Theory Change in Science

Strategies from Mendelian Genetics

LINDLEY DARDEN

Theory Change in Science

Monographs on the History and Philosophy of Biology

RICHARD BURIAN, RICHARD BURKHARDT, JR., RICHARD LEWONTIN, JOHN MAYNARD SMITH

EDITORS

The Cuvier-Geoffrey Debate: French Biology in the Decades Before Darwin
TOBY A. APPEL

Controlling Life: Jacques Loeb and the Engineering Ideal in Biology
PHILIP J. PAULY

Beyond the Gene: Cytoplasmic Inheritance and the Struggle for Authority in Genetics
JAN SAPP

The Heritage of Experimental Embryology: Hans Spemann and the Organizer
VIKTOR HAMBURGER

The Evolutionary Dynamics of Complex Systems: A Study in Biosocial Complexity
C. DYKE

The Wellborn Science: Eugenics in Germany, France, Brazil, and Russia
Edited by MARK B. ADAMS

Darwin without Malthus: The Struggle for Existence in Russian Evolutionary Thought
DANIEL P. TODES

Theory Change in Science: Strategies from Mendelian Genetics
LINDLEY DARDEN

Theory Change in Science

Strategies from Mendelian Genetics

LINDLEY DARDEN

University of Maryland College Park

New York Oxford

OXFORD UNIVERSITY PRESS

1991

Oxford University Press

Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Petaling Jaya Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland

and associated companies in
Berlin Ibadan

Published by Oxford University Press, Inc.,
200 Madison Avenue, New York, New York 10016

Library of Congress Cataloging-in-Publication Data
Darden, Lindley.
Theory change in science : strategies from Mendelian genetics /
Lindley Darden.
p. cm. — (Monographs on the history and philosophy of biology)
Includes bibliographical references (p.) and index.
ISBN 0-19-506797-5 (cloth)
1. Science—Methodology. 2. Science—Methodology—Case studies.
3. Human genetics—Case studies. 4. Science—Philosophy.
5. Science—Philosophy—Case studies.
I. Title. II. Series.
Q175.D268 1991 502.8—dc20 90-22007

2 4 6 8 9 7 5 3 1

Printed in the United States of America
on acid-free paper

Preface

Mississippians, even former Mississippians who have come north toward home, write books. I was born in the same town as William Faulkner (New Albany) and I know the fun of trying to find the real life models for characters in a local's novel. Unfortunately, this book offers no such fun. It is a book about ideas, not personalities. Playing with ideas is fun too, but fun of a different sort. The ''characters'' in this book aren't the southern ladies in the garden club, but genetic characteristics and laws of their inheritance. Before plunging into the academic analysis, I'm going to be a bit self-indulgent and discuss characters in my life and autobiographical influences. Like many an author of a first book, I feel obliged to thank many people who have aided and influenced me over the years. Readers anxious to get on with the topic may prefer to skip to the beginning of the intellectual discussion in Chapter 1.

The book is dedicated to the memory of my father, Leslie Darden. He was a gracious southern gentleman and a lover of learning (as well as a ''country lawyer,'' his own somewhat tongue-in-cheek appellation, issued from the podium of the state bar association). He awakened me to the excitement that comes from learning new ideas and provided a well-stocked library to explore. Moss Hill, the thirteen acres on which I grew up, provided the environment in which I came to appreciate the natural world. I followed my grandmother, ''Mama'' Nabors, around in her garden, and watered the grass with my mother, Inez Darden, and tasted the Lindley grapes grown by ''Miss Josie,'' my other grandmother, who lived cross-town. The grapes, she claimed, were developed by some distant relative in England, who probably worked after our Lindley ancestors settled in Pennsylvania in the 1600s. My strong interests in biology are due to that early environment. Like all precocious but insecure students, the encouragement that I received from various teachers was important. One of my high school teachers, Hilda Hill, still serves as my role model for a woman in front of a class giving a lecture, because I had very few women professors in college or graduate school. Her poetic admonition still echoes: ''you are a part of all that you have met.''

My interest in philosophy of science developed in Larry Lacy's excellent course at Rhodes College (then named Southwestern at Memphis). We read Hanson's (1958) *Patterns of Discovery*. Hanson raised fascinating questions about reasoning in scientific discovery, but so obviously didn't solve them satisfactorily that I've been worrying about them ever since. The discovery of new ideas is exciting, yet methods for discovery have been too little explored. My graduate training in philosophy, biology, and conceptual foundations of science at the University of Chicago set me on my present course in the history and philosophy of science. Academic life in a congenial history and philosophy of science program at the University of Maryland College Park ended my search for a place where I, as a woman with scholarly interests, could feel at home.

At the University of Chicago, my dissertion analyzed the case study of the beginnings of genetics in the nineteenth century (and this book takes up where that case left off). Lasting influences were left by my teachers: Kenneth Schaffner, Stephen Toulmin, David Hull, Bill Wimsatt, the late Arnold Ravin, and especially Dudley Shapere. I'm sure that their work influences this book more than the actual citations indicate.

I've been working on Mendelian genetics as a case study for many years and have written a number of articles using it to address philosophical issues other than the ones taken up in this book. Earlier work on the case was supported by the History and Philosophy of Science Program of the National Science Foundation. This book received support from the American Council of Learned Societies, as well as the Graduate School and the College of Arts and Humanities at the University of Maryland College Park. I am grateful for that support. I have learned much from the historians who have written so cogently about genetics, especially Bob Olby, Onno Meijer, Garland Allen, E.A. Carlson, and Scott Gilbert. And, I am grateful to librarians and archivists for their help at the American Philosophical Society, the University of California at Berkeley, the California Institute of Technology, the Marine Biological Laboratory at Woods Hole, and the interlibrary loan office at McKeldin Library at the University of Maryland College Park.

In 1978 I attended a conference that Ken Schaffner organized on "Logic of Discovery and Diagnosis in Medicine." Hearing Bruce Buchanan's paper "Steps toward Mechanizing Discovery" (published in 1985) was another intellectual turning point for me. I realized that the field of artificial intelligence (AI) was addressing some of the same questions about scientific knowledge and reasoning as I was, from my more historical perspective. I am grateful to Bruce, his colleagues, and students at the Heuristic Programming Project at Stanford University (now the Knowledge Systems Laboratory) for their cordial welcome during my sabbatical visit in 1980. Following developments in AI and computational philosophy of science (as Paul Thagard has dubbed it) has been exciting. Many of the ideas in this book have been influenced by those developments. I am grateful to Roy Rada for his collaboration in constructing an AI LISP program to simulate the discovery of the chromosome theory of heredity. That work convinced me how difficult it is, with current AI techniques, to simulate scientific reasoning. Also, I have enjoyed learning about the work on abduction and diagnostic reasoning by B. Chandrasekaran, John Josephson, Dale Moberg, and their colleagues at Ohio State University. I look forward to further collaboration with them.

I'm sure that a number of my positions on various issues have influenced my analysis. It may be fruitful to mention some of my "biases." Just as that term is used in machine learning, my usage is not pejorative. Biases provide needed focus and direction. One can't do everything at once. Choices must be made, directions taken, within the space of possibilities. My basic interests are metaphysical and epistemological, but I am modest about what evidence I have for drawing conclusions in those grand philosophical subjects. I believe that science has been our most successful knowledge-gaining enterprise. An understanding of the nature of scientific knowledge and reasoning is an important approach to answering basic questions about the nature of reality and the nature of methods for knowing it. As a historian, I am an "internalist," concerned with ideas, not an "externalist," concerned with social and institutional aspects of science. If science doesn't produce reliable (though not infallible) knowledge, but is only a social construct that serves various human interests, then I am not interested in it. I might as well become a gardener, instead of a philosopher of science. If I take gardening more seriously, however, then I'll want to do a soil test to determine the scientific analysis of my soil. If I start breeding new species of grapes, then I will be using Mendelian principles and believing that they give me reliable knowledge. So . . . I simply can't escape being a realist about scientific knowledge.

