchers.

Many scientists in the natural and biological sciences have placed stock in the explanatory standing of theories, with Darwin and Einstein prominent among them (Janssen, 2002). In a well-known passage in the final chapter of *On the Origins of Species*, Darwin declares his confidence in justifying theories by appeal to explanatory considerations:

> It can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified [e.g., the geographical distribution of species, the sterility of hybrid species]. It has recently been objected that this is an unsafe method of arguing. But it is a method used in judging common events of life, and has often been used by the greatest natural philosophers. The undulatory theory of light has thus been arrived at; and the belief of the revolution of the earth on its own axis was until lately supported by hardly any direct evidence. (Darwin, 1958, 452)

It is clear that Darwin set great store by the fact that his theory of natural selection provided a much better explanation of the classes of facts such as those just mentioned than did the rival creationist theory.

In addition, methodologists have been concerned for some time to articulate ways in which we can understand the explanatory worth of theories (e.g., Lipton, 2004; Josephson & Josephson, 1994; Thagard, 1989, 1992). However, although inference to the best explanation (IBE) is used in some sciences and extensively discussed in the philosophy of science, it is seldom heard of in psychology. This is an omission that I believe needs to be put right.

The primary purpose of this chapter is to bring the important idea of IBE to the attention of psychologists while emphasizing that the literature on the topic contains methodological resources that can help researchers evaluate the explanatory worth of their theories (Haig, 2009). I begin by introducing the general idea of explanatory inference. Then I consider a number of different approaches to characterizing IBE; prominent among these is the theory of explanatory coherence , which is the approach to theory appraisal adopted by ATOM. Thereafter I discuss the strengths and limitations of IBE, together with its relationship to other major approaches to theory appraisal and its place in the broader domain of scientific inference. The chapter's penultimate section considers IBE in relationship to psychology, and in the conclusion, I recommend that psychologists use IBE as an appropriate means of evaluating the worth of their explanatory theories.

5.2 Inference to the Best Explanation

In accordance with its name, IBE is based on the idea that much of what we know about the world, in both science and everyday life, is based on considerations of explanatory worth. Scientists often accept theories about the hidden causes of empirical phenomena because they believe them to be the best explanations of those phenomena. This was the reasoning Darwin used in judging his theory of natural selection to be superior to the rival creationist explanation of his time (Thagard, 1978). In contrast to the hypothetico-deductive method, IBE takes the relation between theory and evidence to be one of explanation, not logical entailment. This means that for IBE the ideas of explanation and evidence come together, and explanatory reasoning becomes the basis for evaluating theories. Also, in contrast with the Bayesian approach to theory evaluation, advocates of IBE generally take theory evaluation to be a predominantly qualitative exercise that focuses explicitly on explanatory criteria, not a statistical undertaking in which one assigns probabilities to theories. Given that a primary function of most theories in science is to explain empirical facts, it stands to reason that the explanatory goodness of explanatory theories should count in their favor. Conversely, explanatory failings should detract from their credibility. The major point of IBE is that the theory judged to be the best explanation of the facts is taken to be the theory most likely to be correct. There is, then, a twofold justification for using IBE when evaluating explanatory theories: it explicitly assesses such theories in terms of the important goal of explanatory power, and it focuses on science's goal of maximizing truth. The basis for this second justification is briefly considered later in the chapter.

Methodologists have used a number of different terms for explanatory reasoning. Many have followed Charles Peirce (1931–1958) in calling it *abduction*. Others have adopted Gilbert Harman's (1965) term *inference to the best explanation*, and still others speak of *explanatory induction* (Rozeboom, 1997). However, the tendency in the literature to think of IBE as the generic form of explanatory reasoning can mislead, for it glosses over the fact that there are different forms of explanatory reasoning—or, as one might say, different forms of abduction. The terminological preferences I adopt here acknowledge genuine differences in methodological context. In this chapter, I distinguish between the abductive generation of new theories and the abductive appraisal of existing theories. This is similar to Capaldi and Proctor's (2008) distinction between *novel hypothesis abduction* and *competing theories abduction*. In each case, it is the latter that is more appropriately described as IBE. With Peirce, I take abduction to involve reasoning from claims about puzzling facts to theories that might explain them. As such, abduction is a process of hypothesis or theory generation that can, at the same time, involve an evaluation of the initial plausibility of the hypotheses and theories proposed. In chapter 3, I argued that exploratory factor analysis is an abductive method that helps researchers generate plausible explanatory hypotheses in domains where it is reasonable to suppose that common causes are at work. Abduction in this sense is to be contrasted with IBE, which involves a comparative assessment of rival theories—theories that might have been given to us by the generative process of abduction, as with exploratory factor analysis, and perhaps developed by a modeling process of analogical abduction, which was the subject of chapter 4. Thus the expression *inference to the best explanation* should not be taken to imply that one arrives at the best explanation by reasoning to it. Rather, IBE is a mode of inference by which one judges the best of existing competing explanatory hypotheses and theories that have been generated by other abductive means. This chapter focuses on IBE in the latter sense.

Although scientists often make judgments of IBE, they disagree about how to characterize that process. Accordingly, the characterization of IBE provided in this chapter highlights four major attempts to render this form of inference intelligible. The first of these, often used by philosophers, portrays IBE as a schematic argument. The second, by Peter Lipton (2004), claims that IBE leads to judgments of "explanatory loveliness." The third account is Paul Thagard's (1992), which depicts IBE as a method of determining the explanatory coherence of theories. A fourth characterization presents a variant of the currently popular method of structural equation modeling as a form of IBE.

5.2.1 Inference to the Best Explanation as a Schematic Argument

It is commonly thought that with IBE one infers the likely truth of a hypothesis on the grounds that it better explains a set of data than do competing hypotheses. This characterization of IBE is sometimes presented in the form of a general argument schema like the following (e.g., Josephson & Josephson, 1994; Lycan, 1988):

D is a collection of data.

Hypothesis H explains D.

No other hypothesis can explain D as well as H does.

Therefore H is probably true.

This schematic portrayal of IBE provides some sense of the structure of an IBE argument, but generally speaking, IBE in science does not conform to this schema. Therefore the schema must be amended in light of the following remarks.

First, as emphasized in the discussion of phenomena detection in chapter 2, and as clarified in the abductive depiction of exploratory factor analysis provided in chapter 3, the facts to be explained in science are generally not collections of data but empirical phenomena. Phenomena often take the form of empirical generalizations, and they are not, strictly speaking, observed. Rather, data serve as evidence for phenomena, and phenomena are taken as the usual objects of scientific explanation (Woodward, 1989).

Second, the argument schema refers to hypotheses rather than theories. However, in science, theories are often taken to be the minimum units of theory appraisal. Theories are ramified structures, often comprising several explanatory hypotheses and other factors such as empirical generalizations and models (in the previous chapter, we saw that ATOM is concerned with developing model theories). Typically, IBE is used in science to evaluate theories rather than hypotheses.

Third, the conclusion of the argument schema speaks of the probable truth of the hypothesis. However, although truth is a cardinal aim of science, and although hypotheses are more or less true, the conclusion of the argument does not require talk of truth, let alone probable truth. It is sufficient that the conclusion speaks of the acceptance of the hypothesis in preference to its rivals.

On the basis of these brief remarks about the nature of science, the schematic depiction of the form of an IBE argument just given should be changed to something like the following:

P1, P2, … are surprising empirical phenomena.

Theory T explains P1, P2, … .

No other theory can explain P1, P2, … as well as T does.

Therefore T is accepted as the best explanation.

Note that this schematic depiction of IBE focuses on its form only. However, a more informative characterization of IBE requires one to supplement the schema to capture the complexity of the patterns of reasoning involved. This is especially so when we seek a method of IBE that can help us judge theory goodness. Importantly, a satisfactory account of IBE must be able to say what it means for one explanation to be better than its rivals. We will see that the second and third accounts of IBE meet this requirement by providing a set of criteria that form the basis of the judgments made in IBE. I should also point out here that the notion of an explanation itself remains unclear. However, for the purpose of this chapter, we can assume that a scientific explanation often involves appealing to causes that *produce* their effects, irrespective of how that appeal might be spelled out in detail. Later, I briefly comment on the idea of explanation in dealing with a criticism of Thagard's account of inference to the best explanation.

5.2.2 Inference to the Best Explanation as the Loveliest Explanation

The philosopher of science Peter Lipton has undertaken the most prominent and wide-ranging examination of IBE. In his book *Inference to the Best Explanation* (2004), Lipton articulated and defended IBE as a distinctive kind of inference, which is used in both science and everyday life. With science in mind, Lipton examined and endorsed the related ideas that we often accept a theory on the grounds that it provides a better explanation of the evidence than its rivals do, and the explanatory success of a scientific theory is a good reason to believe or accept that theory as true. Lipton took pains to distinguish between the descriptive task of understanding IBE as it is practiced in science and the normative task of showing how IBE provides a justification for the conclusions reached. His primary concern was the descriptive merits of IBE.

Lipton pointed out that the phrase *best explanation* is ambiguous between what he called the most *likely explanation* and the most *lovely explanation*. Some methodologists take IBE to provide us with the likeliest or most probable explanation. However, Lipton maintained that this approach is not particularly informative because the primary task of IBE is to say what leads to a judgment that one theory is likelier than another. Lipton claimed that it is more informative to regard the best explanation as the loveliest explanation and to use that information to gauge the likeliness of a theory's truth. However, for Lipton, the idea that IBE is the loveliest explanation can stand on its own without analyzing it in terms of probability. Nevertheless he maintained that if IBE is a good model of our inferential practices, then loveliness and likeliness will tend to be coextensive.

For Lipton, the loveliest explanation comprises the various commonly accepted explanatory virtues, and it is these virtues that provide the guide to inference about causes in science. Lipton listed these virtues as unificatory power, precision, and elaboration of explanatory mechanisms. Although he acknowledged that there is a literature on these and other explanatory virtues, he did not articulate his virtues in detail. However, he stressed the importance of doing so in a more fully developed account of IBE. Because of this lack of detail, the sense in which explanatory loveliness determines explanatory likeliness is somewhat unclear.

Lipton depicted IBE as a two-stage process. In the first stage, a set of potential explanations is generated. In the second stage, an inference is made to the best potential explanation, which is accepted as the actual explanation. Each of these stages involves filtering out a reduced set of explanations on the basis of plausibility considerations. At the first stage, judgments of initial plausibility are made on the basis of background knowledge to identify the potential explanations from all possible explanations. At the second stage, the criteria that comprise the loveliest explanation are used to determine the best of the potential explanations. By including a first stage of hypothesis generation in his model of IBE, Lipton took IBE to be a broader notion than the one I adopt in this chapter, which is confined to his second stage. From the perspective of this chapter, Lipton's first stage can be understood as a necessary precursor to IBE proper.

Although Lipton's two-step filtering process undoubtedly points to important features of scientific research, his abstract characterization of the process of IBE constitutes a general strategy rather than a detailed method. Nevertheless Lipton maintained that IBE shares some similarities with, but goes beyond, Mill's methods of induction, in terms of both applicability and scope. He also maintained that it is different from, and superior to, the hypothetico-deductive method, because it avoids various counterexamples or paradoxes of confirmation to which that method gives rise.

Although Lipton maintained that an analysis of IBE can be given without a satisfactory theory of explanation, he adopted a causal model of explanation as an explicit part of his account of IBE. At the same time, Lipton stressed the importance of the notion of contrastive explanation. A contrastive explanation does not attempt to answer the question "Why this event?" It attempts to answer the question "Why this event rather than that event?" That is, it seeks causes to explain not an event by itself but an event together with the absence of another relevant similar event. As an illustration of the contrastive model of explanation, Lipton took Ignaz Semmelweis's (1983) much-discussed investigation of childbed fever. Semmelweis sought to explain why the incidence of childbed fever was higher in one obstetric clinic than another. By creating a range of contrasts that controlled for relevant differences (e.g., the regular washing of hands with a solution of chlorinated lime contrasted with not doing so), he was able to conclude that the women in some wards were infected by the examining medical students, who carried an unknown material on their unwashed hands. Lipton construed this example as a successful case of IBE, where IBE is construed as inference to the best contrastive explanation.[1]

A further aspect of Lipton's treatment of IBE should be mentioned here. This is his recent suggestion (Lipton, 2004) that IBE is broadly compatible with the Bayesian approach to theory evaluation and that IBE might in fact help determine the prior probabilities and likelihoods that are used in Bayes's theorem. I consider the compatibility of these two approaches later in the chapter.

In addition to his positive account of IBE outlined here, Lipton defended IBE against a number of criticisms. I consider the two most prominent of these criticisms in section 5.3.

5.2.3 Inference to the Best Explanation as Explanatory Coherence

Gilbert Harman (1965) provided the first modern reference to IBE. However, Harman gave no informative account of IBE itself. He chose merely to mention simplicity, plausibility, explanatory breadth, and non–ad hocery as the sort of criteria that figure in judgments of best explanation. As noted earlier, Lipton acknowledged the importance of criteria like these, but he did not provide a detailed account of them. As a result, critics of both Harman and Lipton complained that without an informative account of the criteria that would be used in IBE, the idea was little more than a slogan.

Recognizing this deficiency, Paul Thagard developed a method of IBE that helps researchers reliably appraise competing theories. His method is known as the *theory of explanatory coherence* (TEC) (Thagard, 1989, 1992). The theory comprises an account of explanatory coherence in terms of a number of principles, a computer program for implementing the principles, and various simulation studies that demonstrate the theory's promise as a method of IBE. In this section, I provide an overview and a brief evaluation of the method.

According to TEC, IBE is centrally concerned with establishing relations of explanatory coherence. To infer that a theory is the best explanation is to judge it as more explanatorily coherent than its rivals. TEC is not a general theory of coherence that subsumes different forms of coherence, such as the logical coherence of deductively valid arguments, and the probabilistic coherence of Bayes's theorem. Rather, it is a theory of *explanatory* coherence, where the propositions hold together because of their explanatory relations.

Relations of explanatory coherence are established through the operation of seven principles. These are symmetry, explanation, analogy, data priority, contradiction, competition, and acceptability. A theory's explanatory coherence is determined in terms of three criteria: explanatory breadth, simplicity, and analogy (Thagard, 1978). The criterion of explanatory breadth, which Thagard believes is the most important for choosing the best explanation, captures the idea that a theory is more explanatorily powerful than its rivals if it explains a greater range of facts—the idea strongly endorsed by Darwin in the quotation presented earlier in the chapter. The notion of simplicity that Thagard deems most appropriate for theory choice is captured by the idea that we should prefer theories that make fewer special or ad hoc assumptions. Finally, explanations are judged more coherent if they are supported by analogy to theories that scientists already find credible. Within TEC, each of these three criteria is embedded in one or more of the seven principles. Thagard formulated these principles in both formal and informal terms. They are stated here informally in his words (Thagard, 2000, 43). The accompanying comment on the principles closely follows Thagard's (1992) discussion of a more formal statement of those principles.

1. Symmetry

Explanatory coherence is a symmetric relation, unlike, say, conditional probability. That is, two propositions *p* and *q* cohere with each other equally.

The principle of symmetry maintains that both coherence and incoherence are symmetric relations, unlike the nonsymmetric relations of entailment and conditional probability. The sense of coherence conforms to the ordinary sense of coherence as "holding together."

2. Explanation

(a) A hypothesis coheres with what it explains, which can either be evidence or another hypothesis. (b) Hypotheses that together explain some other proposition cohere with each other. (c) The more hypotheses it takes to explain something, the lower the degree of coherence.

Because the principle of explanation establishes most of the coherence relations, it is the most important principle in determining explanatory coherence. Principle 2a, with principle 7, acceptance, subsumes the criterion of explanatory breadth, which is central in determining the best explanation. Principle 2c accommodates the notion of simplicity, which is also an important criterion in theory choice.

3. Analogy

Similar hypotheses that explain similar pieces of evidence cohere.

The principle of analogy is the same as the criterion of analogy in Thagard's (1978) initial account of IBE. It states that if similar propositions explain similar pieces of evidence, then they cohere with each other. The analogy must be explanatory.

4. Data priority

Propositions that describe the results of observations have a degree of acceptability on their own.

The principle of data priority maintains that claims about observations can stand on their own more successfully than explanatory hypotheses. Of course, they can be doubted, but the reliability of their production will often be sufficient grounds for their initial acceptance.

Despite its name, it is clear that Thagard intends the principle of data priority to include statements about empirical generalizations that are based on observations. Thus the principle covers the generalizations that are robust enough to be considered claims about empirical phenomena, in the sense discussed in chapter 2. Because of their robustness, the evidential respectability of such claims will be high, apart from their relationship to explanatory theories.

5. Contradiction

Contradictory propositions are incoherent with each other.

This principle straightforwardly includes syntactic contradictions involving logical inconsistency and semantic contradictions involving inconsistency of meaning. The principle covers the negative relations that hold between contradictory propositions that actively resist cohering and are said to incohere.

6. Competition

If *p* and *q* both explain a proposition, and if *p* and *q* are not explanatorily connected, then *p* and *q* are incoherent with each other (*p* and *q* are explanatorily connected if one explains the other or if together they explain something).

This principle claims that theories that explain the same evidence should normally be treated as competitors. In such cases, theories are regarded as competing if they are not explanatorily related. Noncontradictory theories may compete with each other.

7. Acceptance

The acceptability of a proposition in a system of propositions depends on its coherence with them.

This last principle asserts that propositions are accepted or rejected on the basis of their degree of coherence with other propositions. The overall coherence of a system of propositions, or a theory, is obtained by considering the pairwise coherence relations through use of principles 1 through 6.

The principles of TEC combine in a computer program called ECHO (Explanatory Coherence by Harmany Optimization) to provide judgments of the explanatory coherence of competing theories.[2] In ECHO, propositions about both evidence and hypotheses are represented by units that have excitatory and inhibitory links to each other, and node activation represents the nodes' degree of coherence with all propositions in the network. The system updates itself based on parallel constraint satisfaction. The best explanation consists of the nodes with the highest activation values once the system has settled down.

TEC has a number of virtues that make it a promising theory of IBE. It focuses on criteria and principles that manifestly have to do with explanation; the criteria of explanatory breadth, simplicity, and analogy are explanatory criteria, whereas the principle of explanation is the most important of the seven principles. Further, as its principle of competition makes clear, TEC takes theory evaluation to be a comparative matter in which a theory is evaluated with reference to one or more competing theories. Furthermore, it is instantiated by, and can be implemented in, the purpose-designed computer program ECHO; it is a considerable achievement of TEC that it enables the researcher to compute explanatory coherence. Finally, it accounts for a number of important episodes of theory assessment in the history of science, such as the superiority of Darwin's theory of evolution over the creationist theory, and the superiority of Lavoisier's theory of oxygen over the phlogiston theory. Simulation studies by Thagard and his colleagues on case histories such as these provide empirical evidence that TEC is on the right track with its distinctive conception of IBE. It is largely for these reasons that I have chosen TEC as the method of IBE for ATOM.

Despite these positive features, TEC is a controversial account of IBE. Two publications by Thagard (1989, 1992) contain a number of criticisms of the method, with replies by the author. Some of these criticisms apply to IBE more generally, and I consider three of them in the next two sections. However, before doing so, I want to briefly consider the fact that TEC's key principle of explanation speaks of hypotheses explaining other propositions without indicating what the term *explanation* means. Some commentators have seen this as a deficiency (e.g., Achinstein, 1989; Glymour, 1992). Thagard was aware of this omission at the outset but maintained that TEC is an effective method of IBE, although it operates with a primitive notion of explanation. More recently, Thagard and Litt (2008) claimed that explanation is a complex process that resists characterization in a single account. It can involve features such as deductive arguments, statistical relations, schema applications, analogical comparisons, and linguistic acts, all of which are subordinate to its fundamental causal character. Thus, for them, the focal challenge in characterizing the explanatory relationship between hypotheses and the propositions they explain is to describe the causal relationship between them. Thagard and Litt developed a neurocomputational model of the cognitive processes that underlie scientific explanations. Their model is much more neurologically complex than the simple model of ECHO. Both Thagard's multifaceted characterization of explanation and the new neurocomputational model should therefore be viewed as complementary to, not a part of, TEC and its accompanying methodology.

