Scientific Pluralism

Stephen H. Kellert, Helen E. Longino, and C. Kenneth Waters, Editors

XIX

MINNESOTA STUDIES IN THE PHILOSOPHY OF SCIENCE

MINNESOTA STUDIES IN THE PHILOSOPHY OF SCIENCE

MINNESOTA STUDIES IN THE PHILOSOPHY OF SCIENCE

Also in this series:

Logical Empiricism in North America
Gary L. Hardcastle and Alan W. Richardson, Editors
Volume XVIII

Quantum Measurement: Beyond Paradox
Richard A. Healey and Geoffrey Hellman, Editors
Volume XVII

Origins of Logical Empiricism
Ronald N. Giere and Alan W. Richardson, Editors
Volume XVI

Cognitive Models of Science
Ronald N. Giere, Editor
Volume XV

Minnesota Studies in the PHILOSOPHY OF SCIENCE

C. KENNETH WATERS, GENERAL EDITOR

HERBERT FEIGL, FOUNDING EDITOR

VOLUME XIX

Scientific Pluralism

STEPHEN H. KELLERT, HELEN E. LONGINO,
AND
C. KENNETH WATERS, EDITORS

University of Minnesota Press
Minneapolis
London

Published by the University of Minnesota Press
111 Third Avenue South, Suite 290
Minneapolis, MN 55401-2520
http://www.upress.umn.edu

Library of Congress Cataloging-in-Publication Data

Scientific pluralism / Stephen H. Kellert, Helen E. Longino, and C. Kenneth Waters, editors.
p. cm. — (Minnesota studies in the philosophy of science ; 19)
Includes bibliographical references and index.
ISBN-13: 978-0-8166-4763-7 (hc : alk. paper)
ISBN-10: 0-8166-4763-1 (hc : alk. paper)
1. Science—Philosophy. 2. Pluralism. I. Kellert, Stephen H. II. Longino, Helen E. III. Waters, C. Kenneth, 1956–
Q175.S42367 2006
501—dc22

2006018614

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

12 11 10 09 08 07 06 10 9 8 7 6 5 4 3 2 1

Contents

Stephen H. Kellert, Helen E. Longino, and C. Kenneth Waters

Introduction
The Pluralist Stance

Background and Motivation

In recent years, a number of philosophers as well as some scientists have advanced various forms of pluralism about the theories or methods of science. The general idea is that some natural phenomena cannot be fully explained by a single theory or fully investigated using a single approach. As a consequence, multiple approaches are required for the explanation and investigation of such phenomena. In some cases interest in pluralism is motivated by analysis of particular issues within a science, and in other cases it is motivated by analysis of general philosophical and methodological questions. How pluralism is understood—whether, for instance, it affirms radical ontological or epistemological heterogeneity or merely the diversity of mechanisms in nature—varies from thinker to thinker and topic to topic.

Early discussions of pluralism were usually carried out in the context of debates about the unity of science thesis. In his presidential address to the Philosophy of Science Association in 1978, Patrick Suppes issued a manifesto for pluralism (Suppes 1978). He claimed that the time for defending science against metaphysics (which he took to be the original rationale for the unity of science movement) had passed. A close examination of scientific developments since the heyday of the unity of science movement warranted instead an embrace of pluralism. Suppes argued that neither the languages of scientific disciplines nor their subject matters were reducible to one language and one subject matter. Nor was there any unity of method beyond the trivially obvious such as use of elementary mathematics. With a few notable exceptions, philosophers of science hesitated to take up Suppes's ideas.

Among the exceptions were Nancy Cartwright and her collaborators who explored an alternative vision of the "Unity of Science" offered by the work of Vienna Circle cofounder Otto Neurath, which sees an irreducible variety of scientific disciplines cooperating for concrete purposes (Cat, Cartwright, and Chang 1996; Cartwright et al. 1996). In his contribution to the present volume, Alan Richardson extends this vein by examining the history of unity/disunity themes. He demonstrates that there is greater flexibility in the older ideas than was appreciated in mid- to late twentieth-century philosophy of science.

With regard to more metaphysical issues, Cartwright also has promoted a pluralistic account of a "dappled world" composed of distinct realms (Cartwright 1999). According to this view, laws of limited scope apply to these realms, but the laws form a loose patchwork rather than reducing to a compact, unified set of fundamental laws. John Dupré advanced pluralist positions in both epistemology and metaphysics. His epistemological pluralism seeks to move beyond the search for demarcation of science from nonscience to an account of the epistemic virtues that characterize the variety of scientific enterprises. He argues that the kinds named by conflicting systems of classification are real because they serve the various purposes of the humans classifying things. He calls his metaphysics "promiscuous realism" (Dupré 1993).

Although early work on pluralism tended to focus on issues related to the unity of science thesis, recent discussions have taken up a number of philosophical issues ranging from concrete debates within particular sciences to debates about metascientific concepts to discussions about how philosophical, historical, and sociological accounts of science relate to one another. On the concrete level, pluralism has been invoked to account for the problems concerning interpretations of quantum mechanics (Cushing 1994) and the status of laws in physics (Cartwright 1983), puzzles concerning the relation of quantum mechanics to other branches of physics (Morrison 2001; Teller 2004), the problem of species (Mishler and Donoghue 1982; Mishler and Brandon 1987; Ereshefsky 1998, 2001), the controversy about the level of selection in biology (Dawkins 1982; Waters 1991, 2005), and the relation between genetic and environmental explanations of differences (Longino 2001), to name just a few.

Philosophers of science have begun to advance pluralism at the metascientific level, most notably with respect to epistemic virtues. A variety of views regarding the role, status, and identity of scientific or epistemic virtues has been advanced in the philosophical literature. Some philosophers treat empirical adequacy, consistency, simplicity, explanatory power, and refutability as truth indicators (McMullin 1983). Others treat them as markers of scientificity (Kuhn 1977), while still others emphasize their incapacity to be maximally realized at one time by any given theory (van Fraassen 1989). Despite these disagreements, philosophers typically assumed that there must be one foundational set of virtues, whatever their role or status in science. Hence, some philosophers advocate for one or another within the set (e.g., explanatory power against empirical adequacy and vice versa [Churchland 1985]) or advocate different virtues altogether, such as heterogeneity or social utility (Longino 1996). But now there is talk of a pluralist solution that would claim that which virtues should hold what degree of regulative status in any given research project is a function of features specific to the problem and of the particular aims of the research (Longino 2002).

The appreciation of the need for interdisciplinary approaches in science studies aligns with pluralism at the metaphilosophical level. Because the scientific enterprise is itself a complicated phenomenon, no single disciplinary approach can provide a fully adequate account of its conceptual, technical, cognitive-psychological, social, historical, and normative aspects (Bauer 1990; Stump 1992; Wylie 1995; Giere 1999). The pluralist interpretation might be pushed one step further: no single disciplinary or interdisciplinary approach can provide a full account.

The idea of pluralism is certainly "in the air," but one might ask whether appeals to pluralism, such as the ones mentioned here, are merely opportunistic gestures intended to avoid answering difficult questions. Can pluralism be consistently advanced in philosophical interpretations of science? If so, what are the implications of taking a consistent stand on pluralism? The aim of this book is to answer these questions by investigating a number of topics and areas of the sciences.

Distinguishing between Fact and Interpretation: *Plurality* in the Sciences Contrasted with *Pluralism* about the Sciences

It is useful to distinguish between plural*ity* in the sciences and plural*ism* about the sciences. The former is a feature of the present state of inquiry in a number of areas of scientific research, such as those listed above. These are characterized by multiple approaches, each revealing different facets of a phenomenon. There can be plurality of representational or classificatory schemes, of explanatory strategies, of models and theories, and of investigative questions and the strategies appropriate for answering them. Plural*ism* is a view about this state of affairs: that plurality in science possibly represents an ineliminable character of scientific inquiry and knowledge (about at least some phenomena), that it represents a deficiency in knowledge only from a certain point of view, and that analysis of metascientific concepts (like theory, explanation, evidence) should reflect the possibility that the explanatory and investigative aims of science can be best achieved by sciences that are pluralistic, even in the long run. Philosophers can ground their pluralism in studies of particular cases, in the findings of cognitive science, or in a priori reflection about such matters as the vagueness of scientific predicates. Philosophers who advocate pluralism can and do differ as to the extent of the plurality they attribute to the sciences, the strength of the pluralism they adopt, and the broader philosophical implications they draw from it.