In this book, I will not argue for internal as opposed to external approaches to science, nor will I argue for a realist as opposed to an antirealist interpretation. One book cannot do everything; others who see the need to debate those points will no doubt continue to do so. My task is to find strategies for theory change. I want to understand how to develop new scientific ideas and theories. I want to find the methods that scientists already seem to be using (although they usually are not explicit about their methods). And I want to suggest additional methods, such as more thorough and systematic methods that scientists can use, or methods that a computer can use more readily than humans. Abstract philosophical arguments for the justification of these strategies as methods for producing *true* scientific knowledge about the *real* world are beyond the scope of my discussion here.

I have discussed the ideas in this book with many people over the years. I have thoroughly enjoyed our lively intellectual discussions about these topics. In addition, a number of people read earlier drafts and provided me with insightful comments. I fear that I will forget to thank some of those whose ideas have left a mark here. At the risk of that, however, let me mention a few: Bill Bechtel, Myles Brand, John Clement, Stiv Fleishman, Jonathan Harwood, Jon Hodge, John Josephson, Josh Lederberg, Jerry Levinson, Jane Maienschein, Ed Manukian, Dale Moberg, Tom Nickles, Bob Olby, Phil Pauly, Moreland Perkins, Roy Rada, Bob Richardson, Pat Ross, Ken Schaffner, Dudley Shapere, Frederick Suppe, Paul Thagard, and Marga Vicedo. The members of our local Study Group in History and Philosophy of Biology have been very helpful, reading every chapter, sending me back to the drawing board at times, being encouraging at others. Perhaps to my peril, I have not always taken their advice. My heartfelt thanks to Pamela Henson, Joel Hagen, Joe A. Cain, and Bob Witkin. Thanks also to the students in my seminar on strategies for theory change for their comments on the manuscript, especially Fred Hickok, Ana Simoes, and Jim Antonisse.

I am grateful to Dick Burian for his extensive and knowledgeable editorial comments and for his support of my project with Oxford University Press. William Curtis of Oxford University Press was encouraging over the years; he expressed an early interest in the project and gently nudged me with queries about its progress when we met at conferences. He provided support during contract negotiations and manuscript preparation. My thanks to him. My thanks also to Stiv Fleishman, Pat Ross, and Liz Fannin for their help with the figures, bibliography, and manuscript preparation. Finally, various friends and relatives have been supportive during the years that I have been glued to my computer screen and tolerant of my plea that I would "do X after the book is finished": Shearer Rumsey; Natalie Schmitt; Virginia and George Moryadas; Tom, Betty, David, and Morgan Darden; and various folk dance friends at Buffalo Gap Camp.

I hope the reader will enjoy a quiet discussion with me while reading the book as much as I have enjoyed discussing ideas in it with others.

College Park, MD
July 1990

L.D.

Contents

This page intentionally left blank

Theory Change in Science

This page intentionally left blank

CHAPTER 1

Introduction

New ideas are exciting. Yet how they develop is mysterious. This book attempts to remove some of the mystery from the development of new scientific ideas.

Consider: If new ideas arise from old ones, then their origin isn't so puzzling. If new theories don't arise all at once, but instead develop piecemeal in incremental stages, then scientific theory formation is more comprehensible than it would appear at first glance. If theories often arise with vague new ideas, with implicit components and unclear implications, then the process of theory construction requires stages of refinement. And if theories undergo significant developments, driven by responses to anomalies, then theory change must be studied as an incremental process.

Theory change cannot always be viewed as, first, a brief moment of discovery when a theory springs full-blown from the mind of a single creative scientist, which, then, is followed by the logical task of justifying the fully formed new theory. Instead, the development of a new theory may occur gradually, over a significant period of time, either through the work of a single scientist or through a group process. Methods for effecting such development need to be studied.

My task is to explore the dynamics of theory change. The historical sections of this book focus on the development of a theory as evidenced in the public record, namely the published scientific literature, as the theory changes over time. The philosophical goal is to find *strategies of theory change*, reasoning strategies that *could have* produced the changes that did occur. At first glance, strategies based on the published record may not seem to offer much evidence about ways to develop new ideas. I hope to convince the reader otherwise.

One historical case of theory change will be examined in detail—the theory of the gene, in the period 1900-1926. In 1926, Thomas Hunt Morgan, a prominent American geneticist, succinctly stated the theory of the gene that had developed since the 1900 rediscovery of Mendel's 1865 work on heredity. By 1926, all the major theoretical components of the theory of the gene had been developed. This case is a well-circumscribed episode of the development of a successful theory, with numerous failed hypotheses along the way. The historical sections of the book trace the changes in the major theoretical claims within Mendelian genetics over approximately thirty years, from 1900 to about 1930. By 1930, they had reached the form that has now become textbook knowledge in genetics.

The historical changes provide the basis for an analysis of strategies of theory change that could have produced the historical changes that did occur. In suggesting strategies, I wear my philosopher's hat, not that of the historian. The claim is not a historical one, that any particular person consciously employed a specific strategy. Instead, the claim is that a change between a component of the theory at one time and that component at a later time "exemplifies" a strategy. The philosophical sections of the book suggest strategies that could have produced the historical changes that did occur. The strategies are named and are characterized sufficiently generally that they may be studied to see if they are applicable in other cases, either historical or contemporary. The names of strategies will be in **boldface** type to highlight them in the subsequent discussions.

What, the reader may now be asking (and may well have asked a while back), is a *strategy* for theory change? An example is helpful. A theory may change or be changed in order to resolve an anomaly. One example of a strategy for resolving an anomaly is to *complicate* an existing part of the theory. Suppose one component of a theory is the claim that there exist one-to-one relations between two types of things. The anomaly points (somehow, not specified in this example) to a problem with that simple assumption. One way of changing the theory is to complicate the one-to-one relation and to propose a many-to-one relation. Such a theory change exemplifies the strategy that I call **complicate an oversimplification**. Next, in order to assess the adequacy of such a change, the new theoretical component (many-to-one relations) would be assessed, using strategies for theory assessment, such as whether it adequately explains the anomaly. The strategy of **complication** is an example of a general strategy of taking a part of the theory and changing it slightly in order to account for anomalous data. Complicating existing ideas in small ways (stepwise refinement) is one of the easiest ways of getting (slightly) new ideas; however, other strategies are needed when the change in theory must be more drastic. Assessment of **explanatory adequacy** is one strategy often used in evaluating theories; that is, a theory is usually expected to explain a domain of phenomena. As Chapter 2 will argue, how and when to apply criteria of theory assessment are strategic decisions; thus, such criteria are labeled "strategies." Chapter 2 also will discuss my use of the term "strategy" more extensively.

The historical analysis here aims at an accurate characterization of historical changes in the components of the theory of the gene over time. The philosophical analysis aims at finding some of the general strategies that are exemplified in those conceptual changes. Such strategies can be divided into three types: (1) strategies for producing new ideas, such as the use of analogies or the use of specific interrelations to entities or to processes studied by another field; (2) strategies for theory assessment, such as the assessment of a theory's predictive adequacy or its consistency with other accepted theories; and (3) strategies for anomaly resolution and expansion of scope, such as generalizing, specializing, or complicating a hypothesis.