Toward the end of the chapter, I will suggest that TEC provides psychologists with a valuable method for engaging in the comparative appraisal of explanatory theories.

5.2.4 Inference to the Best Explanation as Structural Equation Modeling

The guess-and-test strategy of the standard hypothetico-deductive method takes predictive accuracy as the sole criterion of theory goodness. However, a close examination of research practice in psychology and the behavioral sciences reveals that the hypothetico-deductive method is sometimes combined with the use of supplementary evaluative criteria such as simplicity, scope, and fruitfulness. When this happens, and one or more of the criteria have to do with explanation, we can reasonably regard the combined approach as a version of IBE, rather than just an augmented account of the hypothetico-deductive method. As noted earlier, this is because the central characteristic of the hypothetico-deductive method is a relationship of logical entailment between theory and evidence, whereas with IBE the relationship is one of explanation. The hybrid version of IBE being considered here will allow the researcher to say that a good explanatory theory will rate well on the explanatory criteria and at the same time boast a measure of predictive success. Most methodologists and scientists will agree that an explanatory theory that also makes accurate predictions will be a better theory for doing so.

Structural equation modeling, now widely used in psychology and related sciences, is a family of multivariate statistical methods that often involves testing models in hypothetico-deductive fashion. Its standard formulation is a combination of insights from multiple regression, path analysis, and confirmatory factor analysis, which enables structural equation modeling simultaneously to test relationships among a multitude of manifest and latent variables. It specifies and tests models of linear structural relations, which are often given a causal interpretation. One or more goodness-of-fit measures provide the means by which one confirms or disconfirms the model in question. Structural equation modeling in this sense is hypothetico-deductive because it is centrally concerned with the predictive testing of models one at a time without regard for competing plausible models.

However, some uses of structural equation modeling combine a commitment to predictive hypothetico-deductive testing with an appeal to one or more explanatory criteria. This latter practice involves the explicit comparison of models or theories in which an assessment of their goodness-of-fit to the empirical evidence is combined with the weighting of the fit statistics in terms of parsimony indices (e.g., Kaplan, 2000). Here goodness-of-fit provides information about the empirical adequacy of the model, whereas parsimony functions as a criterion relating to the explanatory value of the model. Both are used in judgments of model goodness.

Markus, Hawes, and Thasites (2008) recently suggested that in structural equation modeling, model fit can be combined with model parsimony, understood as explanatory power, to provide an operationalized account of IBE. They discussed the prospects of using structural equation modeling in this way to evaluate the comparative merits of two- and three-factor models of psychopathy. In their chosen example, Markus and his coauthors reported a study by Cooke and Michie (2001), which employed confirmatory factor analysis (a limiting case of structural equation modeling) to conclude that the commonly accepted two-factor structural model of psychopathy (comprising negative interpersonal and affective features, and social dominance) fitted poorly to the data. In its stead, Cooke and Michie proposed a better-fitting three-factor model (comprising arrogant and deceitful interpersonal style, deficient affective experience, and impulsive and irresponsible behavioral style). Markus et al. concluded that this is an example of IBE, where the factors of the two models are taken to be latent explanatory variables. They suggested that one can partially operationalize bestness in terms of the popular root mean square error of approximation index, an index that measures the degree of ill fit per degree of freedom in the model. Here poorness-of-fit and degrees of freedom are taken to represent strength of empirical test and extent of parsimony, respectively, which together can be taken as a gauge of explanatory power. The three-factor model of psychopathy is thus accepted on the grounds that it is a better explanation of the data than the rival two-factor model.

Structural equation modeling recommends itself to psychologists, not just as a hypothetico-deductive practice but also as a variant of IBE. Employed as a method of IBE, it brings with it an ability to provide a better justification than orthodox hypothetico-deductive method of the hypotheses, models, and theories it evaluates.

5.3 Two Criticisms of Inference to the Best Explanation

Something of the controversial nature of IBE was seen earlier when I addressed specific criticisms that have been leveled against Thagard's TEC. I now consider the two most prominent general criticisms that have been leveled against IBE. Another criticism, which appeals to the alleged superiority of the Bayesian approach to theory appraisal, will be dealt with shortly.

5.3.1 The Bad Lot Argument

A major criticism of IBE was raised by Bas van Fraassen (1989), who maintained that the approach cannot provide a satisfactory basis for believing in a theory. In a nutshell, van Fraassen argued that the best of competing explanatory hypotheses might be "the best of a bad lot," all of which are false. He reasoned that because IBE can select the best hypothesis only from the set of currently available hypotheses, we have no reason to believe that the truth is to be found there rather than in hypotheses that no one has proposed. Therefore he maintained that IBE provides us with no rational grounds for believing that the hypothesis that is judged best is true.

Proponents of IBE have met van Fraassen's objection head-on. They have argued that scientists appeal to background knowledge to select the best of competing theories, and because this knowledge is approximately true, their selection of the best theory is generally well grounded. Lipton (2004) argued along these lines. He asserted that our rankings of the best of competing theories are fairly reliable, a point with which van Fraassen agreed. Furthermore, Lipton maintained that for accepted background theories to be used in the successful ranking of theories, they must be approximately true. From this, Lipton concluded that our best-ranked theories must be at least approximately true, and consequently van Fraassen's argument is unsound.

Another way of dealing with van Fraassen's bad lot argument is to put the question of truth aside and focus on the methodological strategies involved in carrying out IBE. Often in science, a theory will count as a viable candidate for selection as the best explanatory theory only when it has already been subjected to one or more plausibility assessments. Recall that, in Lipton's general two-stage model of IBE, the first stage involves reducing the set of all possible explanations to the set of plausible explanations, and the second stage determines the best of the actual explanations. In both stages, judgments are based on plausibility considerations. ATOM adopts a similar but more detailed strategy. According to this account of method, theory construction involves two rounds of plausibility assessment before one can make judgments of IBE. First, a theory that is generated to explain one or more empirical phenomena will be judged with respect to its initial plausibility. This first determination of its worth appeals to the soundness of the explanatory argument used in its introduction. As previously noted, psychological theories generated by exploratory factor analysis are evaluated in this way. Second, theories judged to have sufficient initial plausibility then receive a further assessment of their plausibility in terms of the aptness of the models that form the basis of their extension. To recall an earlier example, Darwin's theory of evolution by natural selection gained credibility by analogy to the well-known processes of artificial selection. According to ATOM, IBE can be seen as a third round of plausibility assessment rather than just the first or second effort to evaluate a theory. Although we should acknowledge that the best of competing theories might be a poor theory, an explanatory theory with a record of successive appraisals like the one just mentioned, which is judged to be better than its rivals, is likely to be the best of a respectable lot, not a bad lot.

Although one can defend IBE against the bad lot argument by separating IBE from truth, it remains to be shown how IBE can legitimately be used to evaluate theories with respect to their explanatory goodness in a way that avoids judgments of truth per se while at the same time regarding science as a truth-seeking endeavor. It is important to realize that the assumption that one can secure the truth of theories by making judgments of IBE conflates the different notions of truth and justification. It has already been said that truth, understood as correspondence with reality, functions as a guiding ideal for science (Hooker, 1987; Haig & Borsboom, 2012). As such, it is a highly valued but unattained goal that helps us make sense of science as an attempt to represent and intervene in the world. However, as an ideal, truth (or more precisely, approximate truth) is accessible only indirectly by way of the various criteria we use to evaluate and accept theories. Historically, scientists have regarded predictive accuracy, internal consistency, and explanatory power as important criteria of theory acceptance. As justificatory criteria, they can indicate truth, but they do not constitute truth.

For TEC, the criteria of explanatory breadth, simplicity, and analogy are epistemic criteria used in evaluating competing explanatory theories. However, the question arises whether evaluating competing theories in terms of these criteria entitles us to think that the best theories are closer to the truth than their rivals. Thagard (2007) claimed that this will be so, provided that two conditions are met. First, an increase in explanatory breadth by explanation of more empirical phenomena has to occur; second, an increase in explanatory depth by the successful investigation of causal mechanisms in greater detail must be achieved.

From this it follows that accepting a theory on the basis of a judgment of explanatory coherence alone does not mean that it is likely to be true or is closer to the truth than its rivals. All that Thagard's argument entitles us to say is that TEC contributes to the long-term goal of maximizing true propositions and minimizing false ones. Subsequent evaluations of a theory in relation to its rivals will also contribute to that goal, and it is the track record of these assessments over time that will ultimately decide a theory's fate.

5.3.2 The Subjectivity of Inference to the Best Explanation

The second major criticism of IBE states that the evaluative criteria that make up explanatory goodness are relative to a scientist's judgments about what constitutes a good explanation and are therefore too subjective to properly determine the warrant that it confers on the best of competing theories.

Lipton (2004) made a two-pronged reply to this objection, one having to do with inference, the other having to do with explanation. With the first prong, Lipton granted that reliable inference is relative to variation in evidence and background beliefs from person to person, but maintained that audience relativity alone will not prevent IBE from being a reliable form of inference. He appealed to Kuhn's (1977) work on theory appraisal by noting that rational disagreements sometimes stem from nonevidential factors such as a theory's fruitfulness and that this can serve the useful function of allowing the scientific community to hedge its bets. With the second prong, Lipton suggested that his criteria of explanatory loveliness (unificatory power, precision, and the elaboration of causal mechanisms) are also subject to interest relativity. He reasoned that by adopting his contrastive account of explanation, where the same event can be explained with respect to different contrasts, his model allows a substantial measure of relativity of interest in a way that is not damagingly subjective. To give a psychological example using Lipton's turn of phrase, Jennifer's early negative childhood experiences will explain why she has relationship difficulties for someone who is interested in understanding why she, rather than Peter, who does not have relationship difficulties, has relationship difficulties, but not for someone who wants to know why Jennifer developed relationship difficulties when other people with negative childhood experiences did not. Lipton saw his account of contrastive explanation illuminating interest relativity in two ways: different people are interested in explaining different phenomena, and these differences in interests demand explanations that invoke different but compatible elements of a causal story. Lipton concluded that the present argument from subjectivity does not impugn IBE.

In short, the arguments against IBE from the bad lot and from subjectivity do not undermine the viability of the approach.

5.4 Inference to the Best Explanation and Other Methods of Theory Appraisal

As noted at the beginning of the chapter, IBE, the hypothetico-deductive method, and Bayesianism are generally regarded as the major alternative approaches to theory appraisal. I now consider IBE in relation to the other two approaches. These contrasts should serve to characterize further the nature of IBE and help judge its merits.

5.4.1 Inference to the Best Explanation and the Hypothetico-Deductive Method

It has repeatedly been stated that the hypothetico-deductive method has long been the method of choice for evaluating scientific theories (Laudan, 1981), and it continues to have a dominant place in psychology. The hypothetico-deductive method is usually characterized in an austere manner: the researcher takes an existing hypothesis or theory and tests it indirectly by deriving from it one or more observational predictions that are themselves directly tested. Predictions borne out by the data are taken to confirm the theory to some degree; those that do not square with the data count as disconfirming instances of the theory. Normally the theory is not compared with a rival theory or theories with respect to the data, only with the data.

The hypothetico-deductive method, in something like this form, has been strongly criticized by methodologists on a number of counts. As remarked earlier, one major criticism of the method is that it is confirmationally lax. This laxity arises from the fact that any positive confirming instance of a hypothesis submitted to empirical test can confirm any hypothesis that is conjoined with the test hypothesis, irrespective of the plausibility of the conjunct. This occurs because the method distributes confirmation across all claims involved in the derivation of the prediction; it does not have the resources to bestow confirmation on the central test hypothesis alone. Thus the successful hypothetico-deductive test of a prediction bestows confirmation on, among other claims, the auxiliary hypotheses about the relevant measuring instruments regardless of their reliability and validity. In this way, the use of psychometric tests of doubtful validity receives undeserved confirmation in psychology.

Another major criticism of the hypothetico-deductive method is that it founders on the problem of the underdetermination of theory by empirical evidence. That is, the method is incapable of showing that a theory should be accepted on the basis of empirical evidence alone or that one theory is better than another with respect to the empirical evidence.

The seriousness of these criticisms has prompted calls to abandon the hypothetico-deductive method in favor of either IBE or the Bayesian approach to hypothesis and theory evaluation. However, these criticisms tell only against simplistic versions of the method, for it is possible to amend the method in ways that allow it to do useful theory testing. Regarding the problem of confirmational laxity, Giere (1983) recast the method in a way that enables it to test individual hypotheses. Alternatively, one might insert a Bayesian view of confirmation into a hypothetico-deductive framework (Rosenkrantz, 1977). However, whereas many would see this second alternative as providing the hypothetico-deductive method with a superior account of confirmation, using the method in this augmented form would only be appropriate where it made good sense to assign probabilities to hypotheses and theories.

The problem for the hypothetico-deductive method of the underdetermination of theories by empirical evidence might be resolved by adopting a strategy that combines the method with the use of evaluative criteria in addition to predictive accuracy. I briefly consider this possibility in section 5.6.1, where I discuss IBE in relation to psychology. For now, it suffices to note that, by invoking explanatory criteria, IBE has the resources to reduce the gap between empirical evidence and theory and make determinate judgments of explanatory goodness.

5.4.2 Prediction and Theory Evaluation

Although prediction is obviously an essential feature of the hypothetico-deductive method, and although it retains a place in Bayesian theory evaluation and most versions of IBE, it is not a part of TEC. This suggests that predictive success might play different roles in different approaches to theory evaluation.

Examination of a number of case histories in the history of science reveals that, for scientists, the successful prediction of new facts does not necessarily provide better evidence for a theory than do theoretical criteria. For example, Brush (1989) showed that the commonly held view that Einstein's successful prediction of the gravitational bending of light provided strong confirmation of his general theory of relativity was shared by neither Einstein nor the majority of scientists of his time. Einstein (and other physicists) maintained that the coherence and simplicity of the theory were more important criteria for its acceptance than the relevant predictive tests.

Brush (1989) also pointed out that it is a common practice in science, particularly in physics, to take predictive success to cover both the deduction of previously known facts and the successful prediction of new facts, suggesting that the novelty of a prediction is not necessarily an important factor in gauging the evidential worth of a theory. With respect to the general theory of relativity, the successful deduction of Mercury's known orbit was widely considered to be just about as good a source of evidence as the novel prediction of light bending. Brush concluded that the primary value of a successful novel prediction, when compared with the deduction of a known fact, is to provide favorable publicity for a theory. Such was the additional value of the light-bending forecast for general relativity theory.

In addition, the successful prediction of a new empirical phenomenon can sometimes be taken as weaker evidence for a theory, just because of its novelty. Often a scientific fact can plausibly be explained by more than one theory. Thus the discovery of a new fact is likely to result in efforts to construct plausible alternatives to the explanation offered by the theory that sponsored the relevant novel prediction. In the case of general relativity theory, ten years of unsuccessful efforts to provide a better explanation of the phenomenon of light bending passed before Einstein's supporters could convincingly assert that their theory provided the best explanation (Brush, 1989).

In short, it seems that although prediction has a deservedly important role in theory evaluation, it has been less dominant, and its use more varied, than is commonly supposed. That TEC has sufficient resources to produce reliable decisions about the best of competing explanatory theories without recourse to predictions should be considered neither surprising nor untoward. Although TEC ignores predictive success as a criterion of theory appraisal, we should appreciate that it nevertheless satisfies the essential demand for empirical adequacy by appealing to explanatory breadth instead. A theory that satisfies this criterion of empirical adequacy is adequate to the relevant empirical phenomena by being able to explain them.[3] In the nineteenth century, the ability of a theory to explain the relevant phenomena was taken as an important measure of empirical adequacy. TEC usefully brings this neglected criterion of empirical adequacy into the methodological foreground again.

5.4.3 Inference to the Best Explanation and Bayesianism

Although the Bayesian approach to theory appraisal is looked on more favorably in philosophy of science than is the hypothetico-deductive alternative, it remains a minority practice in psychology.[4] Bayesian theory evaluation is widely viewed as an alternative to IBE in the philosophical literature. However, some methodologists have recently looked at ways to bring the two approaches together. In what follows, I briefly characterize the Bayesian outlook on theory appraisal and then consider different ways in which it might relate to IBE. Although work on this topic is at a formative stage, reasonable grounds exist for regarding IBE as a sufficient approach to appraising explanatory theories without recourse to Bayesian ideas.

I noted in chapter 1 that Bayesians consider probabilities to be central to scientific hypothesis and theory choice. Bayesians claim that an appropriate understanding of theory choice is best provided by probability theory, augmented by the allied Bayesian philosophy of science known as *Bayesianism*.[5] In using probability theory to characterize theory evaluation, Bayesians recommend assigning posterior probabilities to scientific hypotheses and theories in light of relevant evidence. Bayesian hypothesis choice involves selecting from competing hypotheses the one that, given the evidence, has the highest posterior probability. The vehicle through which this process is conducted is Bayes's theorem, which can be stated in a variety of forms. Given that the Bayesian position is being contrasted with IBE here, I present Bayes's theorem for the case of two hypotheses. Bayes's theorem is written for each hypothesis in turn. For the first hypothesis,

$$\mathrm{Pr}\ (H1/D) = \frac{\mathrm{Pr}\ (H1) \times \mathrm{Pr}\ (D/H1)}{\mathrm{Pr}\ (H2) \times \mathrm{Pr}\ (D/H2) + \mathrm{Pr}\ (H1) \times \mathrm{Pr}\ (D/H1)}.$$

This says that the posterior probability of the first hypothesis is obtained by multiplying its prior probability by the probability of the data, given that hypothesis (the likelihood), and dividing the product by the value that results from adding the prior probability of the second hypothesis, multiplied by the likelihood for that hypothesis, to the prior probability of the first hypothesis, multiplied by its likelihood. Bayes's theorem for the second hypothesis is written in a similar way.

Although Bayes's theorem is not controversial as a mathematical theorem, it is controversial as a guide to scientific inference. With respect to theory appraisal, one frequently mentioned problem for Bayesians is that the probabilistic information needed for their calculations on many scientific hypotheses and theories cannot be obtained. As noted in chapter 1, it is difficult to know how one would obtain credible estimates of the prior probabilities of the various hypotheses and evidence statements that made up, say, Freud's psychodynamic theory or Darwin's evolutionary theory. Not only are the necessary probabilistic estimates for such theories hard to come by, but they do not seem to be particularly relevant in appraising such explanatory theories. For example, what would it mean, and how would it be possible, to speak in evolutionary psychology of the probability of an adapted psychological trait being responsible for young children being able to solve theory-of-mind problems?

The problem for Bayesianism presented by explanatory theories such as those just mentioned is that scientists naturally appeal to qualitative theoretical criteria rather than probabilities when evaluating those theories. For example, with TEC, the three criteria of explanatory worth identified in Thagard's (1978) case histories are qualitative, even when they are given a precise formulation in terms of the relevant principles and the computer program ECHO. To reiterate, scientific theories for which IBE is an appropriate assessment strategy typically explain empirical phenomena, and in these cases, explanatory reasoning rather than probabilistic reasoning is appropriate for their assessment.[6]

Although IBE has typically been regarded as a competitor to Bayesian theory evaluation (e.g., van Fraassen, 1989), Lipton (2004) argued that the two approaches are broadly compatible, and in fact their proponents "should be friends." In broad terms, he suggested that judgments of the loveliest explanation, which are provided by the evaluative criteria of IBE, contribute to assessments of the likeliest explanation, which are provided by the probabilities of the Bayesian approach. Specifically, Lipton maintained that the explanatory considerations invoked in IBE guide the determination of the prior probabilities (and the likelihoods) that are inserted in Bayes's theorem. However, although appeal to explanatory matters might well be one way in which Bayesians can determine their prior probabilities, Lipton did not suggest how this might be done. Furthermore, those who hold IBE to be a normative approach to scientific theory evaluation, with its own distinctive character, will worry that Lipton relegates it to a descriptive role within a Bayesian normative framework (e.g., Psillos, 2004).