Interpretations of Plurality

Monism versus Pluralism about Plurality in the Sciences

We take *scientific monism* to be the view that

1. the ultimate aim of a science is to establish a single, complete, and comprehensive account of the natural world (or the part of the world investigated by the science) based on a single set of fundamental principles;
2. the nature of the world is such that it can, at least in principle, be completely described or explained by such an account;
3. there exist, at least in principle, methods of inquiry that if correctly pursued will yield such an account;
4. methods of inquiry are to be accepted on the basis of whether they can yield such an account; and
5. individual theories and models in science are to be evaluated in large part on the basis of whether they provide (or come close to providing) a comprehensive and complete account based on fundamental principles.

Scientific pluralism, in contrast, holds that there are no definitive arguments for monism and that the multiplicity of approaches that presently characterizes many areas of scientific investigation does not necessarily constitute a deficiency. As pluralists, we do not assume that the natural world cannot, in principle, be completely explained by a single tidy account; rather, we believe that whether it can be so explained is an open, empirical question. Although we often write "the world," we also question whether *parts* of the world investigated by different sciences (e.g., the world economy or the system within an organism) can be completely accounted for by a single, comprehensive theory. Treating this tenet of monism (tenet 2 above) as an open question rather than as a metaphysical truth undermines the remaining tenets of monism. It undermines tenet 1 because if we don't know whether the world can be fully accounted for by a single comprehensive account, then it seems unreasonable to assume that the ultimate aim of science is to achieve such an account. If the world cannot be fully accounted for by a single comprehensive account, then there cannot be methods of inquiry that if correctly pursued would yield such an account. Hence, we should not assume that tenet 3 is true. And if we don't know whether the world can be fully accounted for by a single, comprehensive account, then it seems unreasonable to accept or reject scientific methods according to whether they can yield such an account (tenet 4) or to evaluate scientific theories in terms of how close they come to providing a complete and comprehensive account (tenet 5).

Monists might admit that a plurality of approaches and models can meet appropriate scientific standards (or satisfy the corresponding epistemic values) but insist that this is only because today's science is incomplete. The ultimate aim of science, according to the ideal of monism, is to have for any given phenomenon the complete description of its essentials. But we do not believe that the plurality in today's science is necessarily a temporary state of affairs. We think that some phenomena may be such (e.g., so complicated or nebulous) that there can never be a single, comprehensive representation of everything worth knowing, or even of everything causal (or fundamental), about the phenomenon. If this is the case, that is, if the nature of the world is such that important phenomena cannot be completely and comprehensively explained on the basis of a single set of fundamental principles, then the aims, methods, and results of the sciences should not be understood or evaluated in reference to the monist quest for the fundamental grail. Hence, we believe philosophy of science should rethink those of its concepts that rule out the possibility that ultimately the best way to investigate and explain the natural world is through multiple investigative approaches and representational systems. This view is supported by the chapters in this book because they establish the possibility that the world is too complicated or too indeterminate and our cognitive interests too diverse for the monist ideals, and they establish this result across a broad swath of sciences including behavioral, biological, physical, and mathematical sciences.

Modest Pluralist Interpretations

Some forms of pluralism acknowledge the present state of plurality, but treat it as resolvable at least in principle. One form of pluralism in the literature recognizes that the world is patchy and that one model or theory might explain phenomena in one patch while a different model or theory would be necessary to explain similar phenomena in a different patch. Some scientists and philosophers advance this view at the local level of scientific domains. Sandra Mitchell's pluralism seems to be of this type: nature varies in its strategies, using different strategies to achieve the same end, but for each situation in the natural world there is a single complete and comprehensive account that can be given (Mitchell 2002). But this view about the disunity of science seems to reduce to (nonfundamentalist or nonreductionist) monism because it is consistent with the idea that for every particular phenomenon, there is a single, best account. Philip Kitcher accepts a stronger view (Kitcher 2001). He concedes that classificatory concepts and systems reflect different interests and that the legitimate persistence of such varied interests permits the articulation of different theories about one and the same phenomenon. Nevertheless, he seems committed to the assumption that all truths in one theory of X must be translatable into truths in the other theories of X. This view, while nonreductionist, also seems to reduce to monism, because it implies that a single, consistent theoretical system can accommodate all explanatory interests. It is consistent with the idea that different true theories of the same phenomenon are, from a logical point of view, only notational variants of one another. Might the plurality of models and representations arise not simply because some parts of the world are different than others, or because different, but intertranslatable, classification systems respond to different interests, but because some parts of the world are so complicated that they cannot be fully accounted for from the perspective of a single representational idiom? We believe it is metaphysical prejudice to deny this possibility, and we fail to see what is to be gained by this denial. We worry that insistence on this abstract metaphysical point sometimes leads philosophers to three errors: (1) to minimize or overlook important differences among scientific approaches, (2) to dismiss from consideration legitimate scientific approaches that seem to lie outside the mainstream, and (3) to exaggerate the explanatory importance of scientific approaches that are in the mainstream.

Another form of modest pluralism tolerates a plurality of theories, not because there is something importantly right about one that cannot be captured by another (and vice versa), but because it is difficult to predict which research program (or preliminary theory) will lead to a theory that provides a complete account of the phenomena. This view endorses a division of cognitive labor in the short term as a means to achieving the putative long-term goal: a single, all-encompassing, true theory. But again, this form of pluralism seems to reduce to monism. Certainly the division of labor is a good strategy under certain conditions of uncertainty, but we believe the plurality exhibited throughout the sciences involves more than the hedging of bets about which approach will lead to the complete and comprehensive account that supposedly awaits discovery.

Radical Pluralist Interpretations

Constructivists admit an indefinite number of theories, the only constraint being human ingenuity. But they are also anti- or nonrealists. A realist version of the thesis is offered in a radical form of pluralism advanced by John Dupré (1993). According to his "promiscuous realism," there are an indefinite number of ways of individuating and classifying the objects in the world, each of which is responsive to different interests, and no one of which is more correct than the others. All are equivalently referential: any kind term that has a role in a system of understanding refers to a kind. There is an indefinite number of sets of kinds. But just as modest pluralisms are difficult to distinguish from a sophisticated form of monism, promiscuous realism is hard to distinguish from radical relativism. We are committed to the idea that there are constraints that limit the variety of acceptable classificatory or explanatory schemes. It is worth emphasizing that the case studies in this book do not deny the existence of such constraints. An important question they address is whether constraints limit sciences to single schemes for the parts of the world (or individual instances) investigated by particular sciences. We believe the question of whether the constraints on scientific inquiry lead to monism should be treated as an empirical question.

An Empirically Based Interpretation: The Pluralist Stance

The form of pluralism we advance in this essay is not based on metaphysical assumptions. We have no a priori basis for assessing the monist assumption that the nature of the world is such that its parts can be completely described or explained by a comprehensive account grounded on a consistent set of fundamental principles. We also lack an a priori basis for affirming (or denying) universal aims for science such as the monist aim to acquire a single, complete, and comprehensive account of the natural world. The form of pluralism we advance on the basis of the chapters that follow is empirically motivated. The argument we sketch begins with considerations from contemporary psychology about human perception and continues by appealing to detailed case studies about how the plurality in particular sciences ought to be understood. The studies in this book lead, we argue, to a substantial and consistent form of pluralism that is not so much a metaphysical or ideological position about the fundamental character of the world as an approach to interpreting the content and practices of scientific inquiry. We call this approach "the pluralist stance," by which we mean a commitment to avoid reliance on monist assumptions in interpretation or evaluation coupled with an openness to the ineliminability of multiplicity in some scientific contexts. (By "we," we mean the authors of this essay. We are advancing this as our own view, not as a collective view of the contributors to this book. We maintain, however, that the arguments offered by the case studies in this book substantiate the view we are advancing here.)