These types of strategies for theory change cut across, and so are not easily mapped into, the categories of discovery and justification, categories which are often discussed in philosophy of science. It is indeed the case that some strategies for producing new ideas may be characterized as discovery strategies, and most of the strategies for assessing theories play a role in justification. However, strategies for assessment may also function in producing new ideas, as we will see. Moreover, use of some strategies for producing new ideas obviates some ways of assessing the theory. Similarly, using one strategy at an early stage may guide the use of a particular anomaly resolution strategy at a later stage. Furthermore, strategies for anomaly resolution and expansion of scope are not easily categorized as either discovery or evaluative methods.

In explaining what I take my current task to be, namely finding strategies of theory change, it is also useful to say what I am not doing. My task as a philosopher of science is not to do psychology, not to delve into the private mental life of the creative scientist. Furthermore, I am not trying to test a general model of scientific change, such as that of Kuhn (1970), Lakatos (1970), Toulmin (1972), or Laudan (1977), or provide an alternative one. Much of the work in the interdisciplinary area of history and philosophy of science has focused on applying these general models to specific cases and showing the ways they apply, or, more often, do not apply. I am not optimistic about any single model of scientific change being adequate. A piecemeal consideration of numerous types of claims made by the models is a more worthwhile task, as Laudan and others (Laudan et al., 1986) have recently argued. Some of those claims, such as the relations of theory and data, will be mentioned in the context of discussing the gene case. The purpose of this work, however, is not to test those claims. I am engaged in a different enterprise.

Another thing I am not trying to do is to place scientific ideas in their social context, as is the fashion in recent externalist history of science (e.g., Sapp, 1987) and sociology of knowledge (e.g., Latour and Woolgar, 1986). I have chosen to approach the material in an internalist way. I recognize that social and cultural factors influence actual scientific developments, but, as a *philosopher*, my interests focus on scientific reasoning and the development of scientific ideas. In so far as my search for reasoning strategies is successful, those strategies will transcend the social context in which they are found in this one historical case. They can be used by scientists in other social contexts and, possibly, by computer programs that do scientific reasoning relatively independent of a social context.

In addition, my primary purpose is not to write a new historical narrative or provide new historical interpretations of this case, although some new historical interpretations are presented here. My choice of episodes from the history of genetics is selective, chosen to serve my purpose of tracing the most important empirical and conceptual developments leading to the theory of the gene. In Chapter 4, more will be said about my philosophical approach to the history.

The aim of the historical reconstruction is to be faithful to the historical record. I want to portray what the actual data and theoretical components of Mendelism were during specific periods, not what they should have been, given some model of scientific change. My account lays out actual historical changes. The aim of the philosophical analysis is then to find general strategies, which I claim are "exemplified" in such historical changes. The strategies are my own proposals of methods that could have produced those changes. In brief, the historical changes provide my data and the strategies are my working hypotheses about methods for producing such changes. In order to emphasize these differences, the historical sections of the chapters are separate from the philosophical analyses of strategies.

Thus, this book approaches the study of the development of new ideas, not by searching for a single logic of discovery nor by studying the psychology of creative people nor by looking for the social influences on scientists' behavior, but by searching for reasoning strategies for producing new scientific theories.

A guide through the topics covered in the chapters may be helpful:

Chapter 2 discusses philosophical issues, including prior work on the three types of strategies: strategies for producing new ideas, strategies for theory assessment, and strategies for anomaly resolution. It also contains an introduction to some of the metascientific vocabulary to be used in this analysis. The final section introduces an oversimplified diagram of stages and strategies of theory change, a diagram that will be refined in subsequent chapters.

Chapter 3, "The Problem of Heredity," introduces the scientific problem confronted when trying to understand the hereditary process. It was this problem that the theory of the gene solved. The chapter aims at enlivening the reader's historical imagination. It provides a hypothetical path of reasoning to several of the important early discoveries. It challenges the reader to rethink the problems that might have faced the scientists at the time and to engage in a creative problem-solving process so as to reach solutions. It also serves the function of introducing the reader to early Mendelian genetics.

Chapter 4, "Historical Introduction," switches from the hypothetical mode of discussion found in Chapter 3 to actual history. After a brief introduction to the historiographical method that I will be using, the chapter presents recent historical interpretations of Mendel's work and of the beginnings of genetics in 1900. It is based on both primary sources and recent historical interpretations of Mendel's work.

Chapter 5, "Mendelism, 1900-1903," presents an analysis of the theoretical components of Mendelism in the period 1900-1903 and justifies that analysis by extensive evidence from the literature of the period. It thus sets the stage for tracing the changes that followed. Tables 5-1 and 5-2 summarize the domain to be explained and the theoretical components of Mendelism that explain that domain, as of 1903.

Chapters 6 to 10 trace the important changes in Mendelism in the period from 1900 to 1926, including the discovery of multiple factors and multiple alleles, the development of the chromosome theory, the tests of the generality of segregation, the discovery of Mendel's second law of independent assortment as a separate law, the discovery of linkage and crossing-over, and the development of the new problems about mutation and the physical nature, location, and functioning of genes. Failed hypotheses along the way include the presence and absence theory, the hypothesis that genes contaminate each other, three-dimensional (as opposed to linear) linkage relations, the reduplication hypothesis, and the idea that genes were autocatalytic colloidal particles. Separate sections of each chapter analyze strategies that are exemplified in the changes documented in the historical sections.

Many of the developments discussed in Chapters 6 to 10 were driven by efforts to resolve empirical anomalies. In contrast, Chapter 11, "Unit-characters to Factors to Genes," focuses on conceptual problems, problems about the nature of a new theoretical entity—the gene. Conceptual developments included issues about appropriate genetic symbolism and appropriate terminology. Also at issue was the question of what properties were to be ascribed to the newly postulated theoretical entity. The field began with a vague claim about unit-characters and gradually developed explicit claims about the existence of a new theoretical entity. No single moment of discovery produced the new idea of the gene. Changes in the lists of properties ascribed to the gene are depicted. The chapter concludes with a discussion of strategies for finding and solving conceptual problems.

Chapter 12, "Exemplars, Diagrams, and Diagnosis," introduces a more diagrammatic representation of the theory of the gene, in contrast to the analysis in the previous chapters in which the theory is stated in sentences. A contrast is made between diagrammatic and sentential methods for representing theories. The diagrammatic representation shows how exemplary patterns can function in explanation and how "model" anomalies function to provide new patterns. Furthermore, the chapter discusses how this alternative, diagrammatic representation is useful for localizing anomalies that the theory faced, as well as giving indications as to how to resolve them. Analogies between diagnostic reasoning and strategies for anomaly resolution are also explored.

Chapter 13, "Genetics and Other Fields," summarizes the problems that the theory of the gene had solved by 1926 and the problems remaining, both old ones left unresolved and new ones that the theory itself raised. These problems are analyzed in terms of the successful and unsuccessful interrelations between genetics and other fields, including cytology, evolutionary studies, embryology, and biochemistry. The final section of the chapter summarizes and extends the strategy of using interrelations between different bodies of knowledge.

Chapter 14, "Summary of Strategies from the Historical Case," summarizes the strategies exemplified in the development of the components of the theory of the gene between 1900 and 1926. This chapter can provide a "quick read" for the major changes in theoretical components and the strategies they exemplify, which are discussed in Chapters 6 to 13.

Chapter 15, "General Strategies for Theory Change," takes a more "theory-driven" approach to discussing strategies. The previous chapters use a "data-driven" approach: they discuss strategies that emerged from an analysis of historical data supplied by the gene case. Chapter 15 places those strategies into more systematic lists of types of strategies. The more systematic lists were generated by relating the strategies from the gene case to previous work in philosophy of science, cognitive science, and artificial intelligence (AI). The oversimplified diagram of stages and strategies introduced in Chapter 2 is "complicated" in the light of the preceding analysis.

Chapter 16, the conclusion, discusses implications of the results of this case study for analyses of theory structure, theory development, and strategies for theory change. It also discusses possible uses of strategies by scientists, science teachers, and researchers in AI.