Another way of showing the compatibility of IBE and Bayesianism is to translate the evaluative criteria used within IBE into probabilistic terms. McGrew (2003) did this by taking the important theoretical virtue of consilience, or explanatory breadth, and showing that its Bayesian form leads to higher posterior probabilities of the hypotheses being evaluated. Nevertheless, McGrew acknowledged that if one translates consilience into its "flattened" probabilistic form, it no longer remains a genuine explanatory virtue: not only is there no guarantee that consilience will be concerned with an explanation of the evidence, but there is no way that probabilistic translations of the explanatory virtues can refer to the causal connections that are often appealed to in scientific explanations. Furthermore, Weisberg (2009) recently argued that the explanatory loss incurred in such translations will occur for any distinctively explanatory virtue that is given such probabilistic treatment.

One of the reasons that Bayesians have criticized IBE is that its advocates have not been able to spell out its explanatory criteria in a genuinely informative way. However, as seen earlier in the chapter, Thagard's TEC shows that this is not the case: TEC is a detailed theory of IBE in which the explanatory criteria are described, incorporated in the appropriate principles of explanatory coherence, and implemented as part of an integrated method in the computer program ECHO. Although formal, TEC is clearly qualitative, not probabilistic. Therefore it can stand as a method of theory appraisal apart from Bayesianism.

Nevertheless, Thagard (2000) translated ECHO in terms of Pearl's (1988) probabilistic approach to networks, which suggests that the two approaches can be reconciled. However, the probabilistic version of ECHO comes at some computational and conceptual cost, and unsurprisingly, some of the relevant probabilities are hard to come by. In the absence of further relevant comparative work in this domain, Thagard (2000) conjectured that the psychological and technological applicability of explanationist and probabilistic methods will vary depending on the domain of application. He maintained that scientific reasoning is a domain in which explanationism is clearly appropriate, whereas probabilistic reasoning has application in fields such as medicine. Thus although TEC can be clothed in probabilistic dress, it is best used on its own terms for appraising scientific theories.

5.5 The Proper Scope of Inference to the Best Explanation

In this section, I want to briefly challenge two prominent ideas about the proper scope of IBE: first, the belief that IBE is the main account of scientific method; and second, the belief that IBE underlies all forms of ampliative inference, that is, inference involving arguments in which the conclusions contain information that goes beyond the information contained in their premises.

5.5.1 Inference to the Best Explanation as *the* Scientific Method

Some methodologists judge IBE to be the premier account of scientific method. For example, Psillos (2002) compared IBE with what he saw as its two major alternatives—inductive method (understood as enumerative induction) and the standard hypothetico-deductive account of method—and concluded that IBE provides the best description of scientific method. Psillos reasoned that inductive method can satisfactorily justify claims about empirical generalizations but is explanatorily vacuous, whereas the hypothetico-deductive method deals with explanatory hypotheses but offers a poor method for doing so. He concluded that of the three methods, IBE is the only one that is both highly ampliative (i.e., highly content-increasing in the conclusions it draws) and able to provide a decent justification of the explanatory claims it evaluates.

Although IBE is undoubtedly an important scientific method, I believe that it is a mistake to regard it as a rival to the inductive and hypothetico-deductive accounts of method. As discussed in chapter l, all three methods have at various times been proposed as the main account of scientific method. However, I think that it is more realistic to view them as restrictive, domain-specific methods designed to meet particular research goals. Inductive method in the form of enumerative induction, understood as induction by generalization, is used in detecting empirical phenomena, whereas the hypothetico-deductive method tests hypotheses and theories for their predictive accuracy. By contrast, IBE is used to evaluate explanatory theories in terms of both their explanatory power and their predictive success. ATOM explicitly acknowledges the differences in the nature of, and research goals for, the three accounts of scientific method just discussed.

5.5.2 Inference to the Best Explanation and Inductive Inference

A different way of overstating the importance of IBE is to regard it as the superordinate form of ampliative inference. Harman (1965) introduced the idea of IBE to modern philosophy in an effort to show that it underlies all forms of inductive inference. He insisted that enumerative induction is really a special case of IBE. Lipton (2004) also argued for this conclusion.

The problem with this perspective on inference is that enumerative induction does not lead to an explanatory conclusion in any interesting sense of the term, and for this reason, it is fundamentally unlike IBE. Although inductive arguments are ampliative in character, they are descriptive in nature because they reach conclusions about the same types of manifest attributes that are mentioned in the arguments' premises. The widespread practice in psychology of drawing statistical conclusions about populations based on sample characteristics is a case in point. By contrast, IBE is explanatory inference where criteria of explanatory goodness figure centrally in the form of reasoning involved. In TEC, for example, the complex nature of the explanatory reasoning involved is embedded in the principles, the criteria, and the computer program and can be spelled out only with reference to them. IBE and inductive inference, then, are different forms of ampliative inference. In science, IBE is invoked in the explanatory endeavor of theory evaluation. Inductive inference is exemplified in the descriptive tasks of generalizing from statistical samples and establishing claims about empirical phenomena.

5.6 Implications for Psychology

In this penultimate section of the chapter, I consider IBE specifically in relation to psychology. I make some general suggestions about how psychologists might engage with IBE, and recommend ways in which psychology might incorporate IBE into its methods curriculum.

5.6.1 Inference to the Best Explanation in Psychology

Although the standard characterization of the hypothetico-deductive method takes predictive accuracy as the sole criterion of theory goodness, it is plausible to suggest that, in research practice, the hypothetico-deductive method is often combined with the use of supplementary evaluative criteria, such as simplicity, scope, and fruitfulness. This probably explains, at least in part, why the method continues to be widely used in psychology and other sciences. As noted before, it is important to appreciate that, to the extent that these complementary criteria are concerned with explanation, we can appropriately regard the combined approach as a version of IBE rather than an augmented account of the hypothetico-deductive method.

Although psychological researchers do not often discuss the explanatory virtues of their theories, a number of instructive accounts of the virtues of scientific theories in the philosophical literature could help them do so. Perhaps the best-known account is that of Thomas Kuhn (1977), who identified and discussed accuracy, consistency, scope, simplicity, and fruitfulness as five important criteria that are standardly used to adjudicate in theory choice. Another useful account of the theoretical virtues is that of Willard Quine (Quine & Ullian, 1978), who provided a lucid discussion of the notions of conservatism, modesty, simplicity, generality, and refutability. In an important discussion of the place of values in science, McMullin (1983) furnished a different list of virtues: predictive accuracy, internal coherence, external coherence, unifying power, and fertility. To these three accounts of the theoretical virtues, we can add Thagard's (1978) multicriterial account of IBE, discussed earlier in the chapter. If psychologists made a deliberate effort to appraise their explanatory theories by drawing from a number of the criteria just listed, they would be practicing IBE in one of two senses. To the extent that they used several of these nonexplanatory criteria for the purposes of comparative theory appraisal, they would be able to make inferences to the best explanatory theory, even if those criteria were not directly concerned with explanation. To the extent that they used criteria that have to do directly with explanation, they would be engaging in explanatory inference as a basis for deciding between the competing theories. Both approaches are superior to the hypothetico-deductive method as it is traditionally understood and practiced in psychology.

This chapter has given a fair degree of attention to Thagard's TEC. Although TEC is the most codified explicit account of IBE available today, further development of aspects of the approach would make it a genuinely useful method for psychological researchers. These would include developing contemporary case studies of its use in psychology, making a user-friendly version of the computer program ECHO commercially available for the ready implementation of TEC, and augmenting the method of TEC by explicitly linking it to a suitable theory of explanation.

5.6.2 Inference to the Best Explanation in the Methods Curriculum

For IBE to be regularly practiced in psychology, the research methods curriculum will have to broaden its perspective on theory appraisal (see Capaldi & Proctor, 2008). As noted earlier, psychologists should be encouraged to practice IBE in their evaluation of explanatory theories, either by combining an acceptable version of the hypothetico-deductive method with the use of complementary evaluative criteria, as just noted, or by employing TEC. Thagard (1992) is the definitive source for a detailed explication of TEC. An introduction to using the computer program ECHO to compute explanatory coherence can be found at Thagard's Computational Epistemology Laboratory website (http://cogsci.uwaterloo.ca/JavaECHO/jecho.html). The site provides simple examples that show how ECHO deals with the criteria of explanatory breadth, simplicity, and analogy. Substantive examples of scientific theory choice can also be run. In addition, textbooks should present a view of IBE as an important approach to theory appraisal for psychology that is part of good scientific practice. Proctor and Capaldi's (2006) textbook on psychological research methodology, *Why Science Matters*, breaks new ground in this regard.

Although explicit discussions of IBE are rare in psychology, a few methodological articles in the psychological literature will help researchers begin to understand different aspects of IBE. Erwin (1992) argued that debates about the philosophy of scientific realism are relevant to the evaluation of behavior theories and outcome hypotheses, and IBE figures centrally in these debates. Eflin and Kite (1996) demonstrated empirically that instruction and practice in IBE improve the reasoning of psychology students in evaluating competing psychological theories. Rozeboom (1997) compared the hypothetico-deductive, Bayesian, and abductive approaches to theory appraisal and argued that researchers in psychology should use his approach to IBE, known as *explanatory induction*. In Haig (2005a) I proposed ATOM as a broad theory of scientific method in which theory evaluation involves using IBE in the form of TEC (this book covers essentially the same ground). More recently, Capaldi and Proctor (2008) argued, against some popular relativist trends in psychology, for the comparative appraisal of psychological theories through an approach to IBE they call *competing-theories abduction*. In their paper, Capaldi and Proctor provide an example in experimental psychology of the use of IBE to evaluate two formal theories of attention—similarity choice theory and signal detection theory—with respect to the relevant facts. They suggest that considerations of IBE establish the fact that no other theories of attention come close to explaining the range of empirical phenomena explained by these two theories. As noted earlier, Markus et al. (2008) argued for an understanding of structural equation modeling in terms of IBE. Finally, Durrant and Haig (2001) argued that more rigorous evolutionary theories of human psychological phenomena could be achieved by using IBE as a strategy for evaluating adaptationist explanations. Although much work remains to be done to further develop the methods of IBE, these resources should offer both the researcher and the methodologist a sense of the nature of IBE and its relevance to psychology.

5.7 Conclusion

Psychologists, for the most part, evaluate their hypotheses and theories in accord with the dictates of the orthodox account of hypothetico-deductive method. This has resulted in two unfortunate practices: testing psychological theories in isolation without reference to alternative competing theories, and evaluating those theories in terms of their predictive adequacy without regard for relevant explanatory criteria. IBE is a good approach to theory appraisal because it corrects for these malpractices. True to its name, IBE characterizes theory appraisal as an inherently comparative practice, in which two or more theories are evaluated with respect to each other on multiple criteria of explanatory goodness.

The literature on IBE is now sufficiently well developed to offer genuine help to psychologists in explicitly evaluating theories in domains comprising two or more reasonably well-developed competing explanatory theories. I have argued in this chapter that the major criticisms of IBE have not cast doubt on its worth as an approach to theory appraisal. I have presented four different perspectives on IBE. Taken together with other contributions to the literature, they constitute a valuable methodological resource. By acknowledging the importance of explanatory theories in science, one can justifiably use IBE to appraise theories with respect to their explanatory goodness. Psychology is replete with competing theories that might usefully be evaluated with respect to their explanatory worth. With the advent of the methodology of IBE, psychologists can position themselves to make these judgments in a more systematic way than did scientists before them, such as Darwin and Einstein.

However, one should not underestimate the challenges involved in employing IBE. Apart from TEC, and some versions of structural equation modeling, no inferential algorithms are available to help researchers engage in IBE. Researchers who want to employ IBE will have to adopt more of a do-it-yourself attitude than they do in their customary use of the hypothetico-deductive method and classical statistical significance testing. Courses and workshops that focus on IBE simply do not exist at present. Researchers will have to learn from the existing primary literature for themselves what the (somewhat different) approaches to IBE involve. Nevertheless this prospect should appeal to psychologists who want to learn about the comparative explanatory worth of their theories and use those judgments as grounds for accepting or rejecting them.

6 Conclusion

Scientific method, taken as a logical, epistemic, and cognitive process, is certainly at least as complex as, say, the theory of evolution. We do neither of these phenomena justice by failing to appreciate how puzzling they can be.
—James Blachowicz (2009, 306)

6.1 Introduction

In this concluding chapter, I round out my characterization of ATOM. I begin by outlining a promising theory of the nature of research problems and show how it is deployed in ATOM. I then offer some supplementary remarks about the nature of ATOM. This is followed by two applications of ATOM, after which I consider a number of criticisms and misunderstandings of the theory that have surfaced to date. Toward the end of the chapter, I discuss scientific method in relation to science education and conclude with some cautions and caveats about ATOM.

6.2 A Coda on Scientific Problems

The overview of ATOM presented in chapter 1 signaled the theory's serious commitment to the notion of a research problem. This emphasis on the importance of research problems for inquiry contrasts with the orthodox inductive and hypothetico-deductive accounts of method, neither of which speaks of problem solving as an essential part of its characterization.[1]

In an effort to depict scientific inquiry as a problem-oriented endeavor, ATOM deploys the *constraint-inclusion* view of research problems (Haig, 1987; Nickles, 1981). The idea of a problem as a set of constraints has been taken from the problem-solving literature in cognitive psychology (Reitman, 1964; Simon, 1977) and adapted for a methodological role.

Briefly, the constraint-inclusion theory depicts a research problem as comprising all the constraints on the solution to that problem, along with the demand that the solution be found. With the constraint-inclusion theory, the constraints do not lie outside the problem but are constitutive of the problem itself; they actually serve to characterize the problem and give it structure. The explicit demand that the solution be found is prompted by a consideration of the goals of the research program, the pursuit of which is intended to fill the outstanding gaps in the problem's structure. The goals themselves are part of the problem. Problems can only be solved by achieving research goals, and a change in goals will typically eliminate or at least alter those problems (Nickles, 1988).

The constraints that make up research problems are of various sorts. Importantly, many of them are heuristics, but some are rules, and a limited number have the status of principles. These constraints differ in their nature: some are metaphysical, others are methodological, and many are drawn from relevant substantive scientific knowledge. Problems and their constraints also vary in their specificity. Some are rather general and have widespread application (e.g., "Generate a theory that explains the relevant facts"). Others are context specific (e.g., "Employ common factor analysis to generate a common causal explanation of the correlated effects"). Still others are more specific (e.g., "Use both the scree test and parallel analysis when determining the number of factors in an exploratory factor analytic study").

Note that all relevant constraints are included in a problem's formulation. This is because each constraint contributes to a characterization of the problem by helping to rule out some solutions as inadmissible. However, at any one time, only a manageable subset of the problem's constraints will be relevant to the specific research task at hand. Also, by including all the constraints in the problem's articulation, the problem enables the researcher to direct inquiry effectively by pointing the way to its own solution. The constraint-inclusion account of problems enables the researcher to understand readily the force of the adage that stating the problem is half the solution.

The constraint-inclusion account stresses that in good scientific research, problems typically evolve from an ill-structured state and eventually attain a degree of well-formedness, such that their solution becomes possible. From the constraint-inclusion perspective, a problem will be ill structured to the extent that it lacks the constraints required for its solution. Because the most important research problems will be decidedly ill structured, we can say of scientific inquiry that its basic purpose is to better structure our research problems by building in the various required constraints as our research proceeds. It is by virtue of such progressive enrichment that problems can continue to direct inquiry.

Turning now to ATOM, I should emphasize that its problems dimension is not a temporal phase to be dealt with by the researcher before moving on to other phases, such as observing and hypothesizing. Instead the researcher deals with scientific problems all the time. Problems are generated, selected for consideration, developed, and modified in the course of inquiry. This common error in talking about research problems as a temporal phase is noted in the discussion of grounded theory method in the next section.

Across the various research phases of ATOM, there are numerous problems of varying degrees of specificity to articulate and solve. For example, the successful detection of an empirical phenomenon produces an important new general constraint on the subsequent explanatory efforts devised to understand that phenomenon. Until the relevant phenomenon, or phenomena, are detected, one will not really know what the explanatory problem is. At a more specific level, myriad constraints regulate the process of phenomena detection. For example, if one assumed that the appropriate strategy of phenomena detection was a sequence of data analytic activities in the manner of the multistage model of data analysis outlined in chapter 2, then a host of constraints arising from consideration of the various data analytic methods employed would be part of the evolving research problem. These would include constraints such as using an appropriate maximum likelihood technique when data are randomly missing in the stage of initial data analysis; using back-to-back stem-and-leaf displays for close exploratory comparison of similar batches of data; employing a bootstrap resampling technique in the stage of close replication; and adopting an appropriate triangulation strategy as a basis for accepting the validity of a generalization wrought from constructive replication.

Of course, constraints abound in theory construction as well. For example, constraints that regulate the abductive generation of new theories include methodological guides (e.g., "Researchers should give preference to theories that are simpler and have greater explanatory breadth"), aim-oriented guides (e.g., "Theories must be of an explanatory kind that appeals to latent causal mechanisms"), and metaphysical principles (e.g., "Social psychological theories must adopt a rule-governed conception of human behavior").

An orthodox empiricist reconstruction of scientific problems as constraints would normally take them to comprise those constraints that regulate the testing of theories for their empirical adequacy, where empirical adequacy has to do with predictive success. In this view, scientific problems would be regarded as essentially empirical in nature. However, with the underdetermination of theories by empirical evidence occurring in all of ATOM's phases of theory construction, the realist researcher will naturally appeal to conceptual as well as empirical constraints. For example, a plausible nascent theory will have to satisfy one or more empirical constraints in the form of claims about phenomena, but it will also have to satisfy a set of conceptual constraints about its explanatory promise. The appeal to conceptual criteria is also a natural way to deal with underdetermination in the context of theory appraisal. For example, the theory of explanatory coherence promoted in chapter 5 takes explanatory breadth as its criterion of empirical adequacy, but it also appeals to the criteria of simplicity and analogy to make effective judgments about the best of competing theories.

The importance of research problems, viewed as sets of constraints, resides in the fact that they function as the "range riders" of inquiry and thereby provide ATOM with the operational force to guide research. As just noted, the constraints themselves comprise relevant substantive knowledge as well as heuristics, rules, and principles. Thus the constraint-inclusion account of problems serves as a much-needed vehicle for bringing relevant background knowledge to bear on the various search tasks subsumed by ATOM. In turn, ATOM structures the methodological space within which the various constraints can operate. Given that ATOM is considerably broader in scope than either the inductive or the hypothetico-deductive accounts of scientific method, it canvasses a greater array of research problems than those methods do.

It is worth noting here that though I frequently talk of problem solving as a general aim of research, it is the *formulation* of problems that is the overriding concern of ATOM. The real challenge for researchers who adopt ATOM is to formulate ill-structured problems and better structure them so that they are capable of solution. With regard to science education, this focus on problem formulation is a desirable alternative to the prevalent practice of having students routinely obtain known solutions to well-structured problems as a way of learning disciplinary content. It also serves as a natural correction to the currently popular belief that teaching problem-solving skills is a panacea for overcoming uncritical thinking. By standardly presenting students with well-structured problems, science educators, in effect, formulate the problems for them and thereby provide them with ready solutions to the problems. However, articulating problems is a crucial part of the inquiry process, and it provides learners with highly appropriate opportunities to exercise their creative and critical intelligence.