According to the pluralist stance, the plurality in contemporary science provides evidence that there are kinds of situations produced by the interaction of factors each of which may be representable in a model or theory, but not all of which are representable in the same model or theory. Each factor is necessary for the phenomenon to have the various characters it has, but a complete account is not possible in the same representational idiom and is not forthcoming from any single investigative approach (as far as we know). A more complete representation of some phenomena requires multiple accounts, which cannot be integrated with one another without loss of content. We do not hold that for every phenomenon there will inevitably be multiple irreducible models or explanations. We hold that the task of identifying which situations require multiple approaches requires empirical investigation. We believe that the pluralist stance has important implications concerning metascience and the public consumption of scientific knowledge.

The pluralist stance differs from more modest versions of pluralism because it acknowledges the possibility that there may be no way to integrate the plurality of approaches or accounts in a science. For example, we do not believe that the tension among different models can always be resolved by partitioning the domain so that the conflicting models apply to different subdomains. In addition, we do not assume that the plurality of accounts should be consistent, that all truths from one accepted account must be translatable into truths of the other accepted account(s). Perhaps the approaches and accounts within the plurality cannot be combined and perhaps they even disagree with one another about certain points. How might this be possible? In many complicated situations, investigation is not feasible unless investigators parse causes. In some cases, there are alternative ways to parse causes and one cannot parse the causes in the alternative ways at once. Some parsings are advantageous for explaining (and/or investigating) some aspects of the situation, other parsings are advantageous for accounting for other aspects. In such cases, we could say that each account emphasizes some causal aspects of the situation while obscuring others. In fact, an acceptable scientific model might describe some facets of the situation extremely well (e.g., the potential causal influence of some factors) while actually distorting other facets (the potential causal influence of other factors). If this is the case, and if two models distort some of the same aspects, they might distort these aspects in different ways, giving rise to inconsistencies. This is just one kind of situation in which a plurality of inconsistent approaches might be defended.

An epistemology of monism would insist that at most one of the theories is true and scientists should figure out which one. A pluralist would draw this conclusion in some cases, but a pluralist is also open to the possibility that the situation is such that it is impossible to accurately represent all aspects (even all aspects of interest) with a single model. Seeking a proper plurality of models, each of which accurately accounts for some but not all aspects of the situation, might be preferable. What is the advantage of the pluralist interpretation? As some of the chapters to follow show, it provides a means of avoiding senseless controversies that do not lead to progress. It also helps emphasize the partiality of scientific knowledge. The monist interpretation can give rise to inflated confidence about the comprehensiveness of scientific knowledge. A pluralist stance keeps in the forefront the fact that scientific inquiry typically represents some aspects of the world well at the cost of obscuring, or perhaps even distorting, other aspects.

General Considerations in Favor of the Pluralist Stance Including an Argument Based on Perceptual Psychology

Pluralism can be motivated on the basis of abstract considerations: all representations are partial in that any representation must select a limited number of aspects of a phenomenon (else it would not represent, but duplicate). This selective and partial character of representation means that alternative representations of a phenomenon can be equally correct. Hence, it should be obvious that different accounts, employing different representations, might be generated by answering different questions framed by those different representations. Monism holds that all such correct accounts can be reconciled into a single unified account or that there is a single perspicuous representation system within which all correct accounts can be expressed. The related view, *fundamentalism,* holds that there is one (or a very few) law(s) from which all correct accounts (with requisite empirical input) can be derived. The pluralist stance rejects both monism and fundamentalism. The plurality of representations and approaches in science is sustained by the complexity of nature, the employment of highly abstract representational models, and the diversity of investigative, representational, and technological goals.

Ronald Giere, in his contribution to this book, offers a general empirical argument for pluralism drawing on findings in perceptual psychology. Color perception offers a compelling example of perceptual perspectivalism.

Research on human vision identifies two types of light receptor: the rods, which have a uniform range of sensitivity and a common peak sensitivity and which constitute an achromatic system, and the cones. Cones are typically of three types, each containing a different pigment and each characterized by a distinctive peak sensitivity. The chromatic dimensions of human visual experience are produced by differences in activation of the light receptors. These findings about receptors provide explanations of various aspects of human color vision. Visual systems can be, and are, different from that of the typical human. Some species have a four-pigmented system, and many others have only two. Some humans are also only dichromats (e.g., those with red-green color blindness), and some have no conic receptors and thus experience the world only in black and white. Giere argues that there is no way to say which perspective is *correct,* although one perspective might be deemed richer in certain respects than another. He also maintains that there is no translation from one system to another. There is, however, no need to understand claims made about the same object from the vantage of different perspectives as incompatible as long as colors are understood as the product of an *interaction* between objects and particular types of visual systems, rather than as objective properties of objects. Our visual system affords us a particular perspective that may be different from, but is not incompatible with, that of others attending to the same object.

Giere proposes that we understand scientific observation as analogous to unaided vision, mediated by different instruments (e.g., light telescopes versus infrared telescopes), each enabling a different partial perspective and different partial representations of objects. Scientific theorizing could be partial in a similar way, capable of dealing, for example, with mechanical forces or with electromagnetism, but not necessarily capable of dealing with both kinds of phenomena with one set of principles.

Pluralities in Social, Behavioral, Biological, Physical, and Mathematical Sciences

The contributors to this book identify a variety of ways plurality can characterize a particular area of inquiry and the various sources of plurality located within the complex of inquiry and object of inquiry. These include (a) the complexity of the phenomena—whether associated with crossing levels of organization or multiple factors within the same level of organization; (b) the variety of explanatory interests; (c) the openness of constraints—whether from above or below; and (d) the limitations of particular explanatory strategies vis-à-vis the phenomena. These essays span physics, mathematics, biology, and social and behavioral science, showing that plurality is not confined to any particular area of science. The essays also offer a variety of arguments for taking a pluralist stance toward the plurality they document. In this section, we briefly describe the pluralities. In the next section, we will consider why the authors believe pluralism offers the best interpretation of the pluralities they document.

In some cases formal constraints imposed by higher-level theory leave certain options at a lower level, options that are equally supportable given the evidence. Michael Dickson argues that the mathematical constraints of quantum theory are insufficient to pick out one of several dynamics. In their jointly authored essay, Geoffrey Hellman and John Bell show that both classical and intuitionist logics satisfy the basic logical requirement of consistency. In both the physics and the logic case studies, the authors identify pluralities that are preserved because different cognitive interests are satisfied by different formalizations.

Other authors argue that the complexity of the phenomena generates the possibility for a plurality of scientific approaches. Carla Fehr examines the scientific literature on the evolution of sex and identifies a number of different explanations, explanations that are typically viewed as opposing one another. Fehr argues that the multiplicity of explanations stems in large part because sexual reproduction involves processes occurring at multiple levels of organization. Different accounts of the evolution of sex focus on processes at different levels (meiosis at the cellular level and outcrossing at the organismic level) that are subject to different selective pressures and cannot be collapsed into one.

Esther-Mirjam Sent documents an oscillation between monism and pluralism in economics. At the beginning of the twentieth century, both institutionalist and marginalist (neoclassical) approaches coexisted as jointly necessary to fully cover economic phenomena. At midcentury, the neoclassical approach achieved near-hegemonic status (at least in the United States), and its proponents sought to bring all kinds of social phenomena under its uniform explanatory umbrella. The resistance of some phenomena to neoclassical treatment has led a number of economists to think that alternative approaches are necessary for at least some phenomena and thus also to advocate pluralism.

Helen Longino examines scientific studies of behavior. Researchers agree that there is a multiplicity of causal factors involved in behavior. Longino points out that since it is impossible to measure all of them simultaneously, research approaches must inevitably select from among these the ones they will measure. Each selection constitutes a different parsing of the causal universe, creating different effective sets of alternative causes. These support different research approaches, each able to answer a distinctive subset of the possible research questions one might have.

Stephen Kellert's case study is focused on metascience rather than the object sciences themselves. Kellert suggests that interdisciplinarity, understood as the need for multiple disciplinary approaches, arises because of the complexity of the phenomena to be known and the partiality of the individual disciplines.