The readers who will be most familiar with the problems I am trying to solve are philosophers of science, especially those of us who are called "friends of discovery" (Gutting, 1980; Nickles, 1980a). Still, the book is written for a wider audience. I hope that scientists will find the strategies suggest methods for dealing with current problems. Readers not familiar with Mendelian genetics but interested in the broader issues about strategies for theory change are introduced to Mendelian genetics in Chapter 3.

Historians of genetics will be familiar with this case study; the historical sections of the book serve as an updated version of older histories of genetics, incorporating recent historical interpretations of Mendelian genetics and extending those interpretations in some places.

I have found the work of cognitive scientists and AI researchers helpful. For example, I used some results and terminology from their study of general reasoning strategies as I puzzled over what strategy was exemplified in a historical change and then tried to devise terminology for naming that strategy. Chapter 6 discusses the strategy of generalizing and specializing so often used in studies of machine learning (that is, AI work to devise computer programs that can learn and improve their own performance). Chapter 11 attempts to develop a frame-structured representation for the properties attributed to the gene. Frames are one of several structures that AI has developed for representing knowledge so that computer programs can use it to perform tasks to simulate the reasoning of human experts. Chapter 12 draws on the AI literature in diagnostic reasoning and functional representations, in order to devise strategies to diagnose and fix faults (anomalies) in the theory of the gene. Exciting new methods are emerging in the new area called "computational philosophy of science" (Thagard, 1988), which combines cognitive science, AI, and philosophy of science. Colleagues and I have done some preliminary AI experiments to explore ways of implementing the strategy of **using interrelations**. Although discussion of computer implementations of strategies is, for the most part, outside the scope of this more historical book, I hope my less precise characterizations of strategies here will be suggestive for that more formal work. AI implementations hold promise for transforming philosophy of science into an experimental field, as will be discussed in Chapter 16.

Furthermore, I have developed the idea of "interfield theories" (Darden and Maull, 1977; Darden, 1980b) more extensively in this book. Those who found that work suggestive for analyzing interdisciplinary interactions may wish to concentrate on the strategy of **using interrelations** in hypothesis formation, which is discussed in Chapters 7, 9, 10, 13, 14, and 15.

I believe this work has some implications for science education and the teaching of philosophy of science, both by human teachers and via instructional software. I will mention those implications in the conclusion.

Much of my discussion of general strategies in Chapter 15 is preliminary and suggestive. My hope is that others in diverse disciplines will find this effort of interest and will provide constructive criticisms, additional cases of the use of these strategies, and refinements and additions, as we explore methods for developing new ideas.

CHAPTER 2

Philosophical Preliminaries

2.1 Introduction

This book will discuss three types of strategies for theory change: strategies for producing new ideas, strategies for theory assessment, and strategies for anomaly resolution. This chapter briefly discusses some of the previous work relevant to each type of strategy, makes clearer the goals of this analysis, and introduces some of the terminology to be used. The final section of the chapter introduces a diagrammatic representation of stages in theory construction that will be refined as the analysis proceeds.

2.2 Strategies for Producing New Ideas

Some literature in the philosophy of science exists that discusses strategies for producing new ideas, but most work on this subject has been done outside twentieth-century philosophy of science. Such lack of attention is a consequence of philosophers' skepticism about the possibility of producing a logic of discovery. Philosophers of science in the twentieth century have distinguished sharply between the logic of discovery and the logic of justification (or falsification). Most have concluded that there is no (or, more strongly, there cannot be) a logic of discovery. Hence, the task of philosophy of science, on this view, is to analyze methods for assessing the warrant of knowledge claims, in other words, to provide criteria for the justification of theories (e.g., Popper, 1965; Laudan, 1977; Losee, 1987).

The hypothetico-deductive method has dominated many characterizations of science: a hypothesis is guessed (somehow) and a consequence is deduced. If the prediction holds, then the hypothesis is confirmed (or not falsified); otherwise, it is to be discarded and a new one (somehow) proposed (e.g., Hempel, 1965). This characterization, however, leaves as a complete puzzle the reasoning to hypotheses, assuming that they arise in a fully developed form ready for testing. Furthermore, the view that the testing of theories is an all-or-none act neglects the important information that anomalous data may provide for the next stage of theory construction. On a Popperian view, a falsifying instance (an anomaly) requires that the theory be discarded and that a new trial and error process be started to find a replacement (Popper, 1965). This procedure of "conjectures and refutations" neglects the information that the prior refutation can play in guiding the construction of the next conjecture.

Philosophers of science concerned with understanding scientific change have advanced the discussion by postulating relations among successive theories. Lakatos (1970), for example, discussed progressive "research programmes" that consist of a succession of theories, each an improvement on its predecessor. Kuhn (1970) argued for a view of science as successive "paradigm" changes, provoked by an anomaly that (somehow) causes a crisis. Laudan (1977) analyzed changing "research traditions" and suggested adding a "context of pursuit" between discovery and justification. However, none of these philosophers concerned with conceptual change has proposed methods for the development of new ideas for the next stage. Nor have they indicated how anomalies at one stage can guide refinements. Laudan (1980) even argued that a search for such methods should not be the concern of philosophers.

N. R. Hanson was one of the few philosophers who tried to analyze reasoning in discovery. He argued that scientists do not start from hypotheses, as the hypothetico-deductive method suggests. Instead, discovery begins with "problematic phenomena requiring explanation" (Hanson, 1961, p. 34). Reasoning in discovery, Hanson claimed, required a logic different from either deduction or induction. He called this logic "abduction" (using Peirce's term) or "retroduction":

> Schematically, [retroductive reasoning] can be set out thus:
> (1) Some surprising, astonishing phenomena p-1, p-2, p-3 . . . are encountered.
> (2) But p-1, p-2, p-3 . . . would not be surprising were a hypothesis of H's type to obtain. They would follow as a matter of course from something like H and would be explained by it.
> (3) Therefore there is good reason for elaborating a hypothesis of the type of H; for proposing it as a possible hypothesis from whose assumption p-1, p-2, p-3 . . . might be explained. (Hanson, 1961, p. 33)

Hanson said little about the reasoning that occurred between Steps 1 and 2, that is, how the puzzling data suggested a type of hypothesis. He did indicate that analogies might play a role (Hanson, 1961, p. 25). Furthermore, he continued: "We can have good reasons, or bad, for suggesting one kind of hypothesis initially, rather than some other kind" (Hanson, 1961, p. 21). "A logic of discovery, then, might consider the structure of arguments in favor of one *type* of possible explanation in a given context as opposed to other types" (Hanson, 1961, p.30).

Hanson thus gives tantalizing hints about reasoning in discovery, especially his suggestions that puzzling facts provide a beginning point for theory construction and that analogies point to *types* of hypotheses. His work is exciting in that it poses the problem of discovery as a philosophical one. Nevertheless, Hanson's view is not without problems. First, he placed too much emphasis on the role of data and too little on ideas from other sources in hypothesis formation. Second, Hanson omitted discussion of constraints on the generation of possible hypotheses, other than the requirement that the hypothesis explain the puzzling data. Achinstein (1987) criticized retroduction as allowing "crazy" hypotheses that are wildly implausible, based on accepted background knowledge. For example, he proposed, suppose you see that I am happy about news I just received. Suppose you have the background information that receiving news of winning the Nobel prize makes people happy. According to the retroductive schema, since my happiness would not be surprising if I had received news of winning the Nobel prize (Step 2), then you have good reason for elaborating the hypothesis that I received news of winning the

Nobel prize (Step 3). Yet additional background knowledge that I am a philosopher and Nobel prizes are not given for work in philosophy makes the hypothesis wildly implausible. Achinstein's critique shows that additional constraints must be placed on hypothesis formation, other than that the hypothesis explain the puzzling data. Consistency with other accepted claims is one such constraint. More will be said about recent work on abduction in Chapter 15, in the context of developing the strategy of **assess a theory in relation to its rivals**.