6.3 Two Fundamental Commitments of ATOM

I now return briefly to two important methodological contrasts that were introduced in chapter 1 and discussed in chapter 2, because they are part of the deep structure of ATOM. These contrasts are generative and consequentialist methodology, and reliabilist and coherentist justification. I have suggested that consequentialist strategies justify knowledge claims by focusing on their consequences. By contrast, generative strategies justify knowledge claims in terms of the processes that produce them. Although consequentialist strategies are used and promoted more widely than generative strategies in contemporary science, both types of strategy are required in an adequate conception of research methodology. Two important features of ATOM are that the methodology promotes both generative and consequentialist research strategies for the detection of phenomena, and it promotes generative research strategies in the construction of explanatory theories.

Consequentialist reasoning receives a heavy emphasis in psychological research through the use of hypothetico-deductive method, often in tandem with null hypothesis significance testing. Consequentialist methods reason from the knowledge claims in question to their testable consequences. As such, they confer a retrospective justification on the theories they seek to confirm. In contrast to consequentialist methods, generative methods reason from warranted premises to an acceptance of the knowledge claims in question. Exploratory factor analysis is a good example of a method of generative justification. It affords researchers generative justifications by helping them reason from established correlational data patterns to the rudimentary explanatory theories that the method generates. As noted earlier, judgments of initial plausibility constitute the generative justifications afforded by exploratory factor analysis. Generative justifications are forward looking because they are concerned with heuristic appraisals of the prospective worth of theories.

In addition to embracing both generative and consequentialist methodologies, ATOM uses two distinct theories of justification. One of these, reliabilism, asserts that a belief is justified to the extent that it is acquired by reliable processes or methods. ATOM makes heavy use of reliability judgments because they furnish the appropriate type of justification for claims about empirical phenomena.[2] For example, as noted in chapter 2, statistical resampling methods and the strategy of constructive replication, are different sorts of consistency tests through which researchers seek to establish claims that data provide reliable evidence for the existence of empirical phenomena.

By contrast with reliabilism, coherentism maintains that a belief is justified in virtue of its coherence with other accepted beliefs. ATOM also uses coherentist justification (albeit of a special kind), where its approach to theory appraisal is governed by considerations of explanatory coherence.

I should emphasize that although reliabilism and explanationism are different and are often presented as competitors, one can view them as complementary theories of justification. ATOM adopts a broadly coherentist perspective on justification that endorses both reliabilism and explanationism and provides for their interaction. ATOM enjoins researchers first to justify claims about phenomena in terms of reliability considerations, and then to fashion explanatorily coherent theories that will account for the phenomena. Thus, when using the theory of explanatory coherence, one is concerned with delivering judgments of explanatory coherence, but the theory's principle of data priority presupposes that the relevant empirical generalizations have been justified on reliabilist grounds.

Further, the acceptability of claims about phenomena will be enhanced when they coherently enter into the explanatory relations that contain them. Alternatively, the explanatory coherence (specifically the explanatory breadth) of a theory will be reduced as a consequence of rejecting a claim about a relevant phenomenon that was initially accepted on insufficient reliabilist grounds.

6.4 Phenomena Detection and Theory Construction Again

The preceding exposition of ATOM prompts the following remarks about the tandem processes of phenomena detection and theory construction.

Successfully detecting a phenomenon is a major achievement in its own right and is a significant indicator of empirical progress in science.

In fact, the importance of phenomena detection in science is underscored by the fact that more Nobel Prizes are awarded for the discovery of phenomena than for the construction of explanatory theories. From the perspective of ATOM, theoretical progress is to be understood in terms of the goodness of explanatory theories as determined by the theory of explanatory coherence. Research methodology in psychology has placed a heavier professional emphasis on describing empirical regularities than on constructing explanatory theories, though the philosophy of science, until recently, has focused on a theory-centered view of science. However, I know of no good argument that supports the conclusion that one of these endeavors is more important than the other. Accordingly, ATOM takes phenomena detection and theory construction to be of equal worth.

The characterization of phenomena given in chapter 2 helps correct a widely held misunderstanding of science: taking the standard twofold distinction between observation and theory to be of fundamental methodological importance prevents one from being able to conceptualize properly the process of phenomena detection. This holds whether or not one subscribes to a hard-and-fast observation-theory distinction, or whether one accepts a relative observation-theory distinction and the ambiguous idea of theory ladenness that goes with it. To correctly understand the process of phenomena detection, one needs to replace the observation-theory distinction with the threefold distinction between data, phenomena, and theory.

ATOM's account of theory construction is at variance with the way many behavioral scientists understand theory construction in science. Most behavioral scientists use or at least endorse a view of theory construction that is strongly shaped by the guess-and-test strategy of the hypothetico-deductive method. In contrast with this prevailing conception of scientific method, ATOM asserts that theory generation can be a logical or rational affair, where the logic takes the form of abductive reasoning. It insists that theory development is an important part of theory construction—an undertaking that is stifled by a hypothetico-deductive insistence on immediate testing. And it maintains that empirical adequacy, understood as predictive success, is not by itself an adequate measure of theory goodness, there being a need to use additional virtues that focus on explanatory worth.

ATOM's three phases of theory construction have varying degrees of application in the behavioral sciences. Codified methods that generate theories through existential abduction are rare. The use of exploratory factor analysis to postulate common causes is a striking exception, although as remarked in chapter 3, the explicit use of this method as an abductive generator of elementary plausible theory is rarely acknowledged. As I suggest in the next section, grounded theory method (e.g., Strauss, 1987), which is increasingly used in behavioral research, can be understood as an abductive method that helps generate theories to explain the qualitative data patterns from which they are derived. However, grounded theory does not confine itself to existential abduction, and it imposes weaker constraints on the abductive reasoning permitted by the researcher than does exploratory factor analysis. The earlier suggestion that, as human beings, we have an evolved cognitive ability to abductively generate hypotheses leads to the plausible suggestion that scientists frequently reason to explanatory hypotheses without using codified methods to do so. Two prominent examples in the behavioral sciences are Noam Chomsky's (1972) publicly acknowledged abductive inference to his innateness hypothesis about universal grammar, and Howard Gardner's (Walters & Gardner, 1986) self-described use of "subjective factor analysis" to postulate his multiple intelligences. Also, it is likely that behavioral scientists use some of the many heuristics for creative hypothesis generation listed by William McGuire (1997) to facilitate their abductive reasoning to hypotheses.

Researchers in psychology and other behavioral sciences often hypothesize latent causes to explain behavioral phenomena. The challenge of learning about the mechanisms of these hidden causes is sometimes met by employing a strategy of analogical modeling. Unfortunately, the behavioral sciences seldom deal with such a strategy in their methodology and science education practices. Given the importance of such a strategy for the expansion of explanatory theories, methodologists in the behavioral sciences need to promote analogical modeling as vigorously as they have promoted structural equation modeling. Structural equation modeling provides knowledge of causal networks. As such, it does not so much encourage the development of detailed knowledge of the nature of the latent variables as it specifies the range and order of causal relations into which such variables enter. By contrast, analogical modeling seeks to provide more detailed knowledge of the causal mechanisms by enumerating their components and activities. These forms of modeling are different but complementary.

Inference to the best explanation is an important approach to theory appraisal that has not explicitly been tried in the behavioral sciences. Instead, hypothetico-deductive testing for the predictive success of hypotheses and theories holds sway. The theory of explanatory coherence, which is a well-codified method of inference to the best explanation, can be used in domains where two or more reasonably well-developed theories provide explanations of relevant phenomena. By acknowledging the centrality of explanation in science, one can use this method to appraise theories with respect to their explanatory goodness. I hope that behavioral science education will soon add the theory of explanatory coherence to its concern with cutting-edge research methods.

6.5 Two Applications of ATOM

In describing ATOM in the preceding chapters, I have presented it as a framework theory for assembling an array of more specific research methods into a coherent whole. I now provide an overview of two further applications of ATOM: first as a means of reconstructing grounded theory method, and second as the basis for creating an integrated model of clinical reasoning and case formulation. To the extent that these applications are judged successful, they will add to the heuristic worth of ATOM.

6.5.1 A Reconstruction of Grounded Theory Method

The most popular perspective on how to conduct qualitative research in the behavioral and social sciences is known as *grounded theory methodology*. It was introduced in the 1960s by the American sociologists Barney Glaser and Anselm Strauss and has been developed considerably by them and others since that time (e.g., Glaser & Strauss, 1967; Glaser, 1978; Strauss, 1987; Strauss & Corbin, 1998). Grounded theory is employed today by researchers in a variety of disciplines, including sociology, nursing studies, education, and management science. It has a growing influence in psychology, where it is still very much a minority practice.

The grounded theory perspective comprises a distinctive methodology, a particular view of scientific method, and a set of procedures for analyzing data and constructing theories. The methodology provides a justification for undertaking qualitative research as a legitimate—indeed, rigorous—form of inquiry. The original grounded theory conception of scientific method depicts research as a process of inductively generating theories from closely analyzed data. The specific procedures used in grounded theory make up an array of coding and sampling procedures for data analysis and a set of interpretive procedures that assist in the construction of theory. Grounded theory emerges from, and is grounded in, the data. In using these data analytic and interpretive procedures, grounded theorists are expected to meet the established canons of doing good scientific research, such as reproducibility, generalizability, and consistency.

Grounded theory has been presented from a number of philosophical positions.[3] Glaser adopts a general empiricist outlook on inquiry, one leavened by pragmatism, not positivism, as Glaser's critics sometimes mistakenly suppose. Strauss, by contrast, came to prefer a social constructionist position. In contrast with the originators of grounded theory methodology, I offered a reconstruction of grounded theory methodology from a scientific realist standpoint (Haig, 1996). Specifically, I formulated this account of grounded theory as a version of ATOM. Accordingly, we can best regard grounded theory as a broad theory of scientific method concerned with detecting and explaining social and behavioral phenomena. To this end, grounded theory is reconstructed as a problem-oriented endeavor in which theories are abductively generated from robust data patterns, elaborated through the construction of plausible models, and justified in terms of their explanatory coherence.

Glaser and Strauss clearly recognize the importance of understanding method as a problem-solving endeavor. However, although they offer some thoughtful remarks about research problems (Schatzman & Strauss, 1973), they do not give the matter systematic attention. The constraint-inclusion theory of problems employed in ATOM can be adopted by grounded theorists to regulate inquiry. Moreover, this theory of problems helps correct two misconceptions of problems that are evident in writings on grounded theory: the beliefs that problems and method are separate parts of inquiry, and that methods come before problems in a fixed order.

By repeatedly suggesting that theories are grounded in the data, Glaser and Strauss fail to heed the threefold distinction between data, phenomena, and theory. The idea that claims about phenomena, not data, are the appropriate objects of explanation is as relevant to grounded theory methodology as it is to scientific methodology generally. In addition, Glaser and Strauss's general plea to grounded theorists to check their data can be strengthened by acknowledging the important idea of robustness and the concomitant need to reliably establish phenomena in multiply determined ways before they begin to generate grounded theory.

In breaking from hypothetico-deductive orthodoxy, Glaser and Strauss argue that grounded theory emerges inductively from the data. However, the specific nature of the inductive relation that grounds emergent theories in their data is difficult to fathom. For Glaser and Strauss, grounded theory emerges inductively from its data source in accordance with the method of constant comparison. As a method of discovery, the constant comparative method is an amalgam of systematic coding, data analysis, and theoretical sampling procedures. These procedures enable the researcher to make interpretive sense of much of the diverse patterning in the data by developing theoretical ideas at a higher level of abstraction than the initial data descriptions. However, the notion of constant comparison contributes little to figuring out whether the inductive inference in question is enumerative, eliminative, or of some other form. Whatever Glaser and Strauss's view of the matter is, I think that the creative inference involved in generating grounded theory is better thought of as abductive in nature, whereas it is the reasoning from data to phenomena that involves inductive generalization.[4]

Glaser and Strauss hold a developmental perspective on theory construction. This is clear from their claim that "the strategy of comparative analysis for generating theory puts a high emphasis on theory as process; that is, theory as an ever-developing entity, not as a perfected product" (Glaser & Strauss, 1967, 32). In this regard, Glaser and Strauss advise the researcher to be constantly on the lookout for new perspectives that might help them develop their grounded theory, although they do not explore the point in detail. ATOM gives similar advice, but in a more constructive way: because we often do not have knowledge about the causal mechanisms that we abductively probe, we are urged to construct models of those mechanisms by imagining something analogous to mechanisms whose nature we do know. More specifically, theory elaboration in science is frequently a matter of constructing iconic paramorph models through analogical reasoning. There is much to be said for incorporating this perspective on theory development into grounded theory method.

Although Glaser and Strauss do not articulate a precise account of the nature and place of theory testing in social science, they do clarify that theory appraisal involves more than testing for empirical adequacy. Clarity, consistency, parsimony, density, scope, integration, fit to data, explanatory power, predictiveness, heuristic worth, and application are all mentioned as pertinent evaluative criteria. However, Glaser and Strauss do not expound on these criteria, let alone work them into a coherent view of theory appraisal. As with ATOM, inference to the best explanation (specifically, the theory of explanatory coherence) offers the grounded theorist an integrated account of two of the evaluative criteria that Glaser and Strauss deem important for theory appraisal.

ATOM provides a framework for inquiry that takes advantage of realist methodological work on research problems, generative methodology, and coherence justification. These are methodological notions that should be congenial to grounded theorists. Viewed from the perspective of ATOM, we should say that explanatory theory is grounded in phenomena, not data. Moreover, we can reasonably regard ATOM itself as a grounded theory of sorts, one that accommodates both quantitative and qualitative outlooks on research.[5]

6.5.2 Clinical Reasoning and Case Formulation

The scientist-practitioner model of clinical psychology is the most widely used model of professional practice in the Western world today. The model is most commonly satisfied by applying the evidence-based findings of psychological research to clinical practice. However, an important additional way in which the scientist-practitioner model can be realized is to conduct systematic inquiries into clients' problems in a manner that is guided by scientific method. With this approach, clinicians describe and formulate their clients' problems by focusing on their onset, development, and maintenance. To this end, they attempt to systematically collect data that enable them to identify clients' difficulties and their causes. The result of this process is a conceptual representation of each client's various complaints, their causes, and their interrelationships, which clinicians use as a basis to plan and execute treatment in a systematic and effective manner.

Clinical reasoning and case formulation lie at the heart of the work of scientifically oriented clinical psychologists, and from the 1970s onward, researchers have made concerted attempts to understand the nature of clinical reasoning (e.g., Borleffs, Custers, van Gijn, & ten Cate, 2003; Elstein, Shulman, & Sprafka, 1978; Schmidt, Norman, & Boshuizen, 1990) and to apply models of decision making to clinical reasoning (e.g., Galanter & Patel, 2005; Ward, Vertue, & Haig, 1999).

The standard view sees clinical reasoning as the set of decision-making or problem-solving processes employed in describing health problems. The goal of this enterprise is diagnosis, which, in turn, directs treatment. By contrast, a case formulation is the narrative that integrates the description and explanation of health problems. The primary goal of case formulation is to identify causal mechanisms that guide treatment decisions. Clinical psychologists not only describe their clients' functioning but also typically try to understand the causes of their clients' behaviors (Butler, 1998; Garb, 2005). Thus their work involves clinical reasoning (a descriptive process traditionally understood to lead to diagnosis) and case formulation (an explanatory process leading to understanding the causes of the diagnosis and the integration of both in narrative form).

Vertue and Haig (2008; see also Ward, Vertue, & Haig, 1999) argued that the extant literatures on clinical reasoning and case formulation are fragmented and do not provide a broad, coherent perspective that clinical psychologists can use across different theoretical orientations. We further contended that the hypothetico-deductive and Bayesian methodologies cannot provide an adequate framework for clinical reasoning. The hypothetico-deductive method is a weak method of problem solving because, among other things, it operates without regard for relevant background knowledge (Patel, Arocha, & Zhang, 2005). A major problem with the Bayesian alternative is that clinicians do not typically have access to the probabilistic information required for the effective use of Bayes's theorem. Partly in response to these problems, we argued that ATOM provides a suitably broad framework that integrates clinical reasoning and case formulation and can be used by clinicians of varying theoretical orientations (Vertue & Haig, 2008). We maintained that ATOM provides a systematic, coherent, and natural way in which clinical psychologists can reason in diagnosing and formulating a client's psychological difficulties. We showed that, with appropriate supplementation, the method provides a plan of inquiry that can guide the clinician in the reasoning processes involved in developing accurate descriptions of problems, constructing explanations for those problems, and establishing coherent models of the causal mechanisms involved.

From the vantage point of ATOM, the clinical reasoning process is centrally concerned with both the detection of empirical phenomena and their subsequent explanation. However, given that ATOM is a theory of method developed for basic psychological research, it is necessary to add two methodological phases to its standard depiction to complete its suitability for clinical applications. First, ATOM addresses neither the process of data collection nor the process of case formulation. Although ATOM does not deal directly with the methodology of data collection, this is clearly a critical aspect of both scientific research and clinical practice. Second, just as writing up scientific research is an integral part of that research, so writing the case formulation is an integral part of clinical work. However, these two processes can straightforwardly be grafted onto ATOM to produce a comprehensive model of clinical reasoning and case formulation, with data collection as a precursor to ATOM, and the narrative of the case formulation as a successor to ATOM.

In the proposed six-stage model of clinical reasoning and case formulation, the clinician begins with data collection and then proceeds through the four primary phases of ATOM before concluding with a written case formulation. In the beginning, potentially relevant data are gathered using a number of data generation strategies. In turn, the clinician conducts a generic interview to establish a base set of information about a client's functioning across a number of domains; elicits further data that are guided by the nature of the client's referral question; and identifies salient cues or flags that prompt the clinician to probe for possible phenomena associated with the presenting problems. All the while, various steps are taken to ensure that the data are reliably obtained. Regarding data analysis, ATOM's multistage model is systematically worked though in as thorough a manner as possible. Thus systematic attention is given in turn to data quality, pattern suggestion, pattern confirmation, and generalization. For example, to ensure the extent to which the phenomenon claims generalize, constructive replication is sought with respect to different life settings (e.g., home, work, and recreation) and across time (e.g., during childhood, adolescence, early adulthood, the past six months, or the past two weeks). Here the degree to which different, independent sources of information converge on the same conclusion constitutes an important validation strategy. The clinician draws this information from his or her professional database of symptom knowledge and matches the client's current data patterns to that knowledge.

Having identified the empirical phenomena relevant to the client, the next phase involves abductively inferring the psychological causes believed to produce those phenomena. Here it is useful to think of the causes as constituting the psychological makeup of the person, or their psychological strength and vulnerability factors. These causes also have contributing causal conditions, which may be distal, such as heritability, organicity, and learning history, as well as proximal factors, such as stress from a parent's remarriage, or a child starting school. An adequate explanation of the client's difficulties will also need to refer to maintaining factors, including environmental factors.

When a number of plausible explanatory hypotheses have been abductively generated, the next task is to ensure that they are developed to an acceptable degree. Sometimes the research literature or previously formulated cases will present explanatory hypotheses that are at an acceptable level of theoretical development. At other times, the clinician will take responsibility for developing the content of the initial hypotheses about the presence of the causal factors. For the clinician, the major task in developing a causal model is to establish the relationships between these causes in the model.

Once the various relationships are depicted in the causal model, the clinician considers the most coherent way of conceptualizing the client's situation. The developed causal model is evaluated according to its ability to account for the interrelationships between the psychological causes and their phenomena in an explanatorily coherent manner. This is a particularly crucial part of the clinical reasoning process, and it is frequently underemphasized. Within ATOM, the multicriterial perspective on theory appraisal suggested by the theory of explanatory coherence provides an instructive guide. Thus the criteria of explanatory breadth, simplicity, and analogy receive explicit consideration in evaluating the causal model.