In other cases, it is not so much the complexity of the phenomena as a certain fundamental openness or indeterminacy that generates plurality. In mathematics, Hellman and Bell note, it has become standard to treat sets as the fundamental mathematical entities. Category theory, however, proposes a mathematical universe constituted by topoi (toposes) or categories. Suitably enriched, category theory provides an autonomous ontology for mathematics, an ontology free of the constraints required to avoid paradox in set theory. Here again, different mathematical interests will dictate which ontology is preferable in a given situation. Economics, too, can, in fact must, support different ontologies. Sent suggests that, once the no-trade theorems force one to give up any assumption that economic agents are uniform, the variety of distributions of different kinds of agent will determine different kinds of economic structures, no one of which is any more fundamental than any other.

Reasons for Favoring a Pluralist Interpretation of the Pluralities Identified in the Case Studies

The pluralities identified in the case studies can be variously interpreted. A monist or modest pluralist will either treat them as temporary—as stages on the way to a unified treatment of the phenomena—or as steps to a comprehensive resolution that will provide for each instance a single, best way to account for the instance. Philosophers and scientists are inclined to monism or modest pluralism for different reasons, requiring different responses from bolder pluralists. For example, some evolutionary biologists adopt a strictly monist perspective and assume that only one of the diversity of evolutionary explanations of sex is correct and have entered into a debate about which account is the right one. Fehr argues against this monist interpretation by pointing out that the persistence of sexual reproduction in a species involves different and continuing costs. It is often impossible to settle on one account even when limiting the domain to a narrow lineage. Which explanation is appropriate depends on the precise question one is asking. Other biologists have advanced one or another form of modest pluralism. One of these suggests that it is possible to decompose sex into its constituent parts, for each of which a distinct evolutionary account can be given. This interpretation corresponds to Mitchell's abovementioned idea that situations of pluralism are resolvable by separating apparently complex phenomena in such a way that the conflicting explanations apply to different cases that are part of a family of related but distinct phenomena. Fehr argues, to the contrary, that the components of sex cannot be separated in the way demanded by this modest form of pluralism. Other biologists have dealt with the plurality of explanations by holding that the different explanations must be integrated in order to identify a net resultant force responsible for the emergence and persistence of sex. Fehr holds that combining or integrating the explanations would have the effect of decontextualizing them and depriving each of the detail and information that is the source of its explanatory value.

Kenneth Waters argues that the pluralist interpretation applies even to cases where science doesn't exhibit much plurality. His case study involves genetics and molecular biology. Scientific explanation and investigations in these sciences are largely centered on the role of genes. Philosophical critics have advanced an alternative, called developmental systems theory (DST), that treats organisms as systems and genes as just one of many different kinds of equally important developmental resources. They argue that DST should replace the now-dominant gene-centered approaches because the gene-centered approaches leave too much out. Thus, the proponents of DST contend that only an approach that incorporates all the causal factors and their interactions can be correct. Waters maintains, contrary to the critics, that gene-centered accounts are not incorrect. Rather, they are partial accounts of complex processes that could be approached in a variety of ways. Gene-centered accounts provide correct answers to some, but not all, of the questions that can be asked about development. Nongenic factors of a system, e.g., cytoplasmic elements, at the same level of organization as genes (intracellular), could be emphasized in one's research questions, leading to different but not necessarily contradictory accounts of particular developmental processes. Waters claims that the monistic call for comprehensiveness obscures the significant achievements of approaches, like the gene-centered one, that focus attention on only one kind of causal factor.

Waters's case is different from the other cases examined in this book because the other cases argue for pluralist interpretations of sciences exhibiting a plurality of theories or approaches. Waters argues for a pluralist interpretation of a science that does not exhibit a plurality of theories or approaches. He argues that the problem with the monistic interpretation of his case is that it leads proponents of gene-centered science to infer that because the science is successful it must be based on a comprehensive theory that can explain all the essentials of development (genetic determinism). Opponents reject this conclusion because they recognize that the theory behind molecular biology is gene-biased and obscures a lot of factors crucial for development. But monism leads the opponents to conclude that the success of molecular biology is illusory and to seek a replacement. Waters argues that a pluralist epistemology can enable us to acknowledge that gene-centered molecular biology is successful without buying into the idea that the gene-centered perspective offers a comprehensive account of the essentials of development.

Wade Savage takes a somewhat similar line with respect to neuroscience. Although research into sensory and motor capabilities seems to vindicate physicalism (as opposed to dualism), Savage explores the possibility that the psychophysical identity principle that underwrites this research should be interpreted as a methodological principle, consistent with dualism. He proposes that there are multiple senses of identity and that the apparently conflicting conclusions reached about physicalism and dualism can be resolved by distinguishing between empirical identity (the sense at issue in the methodological principle) and logical identity (the sense at issue in contemporary defenses of dualism).

Other contributors (Hellman and Bell, Longino) point out that monism on the part of researchers, especially when motivated by commitment to their chosen theory or approach, fuels sterile and unproductive debates. Adopting a pluralist attitude encourages scientists to pursue interesting research without having to settle questions that cannot, in the end, be settled.

Philosophers advocating monism or modest pluralism worry that tolerating any stronger form of pluralism is equivalent to tolerating contradiction. Thus Kitcher, as noted above, constrains his pluralism by requiring that different languages in which different theories are expressed be intertranslatable so that a truth in one can be translatable into a truth in the other. A pluralism that tolerates inconsistencies is apparently an invitation to incoherence. But Dickson maintains that inconsistencies among different dynamics for quantum theory should be tolerated. He argues that solving the measurement problem requires supplementing quantum theory with a dynamics. Although constraints rule out many dynamics, a number of alternative dynamical accounts are consistent with quantum theory (and with the empirical predictions made on behalf of quantum theory). It turns out that a single dynamics will not serve all the explanatory goals of physicists. To illustrate this point, Dickson considers two different explanatory contexts that call for quantum theory to be supplemented by a dynamics. One explanatory context requires invoking the principle of relativity, and the other context requires the principle of stability. It turns out that no single dynamics is consistent with both principles. Hence, providing an explanation in one context requires supplementing quantum theory with a dynamics that violates the stability principle, and the other context requires supplementing quantum theory with a dynamics that violates the principle of relativity. Dickson argues that physicists should tolerate a contradiction among dynamical accounts because the multiplicity of contradictory accounts is needed for explanatory purposes and because the contradictions do not lead to contradictory predictions about the observables. This is perhaps the clearest example illustrating the following point, which modest pluralism overlooks: there can be a tension within the plurality of accounts even though each account correctly describes, models, or explains an important aspect of the same part of the world toward which it is aimed.

Contributors to this book hold not only that the situations they analyze resist requirements of monism or modest pluralism, but also that scientific knowledge would suffer by their imposition. Scientists sometimes must make decisions about whether to pursue or to defer the quest for comprehensive or convergent accounts. A pluralist approach advocates that such decisions be made on empirical, case-by-case, pragmatic grounds rather than on the basis of a blanket assumption. We expect that decisions made on these grounds will yield more fruitful and effective results.

As has been seen, our contributors have a variety of ways of arguing that the strong pluralism they advocate for their respective areas of investigation does not issue in a debilitating contradiction. They argue further that less ecumenical views would result in a loss to knowledge. Tolerating nonconvergence of approaches avoids the mistake of a priori restricting what can be known and how. Thus, Longino maintains that the approaches to behavior she discusses are not intertranslatable because each parses the (same) causal universe differently. Each is nevertheless capable of producing knowledge, and to restrict research to one or to those that produce intertranslatable sentences is to eliminate avenues of inquiry that have produced important insights. Waters makes a similar case with respect to the demand that an acceptable approach for investigating biological development must include all the causal factors. It is simply not possible to design a research program that takes all factors into account at once. Insisting on a single, comprehensive investigative approach or explanatory account will cut off avenues of knowledge.