The most serious failing of the retroductive schema as a method for reasoning in discovery is that a hypothesis, or at least a type of hypothesis, must already be available (Salmon, 1967, p. 113). The step of constructing such a hypothesis occurs between Hanson's Steps 1 and 2, but actual theory construction was omitted in Hanson's scheme. In his discussion, Hanson did hint that analogies might play a role, but he did not develop methods for using analogies in hypothesis (or hypothesis-type) construction. Schaffner (1980) correctly criticized Hanson for failing to distinguish between (1) a logic of generation and (2) a logic of preliminary evaluation.

Another problem with Hanson's view is that he dichotomized science too sharply into reasoning in discovery and reasoning in justification. Types of inference that play a role in the construction of a theory may also serve to make it plausible (Schaffner, 1974a). Criteria for assessing a theory may be introduced early to provide constraints on generation (Buchanan, 1985). Furthermore, the term "logic of discovery" is misleading. The use of "logic" implies that only one type of inference is found in discovery; also, a comparison with deductive logic suggests that a method guaranteeing the certainty of its conclusions might exist. Strategies for generating possible or plausible hypotheses are more numerous and weaker than the logic Hanson sought. Further, the term "discovery" implies a single event rather than a gradual process of developing a theory over time.

When logic of discovery is recast as the search for strategies for theory development, the work of other philosophers becomes relevant. Shapere's (1974b; 1974a) analysis of the way domain items can provide reasons for proposing a theory of a certain type can be viewed as providing strategies for theory construction; he called such strategies "principles." Philosophers who have studied the role of analogies and metaphors in hypothesis formation, such as Hesse (1966) and Boyd (1979), also can be viewed as discussing strategies for producing new ideas, new hypotheses, and new theoretical terms. Reasoning by analogy will be discussed briefly in Chapters 9 and 15.

The idea that theories may begin as vague ideas that are developed in stages has been suggested by several recent philosophers, such as Shapere (1974b), Monk (1977), Boyd (1979), Gutting (1980), and Nickles (1987b). Moving away from the idea that theories always spring full-blown from the (unconscious?) mind of a lucky guesser to the idea that they may develop in steps over a period of time makes the process of theory development more amenable to analysis. More than the suggestions by philosophers, work by historians and cognitive scientists analyzing the notebooks of scientists provides evidence for the view of an incremental process of theory development in the work of a single scientist. Gruber (1974), Schweber (1977), and Kohn (1980) studying Darwin's notebooks, and Holmes (1985) studying Lavoisier's, have shown the sometimes tortuous twists and turns that were taken before an adequate theory emerged. Similar conclusions have been reached by those studying Faraday's notebooks (Tweney, 1985; Gooding, 1990). Pieces of theory from earlier stages are retained, but new ideas are added along the way.

Furthermore, from the work of computer scientists with interests in artificial intelligence (AI), a new method is emerging to study scientific discovery. The work of Langley, Simon, Bradshaw, and Zytkow showed that methods for discovery could be found that were computationally adequate for rediscovering empirical laws. They said, "Large discoveries take place by the cumulation of little steps, and it is the understanding of these steps and of the processes by which they are accomplished that strips the larger discovery of its aura of mystery" (Langley et al., 1987, p. 58).

Of course, it is an open question to what extent similarities exist between (1) more routine problem solving, (2) an individual scientist struggling to construct a new theory, (3) the incremental development of a theory by a group of scientists over a period of years, and (4) computational methods for reproducing past discoveries. This study of the gene case, an example of (3), will provide further evidence for this new incremental view of discovery.

No one person discovered the theory of the gene. It did not spring full-blown from the mind of a genius or lucky guesser. It began with vague and implicit components that were changed and improved over time by many biologists. Changes in the components of the theory were constrained and directed by a number of factors that show a very intimate relation between the discovery process (or the process of "constructing a new hypothesis," as I prefer to call it) and the process of justification or assessment. Anomalous results often provided strong directives as to where and how to modify the theory. Anomalous data did not supply all that was needed for changing the theory, however. Other factors played roles in suggesting new ideas, and sometimes brought with them a certain amount of justification.

Subsequent chapters will suggest strategies for producing new ideas, all of which are examples of the reasoning pattern: take an old idea and change it to satisfy constraints. Chapter 15 provides a systematic discussion of those strategies.

2.3 Strategies for Theory Assessment

Both discovery and justification will be analyzed as employing strategies. The traditional view, with which I disagree, holds that discovery has only heuristics, while justification has strict rules for deciding the adequacy of a theory. The analysis here will try to make plausible an alternate view: both discovery and justification can be characterized by strategies that may be useful and yet be weaker than a universal method or criterion.

Losee (1987) has called for a descriptive analysis of the strategies for theory evaluation that scientists actually employ. As will be seen, during some of the disputes about modifications to the theory of the gene, different scientists appealed to different criteria of theory assessment. Thus, this case study begins the kind of descriptive analysis that Losee advocates. Determining what criteria a scientist actually employed is easy if explicit appeal is made to a criterion in the published work; at other times, it is necessary to infer what criteria the preferred hypothesis satisfies.

Many criteria of theory assessment have been proposed by philosophers of science; these criteria have been of use in my descriptive task of forming (metascientific) hypotheses about what criteria are exemplified in a given episode. However, the criteria have usually been proposed by philosophers as prescriptions for how to evaluate good scientific theories. The criterion most often proposed has been that a theory be testable or falsifiable (e.g., Popper, 1965). A list of criteria was given by Newton-Smith, which he called the "good-making features of theories." A good theory has: (1) observational nesting—preserving the observational successes of its predecessors; (2) fertility—having scope for further development; (3) a track record—

having a record of past observational success; (4) intertheory support—from being consist with other accepted theories to being able to reduce one to the other; (5) smoothness—making adjustments smoothly in the face of failure; (6) internal consistency; (7) compatibility with well-grounded metaphysical beliefs; and (8) simplicity. Simplicity is a problematic criterion for several reasons, including lack of a criterion for relative simplicity and lack of evidence that simplicity has been a good sign of long-term success in past scientific cases (Newton-Smith, 1981, pp. 226-232). Newton-Smith stresses the role that judgment must play in applying these criteria: "Reasonable men [& women] may be expected to have reasonable disagreements. . . ." That does not mean that anything goes, but at certain times there may be no "knock-down proof of superiority" of one theory over another at those times when a choice must be made (Newton-Smith, 1981, p. 234).

Newton-Smith's discussion is representative, in contemporary philosophy of science, of prescriptions for criteria of theory assessment, although it is a bit more comprehensive than most (see, e.g., McMullin, 1982; Thagard, 1988, ch. 5; Thagard, 1989). Philosophers usually assume that such criteria are to be used to evaluate a well-developed theory at a given time. Like most philosophers, he assumes that theory assessment begins when a theory has been fully developed; like some others, he fails to consider that assessment may play a role in the construction of vague, not yet fully formed, theories. Although Newton-Smith mentions the possibility that compatibility with well-grounded metaphysical assumptions may act as a constraint in theory construction, for the most part he has little to say about the construction of theories.