Finally, the clinician uses information from the preceding phases to write a narrative that constitutes the case formulation. A case formulation is the culmination of the clinical reasoning process and is a comprehensive and integrated conceptualization of a case, encompassing the phenomenology, etiology, maintaining factors, prognosis, and treatment recommendations. The formulation is a set of descriptive and explanatory hypotheses that attempts to explain why a client developed these problems at a particular time, what maintains them, and what should be done about them (Ward, Vertue, & Haig, 1999). The case formulation should demonstrate an accurate and insightful understanding of a unique individual, with vulnerabilities and strengths, and explain how he or she comes to be in the current predicament. The essential task in case formulation is to highlight and make explicit links or connections between different components of the case.

By and large, my concluding remarks in the previous section about the benefits of ATOM for restructuring grounded theory method apply to the suggested rethinking of clinical reasoning and case formulation. However, the bottom-up thrust of ATOM-based inquiry will be a challenging framework for the many clinicians who are used to thinking in accordance with the top-down nature of hypothetico-deductive reasoning.

6.6 ATOM Defended and Clarified

I turn now to a defense and clarification of ATOM. First I defend ATOM against the charge that its view of theory construction is too flexible.

Then I consider some misunderstandings of ATOM that result from fudging the threefold distinction between data, phenomena, and theory.

6.6.1 Is ATOM Too Permissive?

Jan Willem Romeijn (2008) has undertaken a philosophical evaluation of ATOM as it appeared in Haig (2005a, 2005b). Although he judges ATOM's broad framework to be on the right track, he nevertheless thinks that the method is too permissive. Restricting his attention to the theory construction phases of ATOM, Romeijn judges its three components of theory generation, theory development, and theory appraisal to suffer from a problem of underdetermination by empirical evidence. In this regard, he claims that generating theories by exploratory factor analysis leaves us with a superabundance of hypotheses. Further, he claims that the strategy of analogical modeling is underspecified and imposes too few constraints on the process of whittling down this overabundance. Finally, he contends that the evaluation of explanatory theories in terms of their explanatory coherence suffers from two well-known objections that have been raised against inference to the best explanation. Moreover, Romeijn does not think that my combination of these three components of theory construction overcomes the problem of underdetermination. Because of this, he concludes that ATOM has insufficient normative force, and he briefly suggests ways in which this problem might be overcome.

In this section, I examine Romeijn's contention that ATOM's account of theory construction suffers from the problem of the underdetermination of theories by empirical evidence. Roughly speaking, the basic idea of the underdetermination at issue here is that the relevant empirical evidence does not determine the acceptance or rejection of a scientific hypothesis or theory. That is, we do not have the ability to justify accepting one hypothesis or theory from a set of alternatives on the basis of empirical evidence alone. Many philosophers regard such underdetermination to be a serious methodological problem for science. This attitude contrasts with the prevailing view in science, which is that there is no such problem, or if there is a problem, then science has the ability to solve it. My view of this matter is that good scientific practice is often able to exploit appropriate resources that enable scientists to deal with the underdetermination of theories by evidence. Undue philosophical concern about it results, in part, from a tendency of philosophers to underestimate the resources that scientists have at their disposal in determining theory choice (see, e.g., Kitcher, 1993). Following Romeijn's order of treatment, I focus in turn on the issue of underdetermination as it affects the method of exploratory factor analysis, the strategy of analogical modeling, and the theory of explanatory coherence. I will endeavor to show that in each case, sufficient methodological resources are available for scientists to use these methods to good effect. Toward the end of the section, I offer a number of remarks about the normative force of ATOM.

In Defense of Exploratory Factor Analysis ATOM characterizes the process of theory generation as existential abduction. As its name suggests, this form of explanatory reasoning postulates the existence, but not the nature, of new objects or properties. Cases abound in science where hypotheses about new entities have been introduced in this way. Although ATOM is a framework theory in which no one research method provides a general route to new theories, I chose to explicate the nature of theory generation by focusing on psychology's method of exploratory factor analysis—a method that I take to be a rather stylized way of producing existential abductions (Haig, 2005b).

Romeijn worries that exploratory factor analysis suffers from various problems of underdetermination and thus leaves us with an unacceptably large number of latent common factor models. My view is that, despite these underdeterminations, exploratory factor analysis is able to bequeath us a manageable number of plausible factorial hypotheses, which, in ATOM, are subjected to further scrutiny through analogical modeling and judgments of inference to the best explanation. However, Romeijn thinks these additional epistemic appraisals are insufficiently constraining, and recommends replacing exploratory factor analysis with a strategy of experimental intervention, which he thinks will resolve the problem of underdetermination at this point in the research process.

In chapter 3, I suggested that the alleged problem of factor indeterminacy is a special case of the general problem of the underdetermination of theory by evidence (Haig, 2005b; see also Mulaik, 1987). I argued further that if we have appropriate expectations about what exploratory factor analysis can do as a method of theory generation, then we are entitled to think that exploratory factor analysis is not undermined by this particular indeterminacy problem.

However, an important question still remains, a question that may be more important than Romeijn's worry about the indeterminacies of exploratory factor analysis: is the method effective enough in unearthing the common causes it hypothesizes to exist behind the correlated manifest variables? An answer to this question lies at the heart of my defense of the method. I maintain that if exploratory factor analysis proves to be a useful method of generating hypotheses about common causes, then Romeijn's concerns about the various sorts of underdetermination to be found in exploratory factor analysis cannot be too unsettling for the method.

I have two ways of answering this question. One is to examine research programs of theory construction that make heavy use of exploratory factor analysis, and show that the method contributes to the theoretical progress of those programs. We might want to ask, for example, whether the Spearman-Jensen theory of general intelligence is a progressive research program or whether the five-factor theory of personality is currently progressive. This approach would require detailed analyses of the relevant case histories, employing notions of theoretical progress that were or are appropriate to both science generally (a contested matter) and factor analysis more specifically. Space limitations at this point in the book preclude undertaking such a task, and I confine my attention briefly to the second strategy. This strategy involves ascertaining whether exploratory factor analysis succeeds at dimensional recovery as revealed through simulations on artificial data sets where the dimensions of the objects of study are known in advance.

The simulation studies that have been carried out to assess the reliability of exploratory factor analysis in dimensional recovery give mixed results. Some studies support the utility of the method, whereas others show poor dimensional recovery. Consider Armstrong's (1967) influential and widely cited study, which questions the utility of exploratory factor analysis as a method of theory generation. Armstrong analyzed a set of artificial data in a hypothetical scenario where the underlying factors were known, and he concluded from the analysis that exploratory factor analysis did a poor job of recovering the known factor structure. From this he recommended that the method should not be used to generate theories. Subsequently many authors have cited Armstrong's article as grounds for using confirmatory factor analysis rather than its exploratory counterpart in factor analytic research.

However, Preacher and MacCallum (2003) have argued, correctly in my view, that Armstrong's (1967) study represents a poor piece of factor analytic research that gives misleading results, and it provides no real basis for casting doubt on the worth of exploratory factor analysis as a method of theory generation. Preacher and MacCallum's study first replicated Armstrong's factor analysis on an analogous set of data and obtained essentially the same results. They then conducted a further factor analysis of that data set, substituting correct factor analytic procedure for the faulty procedure used by Armstrong. Among other things, this involved using common factor analysis rather than principal components analysis (principal components analysis is not really a method of factor analysis), determining the correct number of factors to retain by employing appropriate multiple methods (the scree test and parallel analysis), and using oblique direct quartimin rotation to simple structure rather than orthogonal varimax rotation. Based on the congruence between the obtained factor pattern and the known structure, Preacher and MacCallum concluded that the proper use of exploratory factor analysis does in fact identify the number and nature of latent variables responsible for the manifest variables. Their exemplary use of exploratory factor analysis and the well-conducted earlier simulations by factor analysts such as Thurstone (1947) and Cattell (Cattell & Dickman, 1962) provide good support for the view that exploratory factor analysis is quite effective at dimensional recovery. Admittedly, these simulations dealt with simple physical systems, but Sokal, Rohlf, and Zang (1980) have shown that exploratory factor analysis can isolate and help identify meaningful biological factors that lie behind correlated physiology-of-exercise variables. The findings from good simulation studies like these, combined with those of a variety of empirical studies on other aspects of the functioning of exploratory factor analysis (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999), suggest that the method can be employed as a useful generator of elementary plausible theories about common causes.

The Strategy of Analogical Modeling As we saw in chapter 4, models serve a variety of functions in scientific research. In ATOM, they play a major role in expanding the rudimentary theories given to us by existential abductive methods, such as exploratory factor analysis. As chapter 4 also showed, this increase in the content of theories is sought through the strategy of analogical modeling, which researchers accomplish by building analogical models of the hypothesized causal mechanisms. The content of the undeveloped theory is expanded by analogy to a well-understood source model, and at the same time, the credibility of the model is provisionally assessed through a process of analogical abduction.

Romeijn believes that this strategy of analogical modeling also suffers from a problem of the underdetermination of theories by empirical evidence because it gives too little guidance in the process of theory development. His criticism is not that ATOM seeks to employ analogical modeling as a strategy of theory expansion, but that it does not specify the notion of analogy in enough detail to prevent our being left with a superabundance of models.

I agree that considerably more needs to be said about the strategy of analogical modeling than the short treatment I gave it in an article-length presentation of ATOM (Haig, 2005a). In the article, I provided a general argument schema that represents the basic structure of the reasoning involved in analogical abduction. Of course, this schema does not fully capture the detailed reasoning required for effective analogical modeling in science. In evaluating the aptness of an analogical model, its structure has to be assessed, and this is done with respect to the aptness of the analogy between the source and subject of the model. In considering the plausibility of the source model, one considers the balance of the positive and neutral analogies. In identifying these analogies and ascertaining their balance, one has to appeal to domain-specific information relevant to the case at hand. Admittedly, there is a dearth of examples of the analogical modeling of explanatory theories in the behavioral sciences. However, as I noted in chapter 4, Harré (1976) gives an informative account of analogical modeling in the social sciences, and Harré and Secord (1972) detail the construction of a role-rule model of microsocial interaction in social psychology that is a source of useful guidance for psychologists.

Although I think that the methodology of analogical modeling is moderately well developed, it clearly needs further work. For example, the work on analogical modeling in cognitive science needs to be integrated with the relevant philosophical modeling in the philosophy of science (e.g., Abrantes, 1999). In addition, detailed case studies of successful analogical modeling in the behavioral sciences should be undertaken to identify exemplars and precepts of good modeling practice that we can use as a basis for further codifying the methodological strategy of analogical modeling.

My hope is that chapter 4 might partly allay Romeijn's worry. I do believe that the strategy of analogical modeling, combined with the constraints provided by methods such as exploratory factor analysis and the theory of explanatory coherence, will result in a manageable pluralism of model-based theories.

Problems for Explanatory Coherence With ATOM, theory appraisal is conducted by employing the theory of explanatory coherence to determine judgments of inference to the best explanation. Although inference to the best explanation is clearly used in science to evaluate scientific theories, and despite determined efforts to explicate its nature, inference to the best explanation has received considerable criticism. Romeijn believes that the theory of explanatory coherence is vulnerable to the two major objections that have been leveled against inference to the best explanation more generally: these have been called *Hungerford's objection* and *Voltaire's objection* (Lipton, 2004). Hungerford's objection is that the evaluative criteria that constitute explanatory goodness are too subjective to determine properly the warrant of inference to the best explanation. Voltaire's objection is that we have no good reason to suppose that sound judgments of the best explanation are likely to be true. Romeijn assumes for the sake of argument that Thagard's (1992) empirical justification of the theory of explanatory coherence takes care of Hungerford's problem. However, I think the justification works as a matter of fact. Not only are the criteria of explanatory coherence (explanatory breadth, analogy, and simplicity) derived from an examination of exemplary cases of theory appraisal in the history of science, but successful simulations of the theory of explanatory coherence by Thagard and his colleagues show that these criteria are successfully incorporated into one or more of the principles of the theory.

However, for Romeijn, Voltaire's problem remains. He reiterates the point that the method of exploratory factor analysis and the strategy of analogical modeling provide us with insufficient reason to think that they bequeath to the theory of explanatory coherence a set of theories that contain a true or truthful theory.

A number of philosophers have criticized proponents of inference to the best explanation for coupling it with truth and maintaining that an inference to the best explanatory theory entitles us to regard that theory as true (e.g., van Fraassen, 1989). Among other things, these critics have pointed out that the history of the various sciences reveals that many theories initially pronounced true on the grounds that they were judged the best of competing theories turned out to be manifestly false (e.g., magnetic ether in physics, phlogiston in chemistry, vital forces in physiology, and Hullian theory in psychology).

My reply to this criticism is the same as the response I gave in chapter 5 to van Fraassen's "bad lot" argument against inference to the best explanation. There I pointed out that inference to the best explanation can legitimately be used to evaluate theories with respect to their explanatory goodness in a way that avoids judgments of truth per se while at the same time regarding science as a truth-seeking endeavor. The crucial point here is that the assumption that one can secure truth by using inference to the best explanation disregards the important distinction between truth and justification. Truth, understood as correspondence with reality, functions as an orienting ideal for science. As such, it is a highly valued, though unattained, goal that helps us make sense of science as an attempt to represent and intervene in the world (Haig & Borsboom, 2012). However, as an ideal, truth (or, more precisely, approximate truth) is accessible only indirectly by way of the various criteria we use to evaluate and accept theories. Historically, scientists have regarded the criteria of predictive accuracy, internal consistency, and explanatory power as important in theory appraisal. As justificatory criteria, they can indicate truth, but they do not constitute truth.

A further point of relevance here is Thagard's (2007) claim that accepting a theory based on explanatory coherence does not mean that it is likely to be true, only that such acceptance is conducive to the long-term goal of maximizing true propositions and minimizing false ones. This line of reasoning is consistent with the endorsement of the idea of possible truth described in chapter 1: that it is realistic to nominate our theories as candidates for truth in the expectation that they will be true in the future, if not the present.

Contrary to Romeijn, I think that the three submethods and strategies employed in theory construction within ATOM can make worthwhile contributions to the development of scientific knowledge. Exploratory factor analysis has proved to be a moderately useful generator of explanatory hypotheses and theories. A number of sciences have successfully employed the strategy of analogical modeling, though its methodology is yet to be fully articulated and systematically used in developing psychological theories. And although the theory of explanatory coherence has not been used as a method of theory appraisal in psychology, it reconstructs an informal approach that has been successfully used in the physical and biological sciences.

If I am right in assessing the effectiveness of these three different parts of ATOM's account of theory construction, then their linking enhances the overall effectiveness of ATOM's prescriptions for theory construction; the initial plausibility judgments of hypotheses in exploratory factor analysis are augmented by judgments of the appropriateness of analogies in model-based theories, before theories are further evaluated in terms of their explanatory coherence. If this extended theory evaluation process goes well, then its outcome should be well-credentialed theories. I therefore conclude that the underdetermination of theories by empirical evidence does not pose a major problem for ATOM.

A Note on Experiments I have repeatedly emphasized that ATOM is to be understood primarily as a framework theory within which one can employ more specific methods. An important point to note here is that many of the methods adopted in the exposition of the framework are optional. For example, my description of phenomena detection in terms of statistical methods is not the only means by which one can detect phenomena. Moreover, although I chose exploratory factor analysis to describe the abductive nature of theory generation, the method is only appropriate when the abductive inferences are to common causes. Options like these give ATOM a degree of flexibility that researchers will want to exploit.

Romeijn accepts the general framework of ATOM but believes that the permissiveness permitted in its account of theory construction can be overcome by replacing my chosen submethods and strategies with a methodology of experimental practice. As Romeijn acknowledges, his suggestions in this regard are brief and speculative, and I find it hard to know what to make of them. I do know that the philosophical methodology of experimentation has made significant gains in the last three decades, and I would not be surprised to learn in the future that ATOM can be given an alternative formulation through experiments. Of course, this alternative rendering of ATOM would not be suitable for the many nonexperimental practices in psychological research.[6]

The Normative Status of ATOM Before concluding this section, I want to make some comments about the normative status of ATOM. The first of these involves a possible misunderstanding of what ATOM is. My other comments are intended to give some indication of how I understand ATOM's normative dimension.

In the introduction to his article, Romeijn (2008) speaks as though ATOM is a scientific methodology. He lists a number of prominent twentieth-century methodologies (e.g., Popper's falsificationism, Bayesianism, and Laudan's normative naturalism) and declares that they all determine a proper scientific method. However, mindful of the distinction between methodology (the study of method) and method (procedures of inquiry), I regard ATOM as a theory of method, not as a methodology. To be sure, ATOM explicitly draws from the literature on methodology (e.g., the ideas of reliabilist and coherentist justification), a point that I emphasized in chapter 1. It also presupposes a number of methodological commitments (e.g., a modified account of Laudan's normative naturalism). Further, as a theory of method, ATOM can be regarded as a part of methodology in the sense that it is an object of methodological scrutiny.

My pedantic insistence on the distinction between methodology and method has a point, for by regarding ATOM as a methodology, Romeijn expects more from it than I think one can reasonably expect from an account of scientific method. For example, and significantly, Romeijn asserts that any scientific methodology should be able to give an account of how the major philosophical problem(s) of induction can be resolved. For him, a resolution requires a philosophical basis for, and justification of, scientific facts. However, being a theory of scientific method, ATOM should not be expected to provide a philosophical justification for inductive inference. It is enough that it provides for the justification of scientific facts (and theories) in research practice.

Twentieth-century philosophers often claimed that scientific methodologies were known a priori and could therefore be presented as radically normative. Popper's falsificationist methodology is a prominent case in point. However, because ATOM presupposes a naturalistic conception of methodology, I do not think it can be so strongly normative. Because it is founded on a naturalistic conception of methodology, ATOM and its components should be tempered by appropriate evidential considerations. For many of the specific research methods that one can employ in ATOM, a developing empirical literature speaks to their effectiveness, but we need more research of this sort. Moreover, psychological researchers need to be more deliberately naturalistic in their methodological behavior and refer to the relevant empirical literature when justifying the methods they use.

Another restriction on the normative force of ATOM is imposed by the conditional nature of the recommendations for research action that accompany it. In effect, such recommendations are subjunctive conditionals that take the form "If you want to reach goal X, then use method or strategy Y." The justification for pursuing goal X rests with the researcher. It is not to be found within ATOM as it is currently formulated. The conditional nature of methodological recommendations is a feature of Laudan's (1996) normative naturalism, a methodology that has been recommended to psychologists by Proctor and Capaldi (2001a).

In an important sense, the normative potential of ATOM resides in its adoption of a problem-oriented view of scientific inquiry. In my exposition of ATOM, I stressed the point that it adopts an account of research problems that depicts them as sets of constraints on their solutions, where the task is to take ill-structured problems and better formulate them so that they are capable of solution. Viewed as sets of constraints, research problems function as the guides of inquiry. In this way, the constraint-inclusion account of problems serves as a vehicle for bringing relevant background knowledge to bear on its various research tasks. In Haig (1987) I provide a more detailed account of research problems and their role in an abductive conception of inquiry.

6.6.2 Some Misunderstandings of ATOM

In a recent article, Mark Orlitzky (2012) recommends a package of reforms designed to help to overcome psychology's heavy reliance on null hypothesis significance testing. One of these reforms involves placing a greater emphasis on abductive research methods. Although I agree with much of what Orlitzky has to say, I think his treatment of abductive methods contains some misunderstandings. Since he gives particular attention to ATOM, I want to correct these misunderstandings.