In addition to avoiding sterile debates, pluralism underwrites the explanatory flexibility that is one of the strengths of the sciences. Fehr notes the loss of information that would perforce accompany attempts to integrate the different explanations of sex. Hellman and Bell note that classical and intuitionist logic each answer to different interests, truth preservation and computability (or constructability) respectively. Neither can be given up, nor should be. Similarly, while not wishing to give up on set theory, they state that the broader ontologies defined by category theory permit the practice of forms of mathematics not possible if sets are taken as the fundamental mathematical entities. Dickson, too, affirms the ineliminability of the explanatory contexts and questions to which the different (and inconsistent) quantum dynamics are addressed. Sent argues that economics will be better able to address the variety of economic phenomena if it embraces a plurality of approaches rather than insisting that one approach must fit all. Finally, as Richardson notes, pluralism enables a deeper connection with social and political concerns than advocacy of a single approach does.

We started from the premise that the world might not yield to the demands of monism. The case studies in this book indicate that science provides good evidence that the world is indeed such that it will not be fully explained on the basis of comprehensive theoretical accounts that identify all the essentials of any given phenomenon. It appears that some parts of the world (or situations in the world) are *such* that a plurality of accounts or approaches will be necessary for answering all the questions we have about those parts or situations. But this raises an important question. What is the "such"? That is, what is the nature of the world such that it, and many of its parts, are not amenable to a single, comprehensive account? The answer seems to differ for different patches of the world. For biological and social patches, the world seems too complicated or complex: many processes involve interaction of multiple causal processes that cannot be fully accounted for within the framework of a single investigative approach. For the domain studied by quantum physics, the situation doesn't seem so much complicated in that sense as perplexing. Our ordinary physical intuitions, which work at the level of the classical physics of midsize objects, seem to fail us at the quantum level (see Morrison 2001). So, while our case studies suggest that interactions of multiple causal processes in the biological and social cases could make it impossible to fully account for the phenomena within a single framework, they do not give a clear indication of what could be making it impossible (if indeed it is impossible) in the quantum domain.

We believe that Dickson's contribution provides evidence that the quantum world is such that a comprehensive, monistic explanatory account is not forthcoming. He makes a strong case that accepting a plurality of dynamics serves divergent interests of physicists that cannot be served by a single dynamical theory (or by leaving out a dynamical theory). He reaches this conclusion while maintaining that these dynamical theories are mutually incompatible and that the formalism of quantum theory does not provide constraints for deciding among them. We admit, however, that we do not know how to describe the nature of the quantum world that makes it resistant to a single, comprehensive account. We are, of course, not alone here. Dickson suggests that the alternative dynamical accounts be thought of as complementary, along the lines that concepts involving observables are said to be complementary. Although this suggestion is promising, it is still not obvious that the need to appeal to alternative complementary concepts or dynamical accounts stems from something akin to the need for plurality in the biological and social contexts. But we do not think our inability to describe the "such" in the case of quantum theory, or other cases for that matter, means that we ought to adopt monism by default.

Although we believe that frameworks for the interpretation of science should not presuppose a metaphysics of monism, it should be clear that we ourselves do not have a general metaphysics. We do not, for instance, insist that all parts of the world are such that they cannot be comprehensively accounted for by a single theory. Furthermore, we do not maintain that there is a common ontology shared by those parts of the world that cannot be fully explained in terms of a single, comprehensive account. Our general thesis is epistemological: the only way to determine whether a part of the world will require a plurality of accounts is to examine the empirical results of scientific research of that part of the world. The case studies in this book are consistent with this general epistemological stance. While contributors concede some of the attractions of monism (e.g., unproblematic commensurability and comparative assessment, singularity of approach, hegemony), they show that in the particular cases being examined, plurality is ineliminable. They argue that a strongly pluralist interpretation of that plurality is more faithful to the scientific situation. In contrast to more radical forms of pluralism, none affirms that nonconvergence is the rule across the sciences. The pluralism advocated is local, rather than universal. The contributors follow the advice from Dewey quoted by Richardson: to avoid being "false to the scientific spirit" by holding a priori to metaphysical doctrines. As Giere puts it, the case studies reject a priori commitments to either unity or multiplicity and allow the evidence and practical success (or failure) to decide.

Consequences of Assuming the Pluralist Stance

The basic point that scientific models generally obscure some aspects of complex phenomena in order to elucidate others has been increasingly accepted in philosophy of science, but the implications we draw from it have not. The implications contradict some deeply held views in the philosophy of science. These are the more difficult aspects of pluralism to accept, as our own occasional unthinking reversion to monist formulations attests. One implication of our pluralist outlook is that scientific approaches and theories should not be evaluated against the ideal of providing the single complete and comprehensive truth about a domain. This implication undermines a good deal of argumentation in the philosophy of science literature. For example, philosophers of biology have often argued that gene-centered explanations should be replaced by DST explanations on the grounds that gene-centered explanations leave out important causal factors. The underlying assumption is that any acceptable theory must include all the causal factors because the aim of science is to identify the single, comprehensive truth about development. Scientists and philosophers should recognize that different descriptions and different approaches are sometimes beneficial because some descriptions offer better accounts of some aspects of a complex situation and other descriptions provide better accounts of other aspects. And this may be the way it will always be.

The pluralist outlook suggests that there are serious limits for drawing metaphysical conclusions from science. While our empirically based pluralism is neutral with respect to realism in the sense that it does not require us to abandon realism, it does imply that realism needs to be tempered. Some philosophers and scientists argue that insofar as we seek answers to metaphysical questions, we should turn to the best contemporary scientific theory related to the question. While modest versions of this project might be sustained (perhaps certain metaphysics can be ruled out), the pluralist stance accepts that science has not and probably will not provide reliable answers to many of the big, interesting metaphysical questions. Is the world fundamentally deterministic? According to the Copenhagen interpretation of standard theory, it is not. According to the Bohm theory, it is. An empirical pluralism is open to the possibility that both accounts of quantum mechanics describe certain aspects of the phenomena well, and both could provide a basis for advancing inquiry. What is the level of selection? According to genic selectionists, it is always exerted at the level of individual genes. According to others, in some cases selection is exerted only at higher levels of organization, in other cases only at lower levels, and in still different cases at multiple levels. An empirical pluralism is open to the possibility that some aspects of a single case of natural selection might be best accounted for by modeling the process only at the genic level, and other aspects of the same selection process might be best accounted for by modeling the process at a higher level (or at multiple levels). If this is right, then science won't answer many metaphysical questions associated with scientific inquiry, such as questions about determinism or *the* level of selection or whether the world is such that a unified comprehensive account of it is possible.

The pluralist stance also has implications for philosophers who draw on philosophy of science to form conclusions about other areas of inquiry. For example, Bernard Williams, in *Ethics and the Limits of Philosophy,* writes, “In a scientific inquiry there should ideally be convergence on an answer” (1985, 136), and he takes this convergent monism in science as a sign of its objectivity, in sharp contrast to the community-bound nature of ethical discourse. Although we are pluralists, we do not assert plurality or the lack of convergence in the sciences. Instead, we deny the presumption of unity made by authors such as Williams. And in this denial, we leave open the possibility that ethical discourse can be as objective as scientific discourse.

Adopting a pluralist stance also has important consequences for the practice of philosophy of science. Pluralists might see the plurality in the local context of a scientific controversy as reflecting the complicated, multifaceted nature of the processes of interest. A monist will look at the same case of plurality and claim that the scientists in this local situation, as scientists in every local context, ought to be guided by the universal goal of uncovering the comprehensive account of the processes being investigated. Monism leads many philosophers to search for the concepts that will enable the pieces to fall into a single representational idiom. For example, philosophers were not content to identify a plurality of fitness concepts that could be drawn on to describe different aspects (or even different instances) of evolution. The explicit aim was to clarify *the* fundamental concept that underwrites all explanations invoking natural selection. The unspoken assumption was that there must be some underlying causal parameter, fitness, that would be the basic cause for all cases of natural selection. Pluralism denies this assumption. Or to be more precise, the pluralist stance refrains from adopting this tenet without empirical evidence. Pluralists do not assume that if we could just “get clear” on essential concepts, biologists could empirically determine how everything can be explained by a single account based on a few fundamental principles. By denying such assumptions, the pluralist stance requires us to revise the way we analyze concepts, both those of science and metascience.