This book takes a different view. The emphasis here is not the usual one of a philosopher attempting to find prescriptions for how to justify a theory. One of my purposes is to analyze how criteria of theory evaluation can function in strategies for the generation of new ideas. On the one hand, such criteria can be imposed as "constraints in the generator" (a phrase from AI) to eliminate the generation of hypotheses not consistent with them. On the other hand, the evaluation criteria can be invoked subsequently, after the generation stage, thereby producing a less constrained generation process. The former method produces a few plausible hypotheses. The latter method produces a greater number of hypotheses, but they are hypotheses that are less plausible than those produced by the former method. Such relations between generation and assessment will be one focus of this discussion of strategies for theory assessment. Another role played by strategies for theory evaluation is to generate anomalies for the theory, such as empirical anomalies, conceptual problems, or problems of relations to other accepted theories. How strategies for evaluation function in producing different kinds of anomalies also will be addressed. In Chapter 15, I will develop my own list of strategies for theory assessment, a list that shares much with the prior work in philosophy of science.

2.4 Strategies for Anomaly Resolution and Change in Scope

Resolving an anomaly for a theory entails generating a new hypothesis, but that generation occurs in a more constrained context than de novo hypothesis construction. Moreover, the nature of a failure can provide guidance as to where and how to modify a theory. Like reasoning in discovery, methods for anomaly resolution have received comparatively little attention in philosophy of science. Popper (1965), e.g., concentrated on falsifying instances as indicators of the inadequacy of a theory, but gave no hints how to use an anomaly to localize and correct the problem to produce an improved version of the theory. Kuhn (1962) discussed the accumulation of anomalies as sometimes provoking crises which, in some mysterious way, led to the proposal of a new theory or "paradigm." Humphreys (1968, p. 248) argued that historical cases support the view that science often changes through the "exploration, definition, and explanation of anomalies," but he had little to say about strategies for resolving them. Laudan (1977) discussed an interesting set of categories for classifying ways anomalies become solved problems, but he provided no strategies for generating such solutions. Instead of searching for a method of localizing a problem within a theory (as a first step to generating a solution), Laudan argued for spreading the blame for an anomaly evenly among the parts of the theory (Laudan, 1977, p. 43). Laudan's concern was only for how to weight the anomaly in theory assessment, not for how to generate new hypotheses to solve it and produce an improved version of the theory. My discussion of the role that an anomaly may play in *generating hypotheses* contrasts with the usual discussion of the role of anomalies in theory assessment, such as Laudan (1977) discussed.

The first step in anomaly resolution is localization; potential sites of failure within the theoretical components need to be found. Some philosophers have been pessimistic about the ability to localize anomalies, since isolating one faulty component in a complex theory appears too difficult (e.g., see Laudan, 1977, and Quinn's, 1974, discussion of Duhem). My discussion will show that for the gene case such pessimism is not warranted, whatever may be the case for knowledge considered from a more global epistemological perspective. The idea of localization of problems within a theory or a theoretical model, in the light of an anomaly, has received some recent attention in philosophy of science (e.g., Glymour, 1980; Nickles, 1981; Darden, 1982b; and Wimsatt, 1987). Furthermore, researchers in AI are developing methods for determining which parts of a complex, explanatory system are involved in the explanation of particular data points. Their techniques for doing "credit assignment" are relevant to the problem of localizing plausible sites (or even all possible sites, given an explicit representation of all knowledge in the system) for modification, in the face of a particular anomalous data point (Charniak and McDermott, 1985, p. 634). Finding one or more locations in such ways is only the first step in resolving an anomaly. Localization must be followed by the redesign of the failing components.

After localization, the next stage in anomaly resolution is the generation of the new hypotheses to replace the failing components. Shapere (1974b) suggested that when theoretical inadequacies arise, simplifications made in the early stages of theory development are likely areas for hypothesis formation. Wimsatt (1987) suggested how mechanical and causal models might aid in forming hypotheses for resolving anomalies that arise for such models; his analysis extends that of Hesse (1966). She suggested that unexplored areas of an analogy used in the original construction of a theory might function in forming hypotheses to resolve an anomaly at a later stage. AI work on diagnostic reasoning (e.g., Sembugamoorthy and Chandrasekaran, 1986) and on redesign in the light of failures (e.g., Karp, 1989) provides fruitful analogies for reasoning in anomaly resolution (see Chapter 12).

In his delightful imaginary dialogues in *Proofs and Refutations*, Lakatos (1976) discussed "monster-barring," a strategy to improve mathematical conjectures in the light of counterinstances. Lakatos proposed monster-barring as a way of preserving a generalization in the face of a purported exception: if the exception could be barred as a monster, that is, shown not to be a threat to the generalization after all, then it could be barred from causing a change in the generalization. Lakatos was concerned with distinguishing between legitimate exception-barring instances and any illegitimate barring of instances when such instances really did require a change in the conjecture. I will discuss Lakatos's monster-barring, in subsequent chapters, when counterparts to his work in mathematics emerge from the gene case.

When Lakatos (1970) switched from discussions of mathematics to natural science, he did not attempt to discuss such strategies. Instead, he spoke vaguely of the "positive heuristic" associated with each "research programme" and thus implied that the heuristics were domain-specific. (For critiques of Lakatos's positive heuristic, see Newton-Smith, 1981; Nickles, 1987b.) In contrast to Lakatos, I am trying to devise general, nondomain- specific, strategies for scientific theory change. I wonder why Lakatos abandoned that task when he moved from mathematics to science. Certainly domain-specific strategies can be more powerful than vague general ones (Langley et al., 1987, p. 46), but I believe general ones also can be developed.

The focus in this analysis will be on finding general strategies for anomaly resolution. The strategies provide methods for localizing the theoretical component(s) at fault and for providing hints about how to modify the faulty component(s). The gene case, as we will see, provides several episodes in which, first, an anomaly was localized in one or more theoretical components, then alternative hypotheses were generated and tested, and then the theory was changed. It is a rich case for developing strategies for anomaly resolution. A series of steps to follow in resolving an anomaly will be developed in subsequent chapters and summarized in Chapter 15. Although the strategies are extracted from the analysis of a single case, they are not specific to genetics. Some instances of their use in other cases will be discussed in Chapter 15; furthermore, Chapter 16 will suggest that they may be amenable to being programmed into AI discovery systems.

2.5 Descriptive, Hypothetical or Normative Strategies?

A question to ask about the strategies I will be proposing is whether they are *descriptive* of reasoning strategies actually used by scientists or whether they are *prescriptive* of the strategies that scientists should have used (or should use in the future). Neither descriptive nor prescriptive is quite the right characterization; I prefer to call the strategies to be discussed in subsequent chapters "hypothetical." They are my (metascientific) hypotheses about strategies that could have produced the historical changes that did occur. They are not descriptive of strategies that I claim a given scientist *consciously* followed (and I would have no idea how to describe what someone does *unconsciously*). Hence, my goal is not to find descriptively adequate strategies (although those are good to have when possible). The extant historical evidence in the gene case does not provide enough evidence to argue that the strategies describe the actual reasoning processes of scientists. Nonetheless, the strategies may serve as hypotheses about what reasoning strategies scientists might have used. Alternatively, they may be considered hypotheses about what strategies could be followed in such circumstances, if scientists explicitly attended to their methods of reasoning, rather than implicitly using some method or other.

Furthermore, I will not argue that a given strategy was *necessary*; some other strategy might have produced the result. I am exploring some of what may be characterized as the "space of possible strategies." I would like to be able to argue that a given strategy is *sufficient* for producing the change. A good proof of sufficiency could be made if computer simulations could be done, showing that the strategy was sufficient for producing the change (more on this topic in Chapter 16). Discussion of such a rigorous proof of sufficiency, however, is beyond the more historical analysis in this book. My more limited goal here is to argue that a given strategy is a plausible hypothesis for a reasoning method that could have contributed to the change that did occur.