Somewhat surprisingly in my view, Orlitzky (2012) takes exploratory data analysis and computer-intensive resampling methods to be basically abductive in nature. However, as its name implies, exploratory data analysis is data analytic in character. As noted in chapter 2, it involves descriptive and frequently quantitative detective work designed to reveal structure or patterns in the data. For this reason, I do not think it can be considered an *explanatory* or *abductive* undertaking in any interesting sense of the terms. I made this same point against Behrens and Yu (2003) in a footnote to chapter 2. Computer-intensive resampling methods are also data analytic in character. They are confirmatory procedures designed to check the reality of the patterns revealed by exploratory data analysis. In ATOM, these methods are used to achieve close replication, not to further explanatory research. As such, they are part of the process of detecting empirical phenomena. By contrast, abductive inference is reserved for constructing explanatory theories, which are introduced to explain empirical phenomena.

Orlitzky (2012) also regards meta-analysis as abductive in nature. He demonstrates this by taking the argument schema for existential abductive inference that I laid out in my characterization of exploratory factor analysis in chapter 3 and instantiating it with a meta-analytic example. In the second premise of this schema, he inserts information of an explanatory kind that explains the empirical phenomenon described in the first premise. However, I do not think that the explanatory information is generated directly by the use of meta-analytic techniques; it is gained instead by abductively hypothesizing plausible causes, sometimes using something akin to informal factor analysis. Meta-analytic techniques are most frequently used to identify empirical phenomena, whereas suggested explanations for phenomena are fashioned abductively, with or without the help of codified methods.

I conclude this short section by briefly considering a related claim that is sometimes made about the explanatory reach of meta-analysis when it is used in theory testing. When meta-analysis enters into the process of testing explanatory theories, it typically does so by contributing to an evaluation of the predictive success of those theories. However, this common strategy of theory evaluation is not directly concerned with their explanatory adequacy. This is not to deny that researchers can employ meta-analytic methods when testing theories, but meta-analysis itself is not an explanatory approach to hypothesis testing. To employ meta-analysis to assist in the predictive testing of an explanatory theory does not thereby confer a direct explanatory role on meta-analysis itself. One does not assign genuine status simply on the basis of association.

6.7 Scientific Method and Education

In some of the preceding chapters, I have offered remarks about the proper place of various research methods in psychology's research methods curriculum. In the penultimate section of this chapter, I want to offer some general thoughts about the importance of an education in research methods.

To begin by considering the nature of education itself, I believe we should follow John Dewey's (1910) lead and embrace an inquiry-oriented conception of education, which accords a central place to scientific method. Scientific method is important to education for at least three reasons: it provides us with a codified way of learning how to learn; it enables us to justify our knowledge claims, both about empirical phenomena and about explanatory theory; and it is a central feature of science itself, which is an enterprise we seek to understand in education. If we accept an inquiry-centered view of education, it is a small step to think of education itself, and scientific research, as broadly the same type of endeavor, where both are essentially concerned with learning. In such a view, students are concerned with learning through inquiry, whereas the major concern of teachers is to lead less experienced inquirers into new areas of learning.

A view of education with scientific method at its heart will clearly emphasize learning about research methods, and their accompanying research methodology, in the science curriculum. This clearly happens in psychology, but considerable evidence suggests that students and researchers do not acquire a deep understanding of these methods. A striking example of this is the low level of understanding among professional researchers in psychology of null hypothesis significance tests. Despite repeated exposure to these procedures in taught courses, and their frequent use in psychological research, psychologists fail to properly understand the logic of the method (e.g., Gigerenzer, Krauss, & Vitouch, 2004; Hubbard, 2004). As we saw in chapter 3 and in section 6.6.1, exploratory factor analysis is another frequently used method in psychology that is not well understood, with respect to its abductive nature (Haig, 2005b) and its procedural implementation (Fabrigar et al., 1999).

As I have stated several times, we cannot have a proper understanding of research methods, both conceptually and procedurally, without a sound appreciation of their accompanying methodology. However, the majority of researchers in psychology are reluctant to think critically about the methodological foundations of the methods they use. Nor are students encouraged to do so in the research methods courses they take. As we saw in chapter 1, methodology is the interdisciplinary field that studies methods. Although it draws from the disciplines of statistics, philosophy of science, and cognitive science, the professional literature of these disciplines does not figure systematically in the content of research methods courses. For example, the philosophy of research methods is an aspect of research methodology that receives limited attention in behavioral science education. The majority of students and research practitioners in the behavioral sciences obtain the bulk of their knowledge about research methods from textbooks. However, a casual examination of these texts shows that they tend to pay little, if any, serious regard to the philosophy of science and its bearing on the research process.[7] As Thomas Kuhn (2012) pointed out more than fifty years ago, textbooks play a major role in dogmatically initiating students into the routine practices of normal science. Seriously attending to the philosophy of research methods would go a considerable way toward overcoming this uncritical practice (Proctor & Capaldi, 2001b). As contemporary philosophy of science increasingly focuses on the contextual use of research methods in the various sciences, let us hope that research methodologists and behavioral scientists will avail themselves of the genuine methodological insights that it contains.

A methods curriculum genuinely concerned with education would profitably consider methods in the light of three primary characteristics of realist methodology outlined in chapter 1. First, greater prominence would be given to generative methodology in which reasoning well to hypotheses and theories would figure in the assessment of those knowledge claims. I have already noted that sound abductive reasoning to factorial hypotheses using exploratory factor analysis, and the abductive generation of grounded theory, are concerned with generative justification. Second, the coherentist justification of explanatory theories using methods of inference to the best explanation would feature much more prominently than it does at present. Third, in adopting methods that are apt for us as knowing subjects, heuristic procedures would receive much more explicit attention in the methods curriculum as serviceable guides to our thinking than is currently the case.

The Association for Psychological Science now takes conceptual and historical issues as one of psychology's seven core areas, and it must be included in degree courses that are accredited by the society. Teaching methods through methodology is the appropriate way to employ this core area in research methods courses. The American Psychological Association and the Association of Psychological Science would do well to follow suit, for it is only by making full and proper use of methodology that we can achieve a genuine education in research methods.

6.8 Final Word

ATOM aspires to be a coherent theory of scientific method that brings together a number of different research methods and strategies that are normally considered separately in the behavioral sciences. The account of phenomena detection I have offered systematically reconstructs a set of practices that are common in science but seldom presented as a whole in methodological writings. That reconstruction is based on the important distinctions between data, phenomena, and theory and the different functions they serve in scientific research. The abductive depiction of theory construction endeavors to make coordinated sense of the way in which science sometimes obtains knowledge about the causal mechanisms that figure centrally in the understanding of the phenomena that they produce. With rare exceptions, the abductive generation of elementary plausible theory, the strategy of analogical modeling, and the method of inference to the best explanation are all yet to receive proper consideration in psychology and the other behavioral sciences. ATOM serves to combine these methodological resources in a broad theory of scientific method.

Although ATOM is a broad theory of scientific method, it should not be understood as a fully comprehensive account. ATOM is a singular account of method appropriate for detecting empirical phenomena and subsequently constructing explanatory theories, where those theories purportedly refer to hidden causes, and where their causes are initially given an indirect, dispositional characterization. However, in dealing with explanatory theories in which the causal mechanisms referred to are more directly accessible than theoretical entities, researchers do not have to use a strategy of analogical modeling to more informatively characterize their theories. The use of functional brain imaging techniques to map neuronal activity in the brain is an obvious case in point. Further, although the evaluation of theories in terms of explanatory criteria deserves a heavy weighting in science, inference to the best explanation will not always be an appropriate or a sufficient resource for evaluating theories. For example, although both scientific methodology and practice have probably overemphasized predictive success (Brush, 1995), it nevertheless remains an important criterion of a theory's worth. It may therefore be sought in a modified hypothetico-deductive strategy that corrects for the confirmational inadequacies of its simple form.

For the sake of consistency, ATOM has to be judged in a way that comports with a naturalist attitude in methodology. In general terms, this comes down to the question of whether ATOM is a genuinely coherent theory of method, and that question is yet to be properly answered. Although it is a fairly comprehensive account of method, and although it captures a natural order of scientific inquiry, and seems to hold together, further development and appraisal is required before we can properly judge ATOM's cohesiveness. My hope is that its current formulation stands as a positive contribution to behavioral research methodology, and that with further work, ATOM might be shown in a reflexive way to be an explanatorily coherent theory.

Notes

1 Method, Methodology, and Realism

1. In a sympathetic appraisal of Laudan's account of the transition from inductive to hypothetico-deductive method, Ernan McMullin (1984) took issue with some of its detail and emphasis. McMullin agreed with Laudan's central contention that since 1700 the philosophy of science had to face the fact that science increasingly appealed to theoretical entities. However, he maintained that the acceptance of the hypothetico-deductive method in the seventeenth century was prompted more by the "corpuscular philosophy" of thinkers such as Robert Boyle and John Locke than the successful use of the hypothetico-deductive method in science.

2. Bert Uchino, Dustin Thoman, and Sari Byerly (2010) sampled over 230 articles from the prominent *Journal of Personality and Social Psychology* from 1982 through 2005 and found that the large majority of articles favored a testing strategy of confirmation. Considerably fewer favored a strategy of falsification, and even fewer favored a strategy of employing crucial tests of multiple hypotheses or theories. These findings square with the author's casual impressions and speak against the claim sometimes made that Popperian falsification is psychology's hypothetico-deductive method of choice.

3. Of course, there are other prominent accounts of scientific method. Two of the best known are T. C. Chamberlin's (1965) method of multiple working hypotheses and John Platt's (1964) advocacy of strong inference. Although they promote important ideas (theoretical pluralism and strong tests, respectively) and receive regular endorsement by methodologists, they seem to have had a limited influence on scientific practice. O'Donohue and Buchanan (2001) provide a thoughtful critique, written for psychologists, of Platt's theory of strong inference.

4. Strictly speaking, the claim that there cannot be a logic for discovering hypotheses is a corollary to the hypothetico-deductive method, not a part of it. Some descriptions of the method speak about the amethodological formulation of

hypotheses to explain the data. However, because hypothesis generation is not part of the method proper, I do not include it in my description and discussion of the method.

5. Erotetic logic, the logic of questions, is the obvious source for a theory of questions, but in my view it is too formal to be readily applicable to most of our scientific problems or to help researchers directly. However, this is not to deny that models of interrogative inquiry may give us some useful insights about inquiry processes generally.

6. More recently, Wimsatt (2007) extended his list of the important properties of heuristics. In addition to the four just mentioned, he noted that heuristics are purpose relative (they are useful for something) and are also derived with modification from other heuristics to better perform a new role.

7. Preliminary results from a 2009 PhilPapers survey of over three thousand philosophers showed that 66 percent either accepted or leaned toward scientific realism, whereas 18 percent favored scientific antirealism. This is in keeping with results from the same survey on a number of more specific philosophical categories, which favored a naturalistic metaphilosophy, a nonskeptical realism about the external world, a correspondence view of truth, and a non-Humean conception of laws.

8. The term *causal mechanism* is ambiguous. In ATOM, the generation of theories involves explanatory inference to claims about the existence of causal entities. It is not until the subsequent development of these theories that the mechanisms responsible for the production of their effects are identified and spelled out. Also, ATOM assumes that the productivity of causal mechanisms is distinct from the regularities that they explain (Bogen, 2005; cf. Woodward, 2003). Importantly, this allows for the methodological use of generalizations that describe natural regularities to help identify the causal mechanisms that produce them.

9. Note, however, that the strategy of analogical modeling is essential for theory development in ATOM, and the theory of explanatory coherence does heavy-duty work in the theory because it is the best-developed method of inference to the best explanation currently available.

2 Detecting Psychological Phenomena

1. When contrasting explanatory theories with claims about phenomena, Bogen and Woodward focus on what they call *systematic theories*. For them, systematic theories properly explain phenomena by showing in detail how the phenomena result from the causal factors appealed to in their explanation, and by unifying, and therefore systematizing, the phenomena claims. Psychology seems to have few well-developed theories of this sort. Although it constructs theories of various kinds, most of them are modest theories with low, but genuine, explanatory power.

2. Bogen and Woodward's work on phenomena detection has received considerable attention in the philosophical literature. It has been endorsed, modified, and

used by Brown (1994), Kaiser (1991), Teller (2010), and Weber (2007), among others. It has also been subjected to criticism, most notably by Glymour (2000), McAllister (1997), and Schindler (2007). Woodward (2011) recently clarified and amended the original formulation of the data-phenomena distinction and defended it against a range of criticisms.

3. I think this practice masks the fact that a number of investigators often contribute to the detection of an empirical phenomenon and receive little or no recognition for it. The Flynn effect was so named by Herrnstein and Murray in *The Bell Curve* (1994). Rushton thinks it should be called the *Lynn-Flynn effect*, after Richard Lynn, who found the upward trend in IQ scores in modern Japanese society. However, as Flynn himself noted, Tuddenham provided clear evidence of large IQ score gains in a comparison of U.S. soldiers in the two world wars, and Flynn stated that if asked, he would have named the effect after Tuddenham. However, it was Flynn who did most of the hard work in establishing the generality of the effect that bears his name. Unfortunately, the practice of giving insufficient intellectual credit to all the people who played an important role in empirical discoveries and theory construction is widespread in science.

4. Cummins (2000) contended that capacities are the primary *explananda* of psychology, whereas empirical regularities are *explananda* of secondary importance. Shapiro (1994) went further and claimed that cognitive psychologists take cognition, not behavior, to be the domain of their true *explananda*. However, I think that regularity phenomena are pursued more frequently, and are generally accorded greater importance, by psychological researchers. Furthermore, not all phenomena have to be detected. As Cummins remarked, phenomena that take the form of capacities are often known to us. In cases such as these, the task is to not to discover the phenomena but to provide an informative specification of them.

5. There are a few exceptions, most obviously when scientists look to explain why a study does not give the expected results, for example, when they suspect that the data are erroneous because they are produced by a faulty instrument.

6. In this chapter, I give little attention to the problem of phenomena decay. However, I do point out that a host of meta-analytic findings supports the view that the behavioral sciences have produced a good number of durable generalizations. It seems that the Flynn effect has ended (and may be in decline) in a few advanced nations (Teasdale & Owen, 2005). However, the effect has persisted for some decades and continues to do so in many countries. Therefore time enough has passed to construct plausible explanatory theories of this effect.

7. Of course, that is not to say that single events, such as the extinction of the dinosaurs, are not the objects of serious scientific investigation.

8. This will not always be the case in science. As Denny Borsboom pointed out to me, self-organizing complex systems produce phenomena that result from many causal influences.

9. In presenting this example, Bogen and Woodward (1988) referred to Ernest Nagel's (1961) discussion of the melting point of lead and indicated a number

of errors in his understanding that resulted from not clearly adhering to the data-phenomena distinction.

10. Bogen (2010) provides an example of phenomena detection where conclusions drawn about brain function analyze pink noise, understood as din, without extracting a signal from it. It is noise, understood as interference, from which signals are extracted.

11. With one exception, the strategies considered here are those discussed by Woodward (1989). Franklin (1990) provided an instructive discussion of an overlapping set of strategies for validating experimental results in physics. Some of his procedures also involve the appeal to explanatory theory.

12. Strictly speaking, it is misleading to speak of common or intervening causes as *spurious* correlations. What we call *spurious correlations* are really genuine correlations, so their existence can hardly be denied by claiming that they are brought about by some underlying third variable (Haig, 2003).

13. For Chatfield, the initial analysis of data has much in common with Tukey's approach to exploratory data analysis, but it is more inclusive. Because these two related data analytic endeavors serve different primary functions (data screening and pattern detection, respectively), I restrict initial data analysis to the preliminary scrutiny of data that occurs before exploratory data analysis (in Tukey's sense) is undertaken.

14. Behrens and Yu (2003) suggested that the inferential foundations of exploratory data analysis lie in the notion of abduction. However, exploratory data analysis is a descriptive pattern detection process that is a precursor to the inductive generalizations involved in phenomena detection. By contrast, abductive inference is employed in the generation of theories that are introduced to explain empirical phenomena. Behrens and Yu's suggestion conflates description and explanation in this regard. That said, one should appreciate that, when describing phenomena, some of the background knowledge presupposed will be the product of abductive reasoning. The true score theory presupposed in determining the melting point of lead, which was mentioned earlier, is a case in point.

3 Theory Generation: Exploratory Factor Analysis

1. An important part of this controversy is the contested nature of the relationship between the methods of exploratory factor analysis and principal components analysis. It is not uncommon in the behavioral sciences to claim that exploratory factor analysis and principal components analysis are similar, conceptually speaking, but different in their mode of calculation. I think that this view of the relation between the two methods is mistaken. It stems from ignoring the relevant interpretive dimension of factor analytic methodology and regarding exploratory factor analysis as a data reduction method on a par with principal components analysis. However, the interpretive part of factor analytic methodology makes clear that exploratory factor analysis is a genuine latent variable method, whereas principal components analysis is a method of data reduction. Factor analysis and principal components analysis are, roughly speaking, computationally similar, but conceptually different (Bartholomew, 2004).

2. Two less-noticed, but still important, parts of the controversy about the methodological status of exploratory factor analysis are the claims that standard factor analysis presupposes that genuine measurement of quantitative structure is possible, and the method's conclusions apply to populations, not individuals. These claims deserve more consideration by factor analysts than they currently receive. A researcher who accepted the force of these claims could defensibly employ a nonmetric form of dynamic factor analysis. This would give the researcher a person-centered method that also avoided the problem of assuming the existence of quantitative structure.

3. My principal reason for assigning a theory generation role to exploratory factor analysis is based on the related beliefs that factors are best regarded as latent common causes and that inference to such causes is abductive in nature (Haig, 2005b).

4. The term *entity* is used as a catchall ontological term to cover a miscellany of properties that include states, processes, and events. Although in the first instance existential abductions in exploratory factor analysis are about properties expressed as the values of variables, not all existential abductions need take this form.

5. The *positive manifold* is a term that is sometimes used to refer to the striking and well-established fact that almost all different tests of ability correlate positively with one another to a significant degree. Despite its historical link to Spearman's theory of general intelligence, the positive manifold can be taken as evidence for the existence of multiple-factor theories of intelligence.

6. The phrase "the knowing subject" comes from Karl Popper (1972), who advocated an objective theory of scientific knowledge that did not refer to cognitive agents and their mental states. Popper's antipsychologism stands opposed to a plausible moderate psychologism in which psychology and cognitive science more generally play an important role in helping us understand how methods are apt for humans as inquirers.

7. Most methods contribute either to claims about empirical phenomena or to claims about explanatory theory and are not mixed methods in this sense. Exploratory factor analysis is unusual in this regard. Second, it is the custom in contemporary methodology to regard mixed methods as a research strategy that combines both quantitative and qualitative research methods. However, I think it is important to understand that a given method will often have both quantitative and qualitative dimensions. Exploratory factor analysis is a good case in point. Although it is standardly viewed as a multivariate statistical method, and therefore quantitative in nature, its centerpiece, the principle of the common cause, can effectively be understood in qualitative terms.

4 Theory Development: Analogical Modeling

1. Popper's falsificationist variant of the hypothetico-deductive method also eschews models. Like the logical positivists, Popper took models to be heuristic devices that belonged in the context of discovery where, in his understanding of that context, heuristics could not play a genuine methodological role.

2. Actually, the distinction here is threefold: the mathematical equations of exploratory factor analysis make up a syntactical model; the model's formal structure is given a methodological interpretation, for example, when the latent variables of the method are considered as markers for theoretical entities; and the methodologically interpreted latent variables are then given a semantic interpretation in discipline-theoretic terms.