Much of the analysis of concepts such as fitness in biology, function in psychology, and force in physics hinges on finding counterexamples against various proposed analyses. The unspoken assumption behind the method of counterexample is that there must be one kind of abstract thing that counts as fitness, function, or force. If one finds something that a proposed analysis can’t account for, then the analysis is taken to be refuted. To return to the fitness example, if a proposed analysis of fitness can't account for the long-term, as opposed to the short-term, evolution of a trait, then it is rejected. The idea behind such an argument is that the counterexample proves that the proposed analysis must not capture "the" concept of fitness because the right interpretation of fitness will be useful for understanding all important aspects of a complex evolutionary process. Scientific pluralism, however, acknowledges that different aspects of a sufficiently complex example of natural selection might be best accounted for by different models, which in turn might employ different concepts of fitness. It is not that any analysis of a term will do. But which analysis is best sometimes depends in part on what aspect of a complex situation is of greatest interest, and hence there might be more than one correct analysis. We believe that terms such as "chaos," "electron," and "function" exhibit the same polysemy.

Philosophers of science have also employed the method of counterexample in their analyses of metascientific concepts such as theory, explanation, cause, and probability. Does consistency require scientific pluralists to be pluralists about the analysis of these concepts? We think it does. Philosophers should not *assume* that the nature of science is such that it can be comprehensively accounted for by a single set of concepts that capture the fundamentals of science. This means, for instance, that the assumption that there is one abstract kind of thing, "scientific explanation," may be mistaken. Perhaps accounting for different aspects of scientific understanding will require different accounts of explanation. The monists' essentialism about metascientific concepts is unjustified. It follows that the familiar method of counterexample needs to be revised for our analysis of metascientific concepts as well as scientific concepts. Conceptual analyses ought to be evaluated on the basis of what they help us understand and investigate, not on the basis of whether they identify the single, essential way of understanding. One might extend this critique to analytic philosophy more generally and challenge the assumption that justice, knowledge, or consciousness must have uniform essential meanings that can be determined by the method of counterexample.

It should be evident that just as we take a pluralist stance on scientists' understanding of complex phenomena in the natural and social world, we also favor a similar stance on our own understanding of the multifaceted nature of scientific knowledge. This means that, like physicists trying to answer the most fundamental questions about the physical world, philosophers should acknowledge that there might not be answers to many of the most fundamental questions about science. Might the debate between Bayesians and their foes be futile, not simply because of lack of compelling evidence, but also because neither approach can (in principle) offer a comprehensive account of the basis of scientific inference?

Reflexivity also raises questions about the relation between philosophy of science and other areas of science studies. What does pluralism imply about the relation between the approaches of philosophers and those of historians, sociologists, and rhetoricians? As Kellert argues in his defense of the cross-training metaphor for interdisciplinarity, different perspectives on science, including the historical, normative-philosophical, and social-scientific, can shed light on different aspects of the multifaceted enterprise. Trying to force them into a convergent viewpoint or demanding a choice among them is counterproductive. Adopting a single approach would obscure certain aspects of science, perhaps limiting the advancement of that approach and certainly limiting our understanding of science as a complex phenomenon. As with our pluralism about science, we are not promoting an "anything goes" view. There are instances of poor research in every branch of science studies. Some of the most glaring examples of substandard work involve promoting a favored approach by trying to demolish what are viewed as opposing approaches for understanding science. These critiques are typically carried out within the perspective of the favored approach and assume the ideal of monism. That is, they assume that we should adopt just one approach, the one that promises to offer a complete account of the "essentials" of science. It is time to reject this ideal, for both science and the study of science. We should acknowledge that whether the world can (even in principle) be explained in terms of a single explanatory idiom or investigated by a single approach is an open question. We should adopt the pluralist stance.

References

Bauer, H. H. 1990. "Barriers against Interdisciplinarity: Implications for Studies of Science, Technology, and Society (STS)." *Science, Technology, and Human Values* 15: 105–19.

Barnes, E. C. 1998. "Probabilities and Epistemic Pluralism." *British Journal for the Philosophy of Science* 49, no. 1: 31–47.

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

———. 1999. *The Dappled World*. Cambridge: Cambridge University Press.

Cartwright, N., J. Cat, K. Fleck, and T. Uebel. 1996. *Otto Neurath: Philosophy between Science and Politics*. Cambridge: Cambridge University Press.

Cat, J. 2005. "Unity and Disunity of Science." In *Philosophy of Science: An Encyclopedia,* ed. Sahotra Sarkar and Jessica Pfeiffer. 2 vols. New York: Routledge.

Cat, J., N. Cartwright, and H. Chang. 1996. "Otto Neurath: Politics and the Unity of Science." In *The Disunity of Science,* ed. Peter Galison and David J. Stump, 347–69. Stanford, Calif.: Stanford University Press.

———. 2005. "Why Genic and Multilevel Selection Theories Are Here to Stay." *Philosophy of Science* 72: 311–33.

Williams, B. 1985. *Ethics and the Limits of Philosophy.* Cambridge, Mass.: Harvard University Press.

Wylie, A. 1995. "Discourse, Practice, Context: From HPS to Interdisciplinary Science Studies." In *PSA 1994: Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association,* vol. 2, ed. David Hull, Micky Forbes, and Richard M. Burian, 393–95. East Lansing, Mich.: Philosophy of Science Association.

Churchland, P. 1985. "The Ontological Status of Observables: In Praise of the Superempirical Virtues." In *Images of Science,* ed. P. Churchland and Clifford Hooker, 35–47. Chicago: University of Chicago Press.
Cushing, J. 1994. *Quantum Mechanics.* Chicago: University of Chicago Press.
Dawkins, R. 1982. *The Extended Phenotype.* Oxford: Oxford University Press.
Dupré, J. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science.* Cambridge, Mass.: Harvard University Press.
Ereshefsky, M. 1998. "Species Pluralism and Anti-Realism." *Philosophy of Science* 65: 103–20.
———. 2001. *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy.* Cambridge: Cambridge University Press.
Giere, R. 1999. *Science without Laws.* Chicago: University of Chicago Press.
Grantham, T. 1999. "Explanatory Pluralism in Paleobiology." *Philosophy of Science* (Proceedings) 66: S223–36.
———. 2004. "Conceptualizing the (Dis)unity of Science." *Philosophy of Science* 71: 133–55.
Kitcher, P. 2001. *Science, Truth, and Democracy.* Oxford: Oxford University Press.
Kuhn, T. 1977. *The Essential Tension.* Chicago: University of Chicago Press.
Longino, H. 1996. "Cognitive and Non-Cognitive Values in Science." In *Feminism, Science, and the Philosophy of Science,* ed. L. H. Nelson and J. Nelson, 39–58. Dordrecht: Kluwer.
———. 2001. "What Do We Measure When We Measure Aggression?*" Studies in History and Philosophy of Science* 32, no. 4: 685–704.
———. 2002. *The Fate of Knowledge.* Princeton, N.J.: Princeton University Press.
McMullin, E. 1983. "Values in Science." *PSA 1982: Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association,* vol. 2, ed. P. D. Asquith and T. Nickles. East Lansing, Mich.: Philosophy of Science Association.
Mishler, B., and R. Brandon. 1987. "Individuality, Pluralism, and the Phylogenetic Species Concept." *Biology and Philosophy* 2: 397–414.
Mishler, B., and M. J. Donoghue. 1982. "Species Concepts: A Case for Pluralism." *Systematic Zoology* 31: 491–503.
Mitchell, S. 2002. "Integrative Pluralism." *Biology and Philosophy* 17: 55–70.
———. 2003. *Biological Complexity and Integrative Pluralism.* Cambridge: Cambridge University Press.
Morrison, M. 2001. "History and Metaphysics: On the Reality of Spin." In *Histories of the Electron: The Birth of Microphysics,* ed. Jed Z. Buchwald and Andrew Warwick. Cambridge, Mass.: MIT Press.
Stanford, K. 1995. "For Pluralism and against Realism about Species." *Philosophy of Science* 62: 70–91.
Stump, D. 1992. "Naturalized Philosophy of Science with a Plurality of Methods." *Philosophy of Science* 59: 456–60.
Suppes, P. 1978. "The Plurality of Science." In *PSA 1978: Proceedings of the 1978 Biennial Meeting of the Philosophy of Science Association,* vol. 2, ed. Peter Asquith and Ian Hacking, 3–16. East Lansing, Mich.: Philosophy of Science Association.
Teller, P. 2004. "How We Dapple the World." *Philosophy of Science* 71, no. 4: 425–47.
van Bouwel, J. 2004. "Explanatory Pluralism in Economics: Against the Mainstream?" *Philosophical Exploration* 7, no. 3: 299–315.
van Fraassen, B. C. 1989. *Laws and Symmetry.* New York: Oxford University Press.
Waters, C. 1991. "Tempered Realism about the Force of Selection." *Philosophy of Science* 58: 553–73.