Nickles (1987c; 1987d) outlines a view of methodology that shares some features with mine. (Some ways in which we differ will be mentioned in the next section.) He suggests that methodology should be "descriptive, critical, and advisory." Methods should be descriptive of what has actually worked in the past; they should be critically assessed for their efficiency; they should provide guidance to inquiry (Nickles, 1987c, pp.126-127). I am less optimistic about finding descriptively adequate strategies than Nickles is; he underestimates the difficulties of using extant historical evidence to find the actual strategies used by scientists. Nonetheless, I share with him the belief that the critical examination of past scientific work can yield valuable insights about reasoning strategies (Darden, 1987). (What I call "strategies," he calls "heuristics"; the next section will discuss the relation between my usage of strategy and his use of heuristic.)

Moreover, I agree that finding inviolable prescriptive rules should not be the goal of philosophy of science. Probably no infallible logic for producing new theories exists; certainly none has yet been found. The search is not for strategies that invariably produce correct hypotheses, but for strategies that produce a range of plausible ones, given the constraints at a particular time. The search is also for strategies that aid in narrowing the range of plausible candidate hypotheses, that is, for strategies for hypothesis assessment, as well as hypothesis generation.

An evaluation of heuristics, as proposed by Nickles, is also an important part of the "normative theory of discovery" outlined by the AI researchers, Langley, Simon, Bradshaw, and Zytkow, in their book *Scientific Discovery: Computational Explorations of the Creative Processes*. They said:

> We think normative statements can be made about the discovery process. Such statements will take the form of descriptions of good heuristics, or of evaluative statements about the relative merits of different heuristics or other methods. The evaluative statements can be generated, in turn, either by examining historical evidence of discovery or failure of discovery or by constructing computer programs that incorporate the heuristics and then testing the efficacy of the computer programs as machines for making discoveries. (Langley et al., 1987, p. 54)

Some evaluation of strategies will be done in subsequent chapters when I believe I have adequate evidence to make an evaluation. (More will be said in Chapter 4 about my use of history in doing this.) The normative task of critically evaluating strategies, as well as the advisory task of arguing for the future use of good ones, are not the primary tasks of this analysis, however. These important tasks must await further work, as Chapter 16 will suggest. I have found it sufficiently challenging for this project just to construct hypothetical strategies that I claim are "exemplified" in the historical changes in the gene case.

Another question to ask about the strategies to be discussed here is the sense in which they are "general" strategies. A contrast can be made between domain-specific strategies and domain-independent ones. Nickles said:

> . . . a useful methodology will go beyond the vapid generalizations characterizing all scientific activity (and indeed almost all intellectual activity) which have dominated recent philosophy of science. This means that methodology no longer will be a single, unitary subject but will, at the more interesting levels, break down into domain- and context-specific rules, practices, and advice. (Nickles, 1987c, p. 127)

Certainly, domain-specific strategies can be powerful ones. Yet I believe more general strategies also can be formulated. They may, nonetheless, be context-specific—for example, in this kind of context, employ this kind of strategy (or perhaps one of several kinds). The strategies to be discussed here are not domain-specific, in the sense of being of use only in Mendelian genetics.

Instead, they are formulated in a general way so that they can serve as candidates for use in other, relevantly similar problem contexts. The strategy of **complicate an oversimplification**, for example, is exemplified in many historical cases; however, it will be of use only in a context in which an oversimplification is (or may be) responsible for anomalous results. Thus, the strategies to be extracted from this case are general in the sense that they are hypotheses about plausible reasoning methods that may be used in analyzing other historical cases or that may be used in solving problems in relevantly similar contexts.

This search for general strategies is not a general model for scientific change of the kind proposed by, for example, Kuhn (1970) and Lakatos (1970). In so far as any general model of science is being developed here, it is the view of science as a problem-solving activity. Scientific changes are viewed as responses to problems; context-specific, explicit, problem-solving strategies are being sought. To say that is not to do much more than make another of what Nickles called "a vapid generalization" about all of science. What is interesting is the search for problem-solving strategies, an open-ended search this book only begins.

2.6 Metascientific Vocabulary

A few words about words are in order because so many (but not all, I think) philosophical disputes hinge on the meaning of terms. Current philosophy of science has a zoo of terms for discussing science. The old dichotomy between "theory" and "observation" has proved too impoverished and problematic to label all the parts of science that we wish to discuss (Suppe, 1977). Out of the possibilities, I am going to be using "domain," "hypothesis," "theory," "anomaly," "exemplar," "field," and "interfield theory." I have also found it (perhaps unfortunately) necessary to appropriate a few more ordinary language terms and give them my own technical meanings: "theoretical component," "interrelation," and (most importantly) "strategy."

Now, some clarifications and worries about the words I will be using:

Domain: A useful term for denoting the data or generalizations that a theory is expected to explain is Shapere's term "domain" (Shapere, 1974b). The domain is a set of items, which, on the one hand, may be data produced by observations or experiments. On the other hand, domain items may be general claims that have previously been labeled "empirical laws" or "explanatory theories"; they become part of the domain when they are judged to be well confirmed and are themselves in need of explanation. All the items of a domain are currently accepted as "facts," are somehow grouped together, and are judged to be in need of a single theory to explain them. The scope of the domain (and thus the explanatory scope of the theory) refers to the number of (kinds of) items in the domain. A more general theory has a domain with a larger scope; a more specific theory a smaller one. As we will see, it may be a matter of dispute which items should be inside and which items should be outside the domain of a given theory at a given time.

Hypothesis, Theory and Theoretical Component: The term "theory" is used in two very different ways in common parlance. First is the sense of being hypothetical, not yet proven, as in "It's only a theory." Second, it can refer to a well-supported, general, explanatory claim, as in "The theory of the gene was widely accepted by 1930." I will use "theory" in the second sense.

Philosophers have taken a very strict attitude about what counts as a change from one theory to another. For some, those with a view that theories can be explicitly stated, any change in any claim is usually characterized as a change from T1 to T2. This notation obscures what has remained the same from one version to the next, and so relations between stages of theory formation are not apparent. Some higher level unit, such as a research program or research tradition, then needs to be introduced, and the different theories are placed in it to show that they are related. I will not take such a strict attitude; changes can occur in components of a single theory without its becoming a different theory.

Another problem arises about what counts as a single theory in analyses of science. Neither philosophers nor scientists are very careful about the level of generality a claim must have to be called a theory; what some might call a theory, others might see as a component of a more general theory. Universal criteria for individuating theories are difficult to formulate; in each case, judgments have to be made about the components of a single theory and how much they can change before a new theory has arisen.

In order to deal with some of these problems in characterizing scientific theories, I will introduce my own usage of several terms. I want to discuss the development of a very general theory over time, so I will be using the term "theoretical component" to discuss parts of the theory that change over time. Some of these theoretical components were historically called theories themselves. For instance, with respect to the presence and absence theory (discussed in Chapter 6), I am doing a little violence to scientists' usage by calling this a "hypothetical theoretical component of Mendelism." I will use the term "hypothesis" for a proposed alternative to a theoretical component. The principal theory that I will be discussing is the theory of the gene; the early ideas that developed into it will be termed "theoretical components of Mendelism." No fully articulated alternative to the theory of the gene was ever formulated, although it was challenged by those who wanted a cytoplasmic theory of heredity and development, as we will discuss. I shall characterize most of the debates discussed here as debates about "hypotheses" that were candidates for being "theoretical components" of the "theory" of the gene.

Thus, I will discuss changes that occurred during the development of Mendelism as changes in theoretical components. All the components were subjected to criticisms and tests during the development of the theory; some components changed more than others because of the debates. The way specific components function in relation to specific items in the domain will be discussed. Some of the components were modular, that is, relatively independent of other components. Other components were not independent but were systematically connected in complex ways, as we will see. The *modular* nature of the theoretical components will be very important in allowing the localization and resolution of anomalies. The *systematicity* among the components limits their modularity, but serves to give unity and coherence to the theory. The trade-off between modularity and systematicity in the relations among theoretical components will be a topic of discussion.