3. My perspective on analogical modeling is influenced by Rom Harré's work on the topic (e.g., Harré, 1976, 1988, 2004); however, we differ on some points. Unlike Harré, I explicitly construe analogical modeling as an abductive undertaking because of its strong concern with explanation. I also distinguish between existential abduction and analogical abduction, where the former is used initially to generate hypotheses and theories, and the latter is used to further their development. Harré assigns analogical models a role in the generation of hypotheses as well as in their development, a practice that I acknowledge does occur in science (in fact, I stated in chapter 3 that this happened to an extent in the early formulation of Spearman's theory of general intelligence). Further, I agree with Harré that both critical description and the construction of explanatory theories are major dimensions of science. However, unlike Harré, I rate critical description just as highly as theory construction, as my heavy emphasis on phenomena detection in ATOM attests. Oddly, the recent upsurge of interest in models by philosophers of science gives little recognition to Harré's work.

4. My treatment of the scientific strategy of analogical modeling has an obvious methodological focus. However, the topic of analogical reasoning has attracted a great deal of attention in cognitive psychology in the last thirty years. The chief value of the resulting literature is that it has extended our understanding of analogies from a concern with arguments to inference more broadly. Thus we now have a considerable amount of knowledge about how human agents reason when they employ analogies. A number of psychological theories recommend themselves for consideration, and the best of them importantly extend our epistemological knowledge of the knowing subject. For example, my treatment of the methodology of analogical modeling is broadly consistent with, and would be enriched by, linking it to Holyoak and Thagard's (1995) multiconstraint theory of analogy. Their theory portrays the creative problem-solving process of analogical reasoning as a mapping between a source and a subject via the multiple constraints of similarity, purpose, and structure. Relatedly, for them, the epistemic justification of analogical reasoning takes the form of analogical coherence, which has some similarities with explanatory coherence. This computational theory of analogical reasoning contributes positively to our understanding of the scientist as an analogical thinker.

5 Theory Appraisal: Inference to the Best Explanation

1. Not everyone agrees that the Semmelweis case exemplifies inference to the best explanation. Carl Hempel (1966) took it as an illustration of the hypothetico-deductive method. Others have likened it to Mill's inductive method of difference.

2. The spelling of *Harmany* is deliberate. It is a tribute to Gilbert Harman (1965), who coined the term *inference to the best explanation* and introduced the corresponding idea to the modern philosophical literature.

3. The philosophical literature on theory appraisal sometimes distinguishes between empirical and superempirical criteria. Predictive accuracy is the standard criterion of empirical adequacy, and explanatory power is often mentioned as an example of a superempirical virtue. However, within the theory of explanatory coherence, explanatory breadth is both an empirical and a superempirical criterion; it is simultaneously a measure of empirical adequacy and explanatory power.

4. Although the Bayesian approach seldom figures in the appraisal of psychological theories, it does often form a template for judging the rationality of laypeople in solving hypothesis-testing problems. By contrast, the model for judging the rationality of hypothesis testing undertaken by psychological scientists is provided by the hypothetico-deductive method plus null hypothesis significance testing. This disparity needs a justification.

5. In its strongest form (e.g., Howson & Urbach, 2006), Bayesianism uses probability theory in an attempt to illuminate scientific reasoning generally. In this chapter, I focus on the Bayesian approach as it applies to the appraisal of scientific theories only.

6. Of course, this does not prevent a defender of inference to the best explanation from acknowledging that Bayesianism can be used in contexts such as legal reasoning and medical diagnosis, where the relevant probabilistic information is often available.

6 Conclusion

1. A reasonable requirement of an adequate theory of inquiry is that it can solve the Meno paradox. Happily, the constraint-inclusion view of problems enables us both to formulate and to solve this age-old dilemma—a dilemma that some regard as the foundation problem of inquiry (Nickles, 1981; see also Simon, 1977). This paradox, which is sometimes called the *learning paradox*, questions the very possibility of inquiry. It claims that we cannot inquire either about what we know or about what we don't know. That is, if we know, we have no need to inquire; and if we do not know, we cannot inquire. But inquiry is neither trivial nor impossible: we can solve the Meno paradox by knowing what counts as an acceptable answer, without having an acceptable answer available. The constraint-inclusion view of problems affords just this possibility. This is because our significant research problems will not be fully structured and therefore will not constitute complete descriptions of their solutions. Yet we articulate our problems in terms of their constituent constraints, and these constraints do serve to direct us toward their problem's respective solutions. When we fill out the structure of our problems by progressively including relevant constraints, our problems better point the way to their own solutions. So by solving the Meno paradox, a constraint-inclusion account of problems indicates in a general way

how inquiry is possible. And by deploying this account of problems within ATOM, we are able to say how inquiry can proceed for a broad range and variety of cognitive pursuits. I note here that the Meno paradox has other solutions, some of which use ideas about abductive reasoning.

2. The use of reliability as a mode of justification or validation differs from the normal psychometric practice in which reliability and validity are presented as contrasts. However, the use of consistency tests to validate knowledge claims on reliabilist grounds is widespread in science.

3. Charmaz (2000) has provided an explicitly constructivist depiction of grounded theory that breaks with the "objectivism" of Glaserian grounded theory. From a constructivist perspective, social reality is not revealed so much as socially constructed in the course of inquiry. Further, Rennie (2000) offers a hermeneutic interpretation of grounded theory method that he believes is able to provide an understanding of the meaning of text and reconcile the tensions that exist between realism and relativism in orthodox accounts of the method.

4. Strauss (1987) depicts the discovery of grounded theory, and theory construction in science more generally, as a sequence of induction, deduction, and verification. However, given the pragmatist influence on the origins of grounded theory method, and given that Strauss mentions Peirce's idea of abduction, it is surprising that he does not see its methodological relevance to the generation of grounded theory.

5. Although grounded theory is almost universally regarded as a perspective on qualitative research, it can be applied to quantitative research. Both Glaser and Strauss acknowledged this possibility in their early writings on grounded theory method. A little-recognized fact is that the first piece of grounded theory research, carried out by Glaser (1964) in his examination of the professional careers of organizational scientists, was quantitative in nature.

6. More recently, Romeijn and Williamson (2013) examined the role that interventions can play in resolving the problem of statistical underdetermination in exploratory factor analysis. I agree with the authors that this differs from my focus on abduction and theoretical underdetermination. However, I am skeptical of their claim that intervention data can replace the practice of using theoretical criteria to resolve the problem of theoretical underdetermination.

7. Cameron Ellis and I recently did a content analysis of a representative sample of sixteen current undergraduate research methods textbooks in psychology. The first chapter in these books standardly addresses the topics of scientific method and the nature of science. However, none of them inform the reader about major theories of scientific method, such as those outlined in chapter 1, and the related methodological literature. These texts are the principal source for psychology students' formal learning about scientific method, so it is disturbing that their treatment of scientific method is so poor.

References

Abrantes, P. (1999). Analogical reasoning and modeling in the sciences. *Foundations of Science*, *4*, 237–270.

Achinstein, P. (1968). *Concepts of science: A philosophical analysis*. Baltimore: Johns Hopkins University Press.

Achinstein, P. (1989). Explanation and acceptability. *Behavioral and Brain Sciences*, *12*, 467–468.

Ackerman, R. (1989). The new experimentalism. *British Journal for the Philosophy of Science*, *40*, 185–190.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Armstrong, J. S. (1967). Derivation of theory by means of factor analysis, or Tom Swift and his electric factor analysis machine. *American Statistician*, *21*, 17–21.

Arntzenius, F. (1993). The common cause principle. *Philosophy of Science Association 1992*, *2*, 227–237.

Baird, D. (1992). *Inductive logic: Probability and statistics*. Englewood Cliffs: Prentice Hall.

Bartha, P. F. A. (2010). *By parallel reasoning: The construction and evaluation of analogical arguments*. New York: Oxford University Press.

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies*. Cambridge: Cambridge University Press.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 421–441.

Behrens, J. T., & Yu, C. H. (2003). Exploratory data analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology* (Vol. 2, pp. 33–64). New York: Wiley.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influence on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425.

Bem, D. J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin*, *115*, 4–18.

Blachowicz, J. (2009). How science textbooks treat scientific method: A philosopher's perspective. *British Journal for the Philosophy of Science*, *60*, 303–344.

Black, M. (1962). *Models and metaphors: Studies in language and philosophy*. Ithaca: Cornell University Press.

Bogen, J. (2005). Regularities and causality: Generalizations and causal explanations. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 397–420.

Bogen, J. (2010). Noise in the world. *Philosophy of Science*, *77*, 778–791.

Bogen, J. (2011). "Saving the phenomena" and saving the phenomena. *Synthese*, *182*, 7–22.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, *97*, 303–352.

Bogen, J., & Woodward, J. (1992). Observations, theories, and the evolution of the human spirit. *Philosophy of Science*, *59*, 590–611.

Borleffs, J. C. C., Custers, E. J., van Gijn, J., & ten Cate, O. T. (2003). "Clinical reasoning theatre": A new approach to clinical reasoning education. *Academic Medicine*, *78*, 322–325.

Boyd, R. N. (1984). The current status of scientific realism. In J. Leplin (Ed.), *Scientific realism* (pp. 41–82). Berkeley: University of California Press.

Boyd, R. N. (1989). What realism implies and what it does not. *Dialectica*, *43*, 5–29.

Bridgman, P. (1955). *Reflections of a physicist*. New York: Philosophical Library.

Brown, J. R. (1994). *Smoke and mirrors: How realism maps the world*. London: Routledge.

Brush, S. G. (1989). Prediction and theory evaluation. *Science*, *246*, 1124–1129.

Burian, R. M., & Trout, J. D. (1995). Ontological progress in science. *Canadian Journal of Philosophy*, *25*, 177–202.

Burnston, D., Sheredos, B., & Bechtel, W. (2011). HIT on the psychometric approach. *Psychological Inquiry*, *22*, 108–114.

Butler, G. (1998). Clinical formulation. In P. Salkovskis (Ed.), *Comprehensive clinical psychology* (Vol. 6): *Adults: Clinical formulation and treatment* (pp. 1–24). Oxford: Elsevier.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Campbell, N. R. (1920). *Physics: The elements*. Cambridge: Cambridge University Press.

Capaldi, E. J., & Proctor, R. W. (2008). Are theories to be evaluated in isolation or relative to alternatives? An abductive view. *American Journal of Psychology*, *121*, 617–641.

Carruthers, P. (2002). The roots of scientific reasoning: Infancy, modularity, and the art of tracking. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 73–95). Cambridge: Cambridge University Press.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

Cattell, R. B. (1966). Psychological theory and scientific method. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 1–18). Chicago: Rand McNally.

Cattell, R. B. (1978). *The scientific use of factor analysis in the behavioral and life sciences*. New York: Plenum Press.

Cattell, R. B., & Dickman, K. (1962). A dynamic model of physical influences demonstrating the necessity of oblique simple structure. *Psychological Bulletin*, *59*, 389–400.

Chalmers, A. F. (2013). *What is this thing called science?* (4th ed.). St. Lucia: University of Queensland Press.

Chamberlin, T. C. (1965). The method of multiple working hypotheses [reprint]. *Science*, *148*, 754–759.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. New York: Oxford University Press.

Charmaz, K. (2000). Grounded theory: Constructivist and objectivist methods. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509–535). Thousand Oaks: Sage.

Chatfield, C. (1985). The initial examination of data [with discussion]. *Journal of the Royal Statistical Society, Series A (General)*, *148*, 214–254.

Chomsky, N. (1972). *Language and mind*. New York: Harcourt Brace Jovanovich.

Clark, J. M., & Paivio, A. (1989). Observational and theoretical terms in psychology: A cognitive perspective on scientific language. *American Psychologist*, *44*, 500–512.

Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment*, *12*, 171–188.

Copi, I. M., & Cohen, C. (1990). *Introduction to logic* (8th ed.). New York: Macmillan.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*, 116–127.

Cummins, R. (2000). "How does it work?" versus "What are the laws?" Two conceptions of psychological explanation. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 117–144). Cambridge, MA: MIT Press.

D'Andrade, R. (1986). Three scientific world views and the covering law model. In D. Fiske & R. Shweder (Eds.), *Metatheory in social science: Pluralisms and subjectivities* (pp. 19–41). Chicago: University of Chicago Press.

Darwin, C. (1958). *On the origins of species* (6th ed.). New York: Mentor.

de Groot, A. D. (1969). *Methodology: Foundations of inference and research in the behavioral sciences*. The Hague: Mouton.

Dewey, J. (1910). Science as subject-matter and as method. *Science*, *31*, 121–127.

Dewey, J. (1938). *Logic*. New York: Holt.

Diaconis, P. (1978). Statistical problems in ESP research. *Science*, *201*, 131–136.

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, *108*, 346–369.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*, 274–290.

Duhem, P. (1954). *The aim and structure of physical theory* (2nd ed.). (Weiner, P. P., Trans.). Princeton: Princeton University Press.

Durrant, R., & Haig, B. D. (2001). How to pursue the adaptationist program in psychology. *Philosophical Psychology*, *14*, 357–380.

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*, 193–242.

Eflin, J. T., & Kite, M. E. (1996). Teaching scientific reasoning through attribution. *Teaching of Psychology*, *23*, 87–91.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Ehrenberg, A. S. C., & Bound, J. A. (1993). Predictability and prediction [with discussion]. *Journal of the Royal Statistical Society, Series A (General)*, *156*, 167–206.

Einstein, A. (1949). Remarks concerning the essays brought together in this cooperative volume. In P. Schilpp (Ed.), *Albert Einstein: Philosopher-scientist* (pp. 665–688). La Salle: Open Court.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Erwin, E. (1992). Current philosophical issues in the scientific evaluation of behavior therapy theory and outcome. *Behavior Therapy*, *23*, 151–171.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299.

Feigl, H. (1956). Some major issues and developments in the philosophy of science of logical empiricism. In H. Feigl & M. Scriven (Eds.), *Minnesota studies* *in the philosophy of science* (Vol. 1): *The foundations of science and the concepts of psychology and psychoanalysis* (pp. 3–37). Minneapolis: University of Minnesota Press.

Feyerabend, P. (1975). *Against method*. London: Verso.

Feynman, R. (1998). *The meaning of it all*. Harmondsworth: Penguin.

Fidell, L. S., & Tabachnick, B. G. (2003). Preparatory data analysis. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology* (Vol. 2, pp. 115–121). New York: Wiley.

Flynn, J. R. (2009). *What is intelligence? Beyond the Flynn effect* (2nd ed.). Cambridge: Cambridge University Press.

Franklin, A. (1990). *Experiment, right or wrong*. Cambridge: Cambridge University Press.

Franklin, A. (1997). Calibration. *Perspectives on Science*, *5*, 31–80.

Funder, D. C. (2009). Naive and obvious questions. *Perspectives on Psychological Science*, *4*, 340–344.

Gage, N. L. (1996). Confronting counsels of despair for the behavioral sciences. *Educational Researcher*, *25*, 5–15, 22.

Galanter, C. A., & Patel, V. L. (2005). Medical decision-making: A selective review for child psychiatrists and psychologists. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *46*, 675–689.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, *1*, 67–89.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychological Review*, *63*, 149–159.

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, *26*, 309–320.

Gergen, K. J. (1994). *Toward transformation in social knowledge* (2nd ed.). Thousand Oaks: Sage.

Giere, R. N. (1983). Testing theoretical hypotheses. In J. Earman (Ed.), *Minnesota studies in the philosophy of science* (Vol. 10): *Testing scientific theories* (pp. 269–298). Minneapolis: University of Minnesota Press.

Giere, R. N. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.

Giere, R. N. (1999). Using models to represent reality. In L. Magnani, N. J. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 41–57). New York: Kluwer Academic/Plenum Publishers.

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 391–408). Thousand Oaks: Sage.

Glaser, B. G. (1964). *Organizational scientists: Their professional careers*. Indianapolis: Bobbs-Merrill.

Glaser, B. G. (1978). *Theoretical sensitivity*. Mill Valley: Sociology Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Glymour, B. (2000). Data and phenomena: A distinction reconsidered. *Erkenntnis*, *52*, 29–37.

Glymour, C. (1980). *Theory and evidence*. Princeton: Princeton University Press.

Glymour, C. (1992). Invasion of the mind snatchers. In R. N. Giere (Ed.), *Minnesota studies in the philosophy of science* (Vol. 15): *Cognitive models of science* (pp. 465–474). Minneapolis: University of Minnesota Press.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Doubleday Anchor.

Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard University Press.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale: Lawrence Erlbaum.

Gower, B. (1997). *Scientific method: An historical and philosophical introduction*. London: Routledge.

Greenwood, J. D. (1992). Realism, empiricism, and social constructionism. *Theory and Psychology*, *2*, 131–151.

Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.

Hacking, I. (1991). Speculation, calculation, and the creation of phenomena. In G. Munevar (Ed.), *Beyond reason: Essays on the philosophy of Paul K. Feyerabend* (pp. 131–157). Dordrecht: Kluwer.

Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.

Haig, B. D. (1987). Scientific problems and the conduct of research. *Educational Philosophy and Theory*, *19*, 22–32.

Haig, B. D. (1996). Grounded theory as scientific method. In A. Neiman (Ed.), *Philosophy of education 1995: Current issues* (pp. 281–290). Urbana: University of Illinois Press.

Haig, B. D. (2003). What is a spurious correlation? *Understanding Statistics*, *2*, 125–132.

Haig, B. D. (2005a). An abductive theory of scientific method. *Psychological Methods*, *10*, 371–388.

Haig, B. D. (2005b). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, *40*, 303–329.

Haig, B. D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *American Journal of Psychology*, *122*, 219–234.

Haig, B. D. (2013). Detecting psychological phenomena: Taking bottom-up research seriously. *American Journal of Psychology*, *126*, 135–153.

Haig, B. D., & Borsboom, D. (2012). Truth, science, and psychology. *Theory and Psychology*, *22*, 272–289.

Hammersley, M. (2011). *Methodology: Who needs it?* London: Sage.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review*, *74*, 88–95.

Harré, R. (1970). *The principles of scientific thinking*. Chicago: University of Chicago Press.

Harré, R. (1976). The constructive role of models. In L. Collins (Ed.), *The use of models in the social sciences* (pp. 16–43). London: Tavistock.

Harré, R. (1979). *Social being*. Oxford: Blackwell.

Harré, R. (1988). Where models and analogies really count. *International Studies in the Philosophy of Science*, *2*, 119–133.

Harré, R. (2004). *Models: Gateway to the unknown*. Amsterdam: Elsevier.

Harré, R., & Secord, P. F. (1972). *The explanation of social behavior*. Oxford: Blackwell.

Hartley, R. E. (1954). Two kinds of factor analysis. *Psychometrika*, *19*, 195–203.

Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist*, *42*, 443–455.

Hempel, C. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.

Hempel, C. (1966). *Philosophy of natural science*. Englewood Cliffs: Prentice Hall.

Herbert, H. (1971). Darwin, Malthus, and selection. *Journal of the History of Biology*, *4*, 209–217.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Hesse, M. (1966). *Models and analogies in science*. Notre Dame: University of Notre Dame Press.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Hooker, C. A. (1987). *A realistic theory of science*. New York: State University of New York Press.

Howson, C., & Urbach, P. (2006). *Scientific reasoning: The Bayesian approach* (3rd ed.). La Salle: Open Court.

Hubbard, R. (2004). Alphabet soup: Blurring the distinctions between p's and α's in psychological research. *Theory and Psychology*, *14*, 295–327.

Hubbard, R., & Lindsay, R. M. (2013). From significant difference to significant sameness: Proposing a paradigm shift in business research. *Journal of Business Research*, *66*, 1377–1388.

Hyman, R. (1985). The Ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology*, *49*, 3–49.

Jaccard, J. (2013). Theory construction, model building, and model selection. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (pp. 82–104). New York: Oxford University Press.

Janssen, M. (2002). COI stories: Explanation and evidence in the history of science. *Perspectives on Science*, *10*, 457–522.

Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). New York: Routledge.