Alan W. Richardson

1 The Many Unities of Science: Politics, Semantics, and Ontology

> One may ask: "What program is common to all the collaborators of the *Encyclopedia*?" A program formed of statements accepted by all the collaborators would be narrow and would be a source of divergences in the near future. . . . The maximum of co-operation—that is the program! (Neurath 1938, 23–24)

> Given the cultural resonance of disunity and unity, it is perhaps no surprise that the various authors here, despite their disagreements, explore not only the disunity of the scientific context but also, implicitly, the context of disunity. (Galison 1996, 33)

Within philosophy of science in the 1990s and into the new century, the themes of disunity of science and of pluralism of or regarding science have been prominent. Work such as John Dupré's *The Disorder of Things* (1993), Nancy Cartwright's *The Dappled World* (1999), and the essays collected in Peter Galison and David Stump's *The Disunity of Science* (1996) make this theme evident in their titles. Other works, such as Helen Longino's *The Fate of Knowledge* (2002), endorse or make room for pluralisms of various kinds without wearing such allegiance on their dustcovers. Such work often draws on (or can be interestingly related to) work in other science and technology studies disciplines such as Donna Haraway's "situated knowledges" (1991), Bruno Latour's recent move to "multinaturalism" (1999), or a whole raft of local, contextual histories of science, some of which draw their antiuniversalist consequences quite explicitly (see, for example, the final chapter of Galison 1987).

Such work has, of course, drawn its critics and thus enabled the next generation of unificationist literature in its wake. Speaking from the very pages of Galison and Stump's disunity manifesto, Richard Creath (1996, 158) has argued that "what our Viennese predecessors were really defending was really more sane and sensible than has been supposed" and that much recent work under the banner of "disunity of science"—and especially Galison's own work—"is in harmony with that classical tradition." More recently, Margaret Morrison (2000, p. 1) has argued:

> Critics of unity claim that when we look at scientific practice we see overwhelming evidence for disunity, rather than the coherent structure we have been led to believe characterizes science. Although some of these arguments are extremely persuasive, the desire to banish unity altogether has resulted, I believe, in a distortion of the facts and a misunderstanding of how unity actually functions in science. It is simply a mistake to deny that science has produced unified theories. So, where does the evidence of disunity come from?[1]

Standing at a more magisterial distance from the disunity debates, Michael Friedman (2001) nonetheless endorses such a strong role for a quasi-Kantian "regulative a priori" that one can confidently draw the conclusion no disunified knowledge system will be able to fulfill the regulative ideal of scientific knowledge.

Now, at the very same time that contemporary philosophers of science have been wrangling over disunity of science, there has been a rekindling of interest in precisely those Viennese predecessors that Creath referred us to above. The history of logical empiricism has become a going concern in the philosophical world.[2] Moreover, since the logical empiricist "unity of science" movement forms the immediate locus of dispute for the new generation of disunity of science scholarship, these areas of interest overlap. Indeed, quite often work in disunity of science proceeds in self-conscious awareness of and respect for the unity of science movement. This is especially true in the work of Peter Galison (1990, 1996), who is the historian of science who has made the most frequent and most valuable interventions in the literature on the history of logical empiricism, and, of course, of Nancy Cartwright, whose scholarship on Otto Neurath finds the main mover behind *The International Encyclopedia of Unified Science* (Neurath, Carnap, and Morris 1938) to have been a great advocate of disunity of science (Cartwright et al. 1996).

There is something philosophically troublesome about this rich stew of current debates and historical antecedents. There is, throughout the disunity literature, a nagging sense that there is a mixing together of issues and concerns. If analytic philosophy has any mortal enemy, it is stews, potages, and other unclear mixtures. Thus, the literature on disunity of science has been regimented on several occasions, perhaps most notably in chapter 1, "The Many Faces of Unity," of Morrison's book (2000) and in Ian Hacking's essay, "The Disunities of the Sciences" (1996). Interestingly, these attempts to induce a bit of conceptual order have a historical bent, exploring the issues through exemplary positions taken—though neither of the authors limit themselves to the logical empiricists. Nonetheless, the logical empiricist unity of science movement finds prominence in these accounts.

This attention to relations to the logical empiricist unity of science movement is right and proper. There can be no question but that disunity and pluralism are key topics for current philosophy of science in large part because current philosophy of science is still working out the ways in which it is not logical empiricism anymore. This was made rather explicit in the letter of instruction to participants in the Minnesota Center for Philosophy of Science "Workshop on Scientific Pluralism" in 2002 that led to this volume. Indeed, here is the account of the goal of the workshop as presented in the letter: "Our goal is to determine whether a consistent, substantive, and philosophically defensible view of scientific pluralism can be developed that goes beyond the mere rejection of the unity of science doctrine."

This essay belongs in the same genre as Morrison's and Hacking's essays. Like them, I shall attempt to clarify issues in disunity or pluralism of science through a bit of attention to history; like them I shall attend, if only briefly, to several different figures in different historical time periods. The logical empiricists shall dominate my discussion, however. This dominance is an expression both of my own limitations as a scholar and of my sense that there are still underdeveloped themes from within the unity of science movement that can help us sort out what is interesting and important in the current disunity literature.

The Unity of Science Movement and Current Vocabulary: Some Preliminary Complications

Here, again, is the goal the workshop was set: "Our goal is to determine whether a consistent, substantive, and philosophically defensible view of scientific pluralism can be developed that goes beyond the mere rejection of the unity of science doctrine." For those of us steeped in the history of philosophy of science, this sentence presents a problem that ultimately becomes a resource and an opportunity. After all, when you have read the *International Encyclopedia of Unified Science* from cover to cover, there is perhaps one theme in it that is more striking than any other: there is no unity of science doctrine. I mean this in two senses. First, it is quite clear that the unities of science either defended, advocated, or proposed by the various authors in the *Encyclopedia* are quite as numerous as they are. But, second, it is not simply that the *Encyclopedia* presents no unified understanding of unity. No, the lack of doctrinal unanimity on just this point was often referred to with pride as one of the principal virtues of the project: the unity of science was a thematic and procedural commitment but did not entail any substantive agreement on philosophical doctrine. Thus, my motto from Otto Neurath's introductory essay to the *Encyclopedia* and thus also John Dewey:

> It follows that a movement in behalf of the unity of science need not and should not lay down in advance a platform to be accepted. It is essentially a co-operative movement, so that detailed specific common standpoints and ideas must emerge out of the processes of co-operation. To try to formulate them in advance and insist upon their acceptance by all is both to obstruct co-operation and to be false to the scientific spirit. (Dewey 1938, 33–34)

The lack of unanimity regarding unity of science in logical empiricism and its fellow travelers is, when rightly understood, a resource for advocates of pluralism. If pluralism were merely the negation of a single philosophical doctrine that turned out to be false or barren, pluralism would have all the virtues but also all the excitement of honest intellectual toil. What makes pluralism promising is that it is more than a negative doctrine accepted, either in disappointment or in relief, in the wake of a failed doctrine. It has the excitement not of honest toil or yet of intellectual theft but of a forensic and deliberative project that seeks to understand what was right and what was wrong about a complicated philosophical project that went unfulfilled—pluralism and disunity are themes in the complex working-through of what was wrong and what was right in logical empiricism. In any case, this is what I will argue by giving attention to some contending doctrines of the unity of science and to the philosophical motivations that underpinned what advocates preferred to call not "the unity of science doctrine" but rather "the unity of science movement."