Anomaly: An "anomaly" is a problem that is a difficulty for an existing theory. Different ways of assessing a theory can produce different kinds of anomalies for it. An anomaly is often generated by data, when a prediction made by the theory fails, that is, when the strategy of assessing a theory as to its predictive adequacy is applied and the theory fails. Other strategies for theory assessment, however, can produce anomalies that do not result from a failed prediction. Laudan (1977) recognized that problems often arise from factors other than anomalous data. Laudan called these "empirical" and "conceptual" problems. The way I am using "conceptual problem," in contrast to Laudan's usage, will be discussed in Chapter 11. In addition to anomalies that suggest an incorrectness in a theory, a theory also can be incomplete (Shapere, 1974b); several instances of incompleteness versus incorrectness will be discussed.

Exemplars, Exemplary Patterns, and Mechanistic Explanation: Kuhn (1970; 1974) discussed the importance of "exemplars," which he characterized as concrete problem solutions in which a formalism (such as a mathematical equation) is applied and given empirical grounding. Chapter

12 will discuss the exemplars of Mendelian genetics and expand Kuhn's analysis. It will be argued that exemplars may serve in the construction of abstract explanatory patterns or schemas (Kitcher, 1981; Schank, 1986; Thagard, 1988). The patterns abstractly characterize mechanisms, which, when they are operating, produce observable data-points as output. Thus, fitting an observation into a pattern is a way of explaining it. A set of exemplary patterns constitutes the explanatory repertoire of Mendelian genetics; diagrams of such patterns provide one way of representing the theory of the gene, as Chapter 12 will discuss.

Field, Interrelation, and Interfield Theory: Philosophers have recognized the need for discussing metascientific units larger than a single theory. Among the many terms, all embedded within a general model of scientific change, are the following: Toulmin's (1972) "discipline"; Kuhn's (1962) "paradigm" and his later (1970) refinement, "disciplinary matrix"; Lakatos's (1970) "research programme"; Laudan's (1977) "research tradition." Although I did not embed it in a general model of scientific change, I contributed to this menagerie the term "field" (Darden, 1974; Darden and Maull, 1977). These terms are not coextensive and deciding which is appropriate in a given historical case has proved difficult. In fact, I think the research program of testing the adequacy of the models and their terms to find an adequate general account of scientific change is a degenerating exercise in history and philosophy of science, even though articles continue to be published doing so. Furthermore, historians and sociologists have coined their own terms for the social units associated with these conceptual units, including such terms as "research school," "speciality," "invisible college," and "discipline" (see,e.g., Geison, 1981; Crane, 1972; Kohler, 1982). Sociologists have tried to devise precise criteria to delineate instances of these groups, such as citation analysis (e.g., Price, 1965). In addition, hierarchical relations between the units have been introduced; one example of such sociological relations in science is that specialities make up disciplines (Zuckerman, 1988). How the sociologists' categories relate to the philosophers' conceptual units is often problematic (Kupferberg, 1989).

I believe that actual areas of science are too various, too fluid, at too many different levels of organization to expect that a single term (or even a hierarchy of terms) can be defined with necessary and sufficient conditions that will be adequate for discussing all of science. I am not trying to give a general model of the organization of all of science, but I need a term to apply to genetics, to cytology, and to more loosely organized areas, such as evolutionary studies and colloidal chemistry studies. I am going to continue to call the conceptual aspects of these areas "fields." In the sense developed in Darden and Maull (1977), a field has: a central problem; a domain to be explained; techniques and methods (unique to it or shared with other fields); concepts, laws, and theories; special vocabulary; and more general assumptions and goals more or less shared by those scientists using the techniques in trying to solve the central problem.

One of the most important ideas in Darden and Maull (1977) (I now believe with the benefit of hindsight), was the list of kinds of relations between two bodies of knowledge, such as identity, part-whole, structure-function, and causal. I will use the term "interrelation" for such relations. Analogical relations will not be designated interrelations; when analogical borrowing of models or techniques from another field occurs, that is a weaker relation between the fields than the more "ontological" relations that I designate interrelations. Analogical relations provide a much weaker form of relation between scientific fields than do specific interrelations: an analogy merely indicates that there is some similarity between them. More will be said, in subsequent chapters, about the contrast between analogy and specific interrelations.

When the two bodies of knowledge can be adequately classed as fields, then the interrelation between them may be called an "interfield" relation or "interfield theory" (Darden and Maull, 1977; Staats, 1983; Bechtel, 1984, 1986, uses "interdisciplinary"). The chromosome theory of heredity is an example of an interfield theory that relates genetics and cytology by claiming that genes are parts of chromosomes. If the two bodies of knowledge can be analyzed as being at different levels of organization in some sort of hierarchy, then the interrelation is an "interlevel" relation (Darden, 1986a; 1986b). Hypothesizing an interrelation between two bodies of knowledge and using one as a guide to form hypotheses in the other is a very powerful strategy for producing new hypotheses, as we will see.

Strategy: A strategy is a method, a procedure, a practice, a principle. Strategies can be codified and taught. In military parlance, most of the strategies I will discuss are a bit more like tactics; that is, I discuss the more localized maneuvers to achieve local ends, rather than general strategies for winning the war or forming a complete theory. Nonetheless, I do not wish to use the infelicitous phrase "tactics for theory change." Strategy has the nonmilitary usage of "method" that is more appropriate here. Pushing the military metaphor a little further, consider that, for example, General Lee's maneuvers can be analyzed in retrospect to find successful and unsuccessful military tactics or strategies. What actually occurred is documented in the historical record. The history "exemplifies" and can be used to extract strategies or tactics, whether Lee consciously planned the maneuvers using such strategies or tactics. (My thanks to Moreland Perkins for this analogy.) Similarly, the theoretical changes, recorded in published scientific literature, can be analyzed to devise and assess strategies for theory change, independently of what occurred in scientists' minds. I will use the term "exemplified" to designate the relationship between a strategy and the historical change that could have been produced by it.

The relation between "strategy" and "heuristic" is problematic. The *Oxford English Dictionary* attributes the first usage of heuristic in English to Whewell in 1860: "If you will not let me treat the Art of Discovery as a kind of Logic, I must take a new name for it, Heuristic, for example" (*OED*, 1971, p. 259). Polya too characterized heuristic as a "branch of study," and said: "The aim of heuristic is to study the methods and rules of discovery and invention." He added: "Heuristic reasoning is reasoning not regarded as final and strict but as provisional and plausible only, whose purpose is to discover the solution to the present problem" (Polya, 1957, pp. 112-113).

The term "heuristic" has come to refer, not to a field of study, but to a reasoning method that produces plausible results. Buchanan discussed heuristic computer programs for forming plausible scientific hypotheses in his "Steps Toward Mechanizing Discovery":

> The traditional problem of finding an effective method for formulating true hypotheses that best explain phenomena has been transformed into finding heuristic methods that generate plausible explanations. The problem of giving rules for producing true scientific statements has been replaced by the problem of finding efficient heuristic rules for culling the reasonable candidates for an explanation from an appropriate set of possible candidates. (Buchanan, 1985, p. 110-111)

Sometimes heuristics are contrasted with algorithms; algorithms are guaranteed to produce correct results, heuristics are not. On the other hand, heuristics that are implemented in computer programs may be considered algorithms (in another sense) because they are procedures that terminate, although they do so without guaranteeing a correct result. (see Barr and Feigenbaum [eds.], 1981, pp. 28-30 for more on "heuristic" in AI).

My use of "strategy" shares much with the use of "heuristic" by AI researchers. I have been reluctant, however, to use the term "heuristic" and have chosen the more neutral "strategy." My strategies are often more general than the heuristics that are built into computer programs. One