Jones, M. R. (2005). Idealization and abstraction: A framework. In M. R. Jones & N. Cartwright (Eds.), *Idealization XII: Correcting the model: Idealization and abstraction in the sciences* (pp. 173–217). Amsterdam: Rodopi.

Jones, S. R. G. (1992). Was there a Hawthorne effect? *American Journal of Sociology*, *98*, 451–468.

Josephson, J. R., & Josephson, S. G. (1994). *Abductive inference: Computation, philosophy, technology*. New York: Cambridge University Press.

Kaiser, M. (1991). From rocks to graphs: The shaping of phenomena. *Synthese*, *89*, 111–133.

Kaiser, M. (1996). Empirical versus theoretical progress in science. In R. S. Cohen, R. Hilpen, & Q. Renzong (Eds.), *Realism and anti-realism in the philosophy of science* (pp. 171–193). Dordrecht: Kluwer.

Kaplan, A. (1964). *The conduct of inquiry: Methodology for the behavioral sciences*. Scranton: Chandler.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks: Sage.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research: Educational and psychological inquiry* (4th ed.). Fort Worth: Harcourt College Publishers.

Kievit, R. A., Romeijn, J. W., Waldorp, L. J., Wicherts, J. M., Scholte, H. S., & Borsboom, D. (2011). Modeling mind and matter: Reductionism and psychological measurement in cognitive neuroscience. *Psychological Inquiry*, *22*, 139–157.

Kim, J. O., & Mueller, C. W. (1978). *Introduction to factor analysis*. Beverly Hills: Sage.

Kincaid, H. (2000). Global arguments and local realism about the social sciences. *Philosophy of Science* [supplement], *67*, 667–678.

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. New York: Oxford University Press.

Kline, R. B. (2013). *Beyond significance testing: Reforming data analysis methods in behavioral research* (2nd ed.). Washington, DC: American Psychological Association.

Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change*. Chicago: University of Chicago Press.

Kuhn, T. S. (2012). *The structure of scientific revolutions* (4th ed.). Chicago: University of Chicago Press.

Kukla, A. (2001). *Methods of theoretical psychology*. Cambridge, MA: MIT Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge: Cambridge University Press.

Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Beverly Hills: Sage.

Laudan, L. (1977). *Progress and its problems: Towards a theory of scientific growth*. Berkeley: University of California Press.

Laudan, L. (1981). *Science and hypothesis: Historical essays on scientific methodology*. Dordrecht: Reidel.

Laudan, L. (1996). *Beyond positivism and relativism: Theory, method, and evidence*. Boulder: Westview Press.

Lee, M. D., & Wagenmakers, E. J. (2005). Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological Review*, *112*, 662–668.

Levins, R. (1968). *Evolution in changing environments*. Princeton: Princeton University Press.

Lindsay, R. M., & Ehrenberg, A. S. C. (1993). The design of replicated studies. *American Statistician*, *47*, 217–228.

Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). London: Routledge.

Lycan, W. G. (1988). *Judgement and justification*. Cambridge: Cambridge University Press.

Lykken, D. T. (1971). Multiple factor analysis and personality research. *Journal of Experimental Research in Personality*, *5*, 161–170.

Lykken, D. T. (1991). What's wrong with psychology anyway? In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1): *Matters of public interest* (pp. 3–39). Minneapolis: University of Minnesota Press.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research*, *38*, 113–139.

Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*, 1–25.

Magnani, L. (2001). *Abduction, reason, and science: Processes of discovery and explanation*. New York: Kluwer/Plenum.

Mäki, U. (2005). Reglobalizing realism by going local, or (how) should our formulations of scientific realism be informed about the sciences? *Erkenntnis*, *63*, 231–251.

Mäki, U. (2011). Models and the locus of their truth. *Synthese*, *180*, 47–63.

Manicas, P. T., & Secord, P. F. (1983). Implications for psychology of the new philosophy of science. *American Psychologist*, *38*, 399–413.

Markus, K. A., Hawes, S. W., & Thasites, R. J. (2008). Abductive inferences to psychological variables: Steiger's question and best explanations of psychopathy. *Journal of Clinical Psychology*, *64*, 1069–1088.

Marquardt, D. W. (1987). The importance of statisticians. *Journal of the American Statistical Association*, *82*, 1–7.

McAllister, J. W. (1997). Phenomena and patterns in data sets. *Erkenntnis*, *47*, 217–228.

McArdle, J. J. (1996). Current directions in structural factor analysis. *Current Directions in Psychological Science*, *5*, 11–18.

McDonald, R. P., & Mulaik, S. A. (1979). Determinacy of common factors: A nontechnical review. *Psychological Bulletin*, *86*, 297–306.

McGrew, T. (2003). Confirmation, heuristics, and explanatory reasoning. *British Journal for the Philosophy of Science*, *54*, 553–567.

McGuire, W. J. (1997). Creative hypothesis generating in psychology: Some useful heuristics. *Annual Review of Psychology*, *48*, 1–30.

McMullin, E. (1983). Values in science. In P. D. Asquith & T. Nickles (Eds.), *Proceedings of the 1982 biennial meeting of the Philosophy of Science Association* (Vol. 2, pp. 3–28). East Lansing: Philosophy of Science Association.

McMullin, E. (1984). On hypothetical approaches. *Isis*, *75*, 372–377.

McMullin, E. (1995). Underdetermination. *Journal of Medicine and Philosophy*, *20*, 233–252.

Meehl, P. E. (1990). Why summaries of research on theories are often uninterpretable. *Psychological Reports* [supplement], *66*, 195–244.

Meehl, P. E. (1993). Philosophy of science: Help or hindrance? *Psychological Reports*, *72*, 707–733.

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin*, *125*, 387–391.

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators: Perspectives on natural and social science* (pp. 10–37). Cambridge: Cambridge University Press.

Mulaik, S. A. (1985). Exploratory statistics and empiricism. *Philosophy of Science*, *52*, 410–430.

Mulaik, S. A. (1987). A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research*, *22*, 267–305.

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton: Chapman & Hall/CRC.

Mulaik, S. A., & McDonald, R. P. (1978). The effect of additional variables on factor indeterminacy in models with a single common factor. *Psychometrika*, *43*, 177–192.

Mulaik, S. A., & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling*, *7*, 36–73.

Muller, K. E., Otto, D. A., & Benignus, V. A. (1983). Design and analysis issues and strategies in psychophysiological research. *Psychophysiology*, *20*, 2012–2018.

Mumford, S. (1998). *Dispositions*. Oxford: Oxford University Press.

Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. New York: Harcourt, Brace, and World.

Nickles, T. (1981). What is a problem that we might solve it? *Synthese*, *47*, 85–118.

Nickles, T. (1987a). Methodology, heuristics, and rationality. In J. C. Pitt & M. Pera (Eds.), *Rational changes in science* (pp. 103–132). Dordrecht: Reidel.

Nickles, T. (1987b). 'Twixt method and madness. In N. J. Nersessian (Ed.), *The process of science* (pp. 41–67). Dordrecht: Martinus Nijhoff.

Nickles, T. (1988). Questioning and problems in philosophy of science: Problem-solving versus directly truth-seeking epistemologies. In M. Meyer (Ed.), *Questions and questioning* (pp. 43–67). Berlin: de Gruyter.

Nickles, T. (1997). Methods of discovery. *Biology and Philosophy*, *12*, 127–140.

Nickles, T. (2009). The strange story of scientific method. In J. Meheus & T. Nickles (Eds.), *Models of discovery and creativity* (pp. 167–207). Dordrecht: Springer.

Nola, R., & Sankey, H. (2007). *Theories of scientific method*. Montreal: McGill-Queens University Press.

Norris, S. P., Phillips, L. M., & Korpan, C. A. (2003). University students' interpretation of media reports of science and its relationship to background knowledge, interest, and reading difficulty. *Public Understanding of Science*, *12*, 123–145.

Norton, J. D. (2003). A material theory of induction. *Philosophy of Science*, *70*, 647–670.

O'Donohue, W., & Buchanan, J. A. (2001). The weaknesses of strong inference. *Behavior and Philosophy*, *29*, 1–20.

Orlitzky, M. (2012). How can significance tests be deinstitutionalized? *Organizational Research Methods*, *15*, 199–228.

Patel, V. L., Arocha, J. F., & Zhang, J. (2005). Thinking and reasoning in medicine. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 727–750). New York: Cambridge University Press.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufman.

Peirce, C. S. (1931–1958). *The collected papers of Charles Sanders Peirce* (Vols. 1–8). C. Hartshorne & P. Weiss (Eds., Vols. 1–6), & A. W. Burks (Ed., Vols. 7–8). Cambridge, MA: Harvard University Press.

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Oxford University Press.

Popper, K. R. (1983). *Realism and the aim of science*. Totowa: Rowman & Littlefield.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*, 13–43.

Proctor, R. W., & Capaldi, E. J. (2001a). Empirical evaluation and justification of methodologies in psychological science. *Psychological Bulletin*, *127*, 759–772.

Proctor, R. W., & Capaldi, E. J. (2001b). Improving the science education of psychology students: Better teaching of methodology. *Teaching of Psychology*, *28*, 173–181.

Proctor, R. W., & Capaldi, E. J. (2006). *Why science matters: Understanding the methods of psychological research*. London: Blackwell.

Psillos, S. (1999). *Scientific realism: How science tracks the truth*. London: Routledge.

Psillos, S. (2002). Simply the best: A case for abduction. In A. C. Kakas & F. Sadri (Eds.), *Computational logic* (pp. 605–625). Berlin: Springer.

Psillos, S. (2004). Inference to the best explanation and Bayesianism. In F. Stadler (Ed.), *Induction and deduction in the sciences* (pp. 83–91). Dordrecht: Kluwer.

Quine, W. V. O. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.

Quine, W. V. O., & Ullian, J. S. (1978). *The web of belief* (2nd ed.). New York: Random House.

Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Chicago Press.

Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.

Reitman, W. (1964). Heuristic decision procedures, open constraints, and the nature of ill-defined problems. In M. W. Shelley II & G. L. Bryan (Eds.), *Human judgments and optimality* (pp. 282–314). New York: Wiley.

Rennie, D. L. (2000). Grounded theory methodology as methodological hermeneutics: Reconciling realism and relativism. *Theory and Psychology*, *10*, 481–502.

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, *65*, 1–12.

Rodgers, J. L., & Rowe, D. C. (2002). Theory development should begin (but not end) with good empirical fits: A comment on Roberts and Pashler (2000). *Psychological Review*, *109*, 599–604.

Romeijn, J. W. (2008). The all-too-flexible abductive method: ATOM's normative status. *Journal of Clinical Psychology*, *64*, 1023–1036.

Romeijn, J. W., & Williamson, J. (2013). Intervention, underdetermination, and theory generation. (Unpublished manuscript.)

Rorer, L. G. (1991). Some myths of science in psychology. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1): *Matters of public interest* (pp. 61–87). Minneapolis: University of Minnesota Press.

Rosenkrantz, R. D. (1977). *Inference, method, and decision: Toward a Bayesian philosophy of science*. Dordrecht: Reidel.

Ross, S. D. (1981). *Learning and discovery*. New York: Gordon & Breach.

Rozeboom, W. W. (1961). Ontological induction and the logical typology of scientific variables. *Philosophy of Science*, *28*, 337–377.

Rozeboom, W. W. (1972). Scientific inference: The myth and the reality. In R. S. Brown & D. J. Brenner (Eds.), *Science, psychology, and communication: Essays honoring William Stephenson* (pp. 95–118). New York: Teachers College Press.

Rozeboom, W. W. (1973). Dispositions revisited. *Philosophy of Science*, *40*, 59–74.

Rozeboom, W. W. (1984). Dispositions do explain: Picking up the pieces after Hurricane Walter. *Annals of Theoretical Psychology*, *1*, 205–223.

Rozeboom, W. W. (1991a). Hyball: A method for subspace-constrained factor rotation. *Multivariate Behavioral Research*, *26*, 163–177.

Rozeboom, W. W. (1991b). Theory and practice of analytic hyperplane optimization. *Multivariate Behavioral Research*, *26*, 79–97.

Rozeboom, W. W. (1997). Good science is abductive, not hypothetico-deductive. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 335–391). Hillsdale: Erlbaum.

Rozin, P. (2009). What kind of empirical research should we publish, fund, and reward? *Perspectives on Psychological Science*, *4*, 435–439.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Samelson, F. (1980). J. B. Watson's little Albert, Cyril Burt's twins, and the need for a critical science. *American Psychologist*, *35*, 619–625.

Sankey, H. (2008). *Scientific realism and the rationality of science*. Aldershot: Ashgate.

Schatzman, L., & Strauss, A. L. (1973). *Field research: Strategies for a natural sociology*. Englewood Cliffs: Prentice Hall.

Schindler, S. (2007). Rehabilitating theory: Refusal of the "bottom-up" construction of scientific phenomena. *Studies in History and Philosophy of Science*, *38*, 160–184.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, *47*, 1173–1181.

Schmidt, F. L. (1993). Meta-analysis and cumulative knowledge. *Contemporary Psychology*, *38*, 1163–1165.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine*, *65*, 611–621.

Searle, J. R. (1995). *The social construction of reality*. New York: Free Press.

Sechrest, L., McKnight, P., & McKnight, K. (1996). Calibration of measures for psychotherapy outcome studies. *American Psychologist*, *51*, 1065–1071.

Semmelweis, I. (1983). *The etiology, concept, and prophylaxis of childbed fever*. K. C. Carter (Trans. and Ed.). Madison: University of Wisconsin Press.

Shapiro, L. A. (1994). Behavior, ISO functionalism, and psychology. *Studies in History and Philosophy of Science, Part A*, *25*, 191–209.

Sherman, R. A., & Funder, D. C. (2009). Evaluating correlations in studies of personality and behavior: Beyond the number of significant findings to be expected by chance. *Journal of Research in Personality*, *43*, 1053–1063.

Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Simon, H. A. (1977). *Models of discovery and other topics in the methods of science*. Dordrecht: Reidel.

Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press.

Skinner, B. F. (1956). A case history in scientific method. *American Psychologist*, *11*, 221–233.

Skinner, B. F. (1984). Methods and theories in the experimental analysis of behavior. *Behavioral and Brain Sciences*, *7*, 511–546.

Slife, B. D., & Williams, R. N. (1997). Toward a theoretical psychology: Should a subdiscipline be formally recognized? *American Psychologist*, *52*, 117–129.

Sober, E. (1982). Dispositions and subjunctive conditionals, or dormative virtues are no laughing matter. *Philosophical Review*, *91*, 591–596.

Sober, E. (1988). The principle of the common cause. In J. H. Fetzer (Ed.), *Probability and causality* (pp. 211–229). Dordrecht: Reidel.

Sohn, D. (1996). Meta-analysis and science. *Theory and Psychology*, *6*, 229–246.

Sokal, R. R., Rohlf, F. J., & Zang, E. (1980). Reification in factor analysis: A plasmode based on human physiology-of-exercise variables. *Multivariate Behavioral Research*, *2*, 181–202.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge, MA: MIT Press.

Sprenger, J. (2011). Hypothetico-deductive confirmation. *Philosophy Compass*, *6*, 497–508.

Stam, H. J. (2006). On the uses of theory. *General Psychologist*, *41*, 30–32.

Stephenson, W. W. (1961). Scientific creed—1961. *Psychological Record*, *11*, 1–25.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.

Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks: Sage.

Suppes, P. (1962). Models of data. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology, and philosophy of science: Proceedings of the 1960 International Congress* (pp. 252–261). Stanford: Stanford University Press.

Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn effect in reverse. *Personality and Individual Differences*, *39*, 837–843.

Teller, P. (2010). "Saving the phenomena" today. *Philosophy of Science*, *77*, 815–826.

Thagard, P. (1978). The best explanation: Criteria for theory choice. *Journal of Philosophy*, *75*, 76–92.

Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*, 435–467.

Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Thagard, P. (2007). Coherence, truth, and the development of scientific knowledge. *Philosophy of Science*, *74*, 28–47.

Thagard, P., & Litt, A. (2008). Models of scientific explanation. In R. Sun (Ed.), *The Cambridge handbook of computational cognitive modeling* (pp. 549–564). Cambridge: Cambridge University Press.

Theunissen, B. (2012). Darwin and his pigeons: The analogy between artificial and natural selection revisited. *Journal of the History of Biology*, *45*, 179–212.

Thompson, B. (1994). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. *Journal of Personality*, *62*, 157–176.

Thurstone, L. L. (1947). *Multiple-factor analysis* (2nd ed.). Chicago: University of Chicago Press.

Trout, J. D. (1998). *Measuring the intentional world: Realism, naturalism, and quantitative methods in the behavioral sciences*. New York: Oxford University Press.

Tucker, W. H. (1994). Fact and fiction in the discovery of Sir Cyril Burt's flaws. *Journal of the History of the Behavioral Sciences*, *30*, 335–347.

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, *24*, 83–91.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.

Tukey, J. W. (1980). We need both exploratory and confirmatory. *American Statistician*, *34*, 23–25.

Tuomela, R. (Ed.). (1978). *Dispositions*. Dordrecht: Reidel.

Uchino, B. N., Thoman, D., & Byerly, S. (2010). Inference patterns in theoretical social psychology: Looking back as we move forward. *Social and Personality Psychology Compass*, *4*, 417–427.

van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford: Oxford University Press.

Vertue, F. M., & Haig, B. D. (2008). An abductive perspective on clinical reasoning and case formulation. *Journal of Clinical Psychology*, *64*, 1046–1068.

Wachtel, P. L. (1980). Investigation and its discontents: Some constraints on progress in psychological research. *American Psychologist*, *35*, 399–408.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology*, *100*, 426–434.

Walters, J. M., & Gardner, H. (1986). The theory of multiple intelligences: Some issues and answers. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 163–182). Cambridge: Cambridge University Press.

Ward, T., Vertue, F., & Haig, B. D. (1999). Abductive method and clinical assessment in practice. *Behaviour Change*, *16*, 64–73.

Wartofsky, M. (1979). *Models: Representation and the scientific understanding*. Dordrecht: Reidel.

Weber, B. H. (2007). Fact, phenomenon, and theory in the Darwinian research tradition. *Biological Theory*, *2*, 168–178.

Weinert, F. (1992). Vicissitudes of laboratory life. *British Journal for the Philosophy of Science*, *43*, 423–429.

Weisberg, J. (2009). Locating IBE in the Bayesian framework. *Synthese*, *167*, 125–143.

Weisberg, M. (2007). Who is a modeler? *British Journal for the Philosophy of Science*, *58*, 207–233.

Whitt, L. A. (1992). Indices of theory promise. *Philosophy of Science*, *59*, 612–634.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604.

Wimsatt, W. C. (1981). Robustness, reliability, and overdetermination. In M. Brewer & B. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 124–163). San Francisco: Jossey-Bass.

Wimsatt, W. C. (1986). Heuristics and the study of human behavior. In D. W. Fiske & R. A. Shweder (Eds.), *Metatheory in social science* (pp. 293–314). Chicago: University of Chicago Press.

Wimsatt, W. C. (1994). The ontology of complex systems: Levels of organization, perspectives, and causal thickets. *Canadian Journal of Philosophy* [supplement], *20*, 207–274.

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University Press.

Woodward, J. (1989). Data and phenomena. *Synthese*, *79*, 393–472.

Woodward, J. (2000). Data, phenomena, and reliability. *Philosophy of Science* [supplement], *67*, 163–179.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Woodward, J. (2010). Data, phenomena, signal, and noise. *Philosophy of Science*, *77*, 792–803.

Woodward, J. (2011). Data and phenomena: A restatement and a defense. *Synthese*, *182*, 165–179.

Yu, C. H. (2006). *Philosophical foundations of quantitative research methodology*. Lanham: University Press of America.

Yu, C. H. (2008). Resampling: A conceptual and procedural introduction. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 283–298). Los Angeles: Sage.

Index