Before attending to the main business, allow me to make some points, as necessary as they are pedantic, about names of philosophical projects. Let us think a bit about "unity of science"—the words, as keys to the thing. Pluralism—about scientific theories, representations, cultures, objects—seems badly defined as "plurality of science." A doctrine of the plurality of science sounds like a doctrine that says that there is more than one science. That doctrine, however, is not honest toil; it is not even theft; it is common property. It is not news that there is more than one science, and one does not need philosophy to arrive at this firm truth. (Perhaps only a philosopher would undertake to deny it.) The unity of science movement surely did not deny that there are many sciences—indeed, most of the articles in the *Encyclopedia* were on individual sciences. What lent to the unity of science project a real intellectual bite was the claim that this historical distinctness among the sciences did not express an important underlying epistemic or methodological disunity—that it sought to lay bare a unity that might have been missed. The question before pluralists is whether there is a *philosophically important* lack of unity that has somehow been missed, one that is not identical with the lack of unity that everyone already knows about.[3]

Before leaving this train of thought, we might attend also to the phrase "unified science," since "unity of science" was the goal for which "unification" was the means. The unity of science movement, in laying out and promoting a unity of science that might otherwise be missed or even not come into existence, was seeking to foster something—a unifying and cooperating tendency among the sciences and between science and philosophy, a science made more unified. One thing to attend to in reading current pluralists is whether they are advocating a "plurified" or "pluralized" science and if so, on what grounds. One could, for example, imagine seeking less cooperation between evolutionary theory and anthropology, on the belief that we have had enough sociobiology already and it is time to do other things that are more explanatorily promising. But in such a case, is one's sense of the need for new things in anthropology due to a commitment to "pluralized" science or to a commitment to better science? This is where the anxieties about negation are most well placed, I believe. In good scholastic vocabulary, one might worry that pluralism is genuinely a negation or a privation since *advocating* a lack of cross-disciplinary unity or cooperation as a way forward methodologically seems rather peculiar advice. "Anthropologists, do not listen to biologists!" seems an injunction in need of a further reason. Moreover, it seems to need completion with a more substantive reason than "Science is plural!"

The example, however, is useful for uncovering a problem that many have sensed within certain scientific projects and that motivates at least some pluralists.[4] Many believe that a Dennett-style Darwinism is not a *cooperative* enterprise in the realm of human psychology or anthropology; evolutionary theory in those domains is an imperialist doctrine. Here we enter ground well trod already in the 1970s when disunity of science was the call of those nascent cognitive scientists such as Fodor (1975) who wanted to argue for the predictive and explanatory autonomy of the special sciences. Of course, the flip side of imperialism is co-optation. Some biologists worry that anthropologists and psychologists move from their felt need for evolutionary explanations to the claim that they have such explanations too quickly. The response by biologists such as Lewontin (1991) to projects like "evolutionary psychology" is less the distaste of an imperialist looking on awkward aboriginal attempts to become civilized and more the horror with which, for example, inner-city black rappers greet the white, suburban co-optation of their music.

Early Logical Empiricism and a Rickertian Form of Pluralism

Let us leave that where it is for the moment and return to the main business of this essay, which is teasing out pluralisms worth having by thinking about various unities heretofore proposed. Let us begin where many have located the point of the unity of science project: in ontology. While this is not the best reading of the mature project (see below), it is certainly the case that early on the unity of science was often presented in something like ontological terms. In the *Aufbau* of 1928, Rudolf Carnap ([1928] 1967, 29) claims to establish the unity of "the object domain of science." He does this by presenting a language in which all significant scientific discourse can be formulated. Putative metaphysical things such as essences, however, cannot be constructed—that is, they cannot be defined in the language—and this is the fact that Carnap uses to expunge metaphysical talk. Metaphysics does not speak of things in the object domain of science; there is only one such domain, and it contains all the objects that can be referred to, so metaphysics strictly does not speak of anything at all.

I have been deliberately sloppy about ontological and semantic issues in what has just been said, precisely because Carnap is sloppy in the same way. This points to his own subsequent move to semantics, but in part it is because even in 1928 the point of insisting on the unity of the object domain of science was not exactly ontological. That is, the point was not to deny that science investigates many different things, indeed, many different kinds of things; Carnap would not deny that rabbits are different from electrons even though both are studied by science. Rather, the point of the unity of science for Carnap in 1928 was to take sides with some neo-Kantians against other neo-Kantians in a dispute that mixed together ontological, logical, and methodological issues. This is, of course, the debate over the difference in kind that some claimed obtained between the *Geistes*- and the *Naturwissenschaften*.

Heinrich Rickert ([1929] 1986) and others in the historically minded Southwest school of neo-Kantianism had claimed that the cultural and the natural sciences were different in principle because the former were, but the latter were not, interested in knowing and understanding the individual as an individual. That is, Rickert claimed that the mathematical form of natural scientific knowledge meant that the concepts of natural science could not uniquely specify individual objects and their activities, but the cultural sciences would lose their point if one could not speak of individual actions of individual agents—no historian is interested in Caesar's crossing of the Rubicon as an instance of a general law of events of such and such an abstract type attending on circumstances similarly abstractly and generically formulated; historical events are historically interesting precisely because they are uniquely meaningful. (This is not to deny that there are many law-governed actions of human beings; it is to deny that those are the actions that are historically interesting.) The two types of science, then, differ in relation to individual things, according to Rickert. Also, they differ according to their need to attribute action based on intentions or reasons. Rickert and others concluded from this that the natural sciences had a different method of knowing than did the cultural sciences. Rather than trying to find the proper universal law under which to subsume Caesar crossing the Rubicon, a proper explanation in history would speak to the significance for Caesar of this action. It would posit plausible beliefs, reasons, and intentions that would make sense of his action in its context. The whole epistemic structures of natural and cultural sciences were different; there was no one way to know everything worth knowing about the world.

It is this joint ontological and methodological distinctness of natural and cultural science that Carnap, in this case quite explicitly following Ernst Cassirer ([1910] 1953), sought to reject by stressing the unity of the object domain.[5] Cassirer and Carnap argued that Rickert had misunderstood the role of mathematical form in natural science, had misunderstood the formal elements as universals rather than as relational forms constitutive of individual content. Not only does formalism not deny individuality, but the individual can only be objectively specified via the formal conditions that uniquely determine it.[6]

We need not go into details here. The point in the current context is that by supporting Rickert against Carnap, we might find a pluralism worth having: Rather than insisting on there being a variety of kinds of things in the world—something not many would deny—the pluralist could go as far as Rickert, even if not doing so to enforce exactly the distinction Rickert himself wanted. That is, the pluralist could insist that there are different kinds *that can be known only in different ways.* We all know that people are not quarks, but perhaps the proper ways of coming to know things about people are not the proper ways of coming to know things about quarks. (Perhaps this is true of zebras and quarks or toasters and humans also.)

Now, I do not have an argument that this form of pluralism is true, or if it is true, into how many epistemologically distinct categories kinds would fall. And of course, one would like there to be some compelling reason why the knowing relation is different for kind x and kind y. But we do seem to have a genus of useful pluralisms; one that has various different varieties. For example, Hacking's (1996, 2001) plurality of styles of reasoning seems, insofar as he does not advocate for the kinds *within* any style of reasoning, to be a way of indicating that what in the world we can know depends on what sort of knowing we bring to the world. Since, on Hacking's view, styles of reasoning can and do coexist, he has a robust pluralism—not the pluralism of the historian who can say "those were the kinds available given the style of 1650; these are our kinds," but the pluralism of a theorist who can say, "We have experimental, algorithmic, statistical, . . . kinds in our world."

Of course, if we push further along in this direction and are more uneasy with relativism than Hacking seems to be, we arrive at a place where an internal ontology within any style of reasoning is our only place to speak of ontology at all. We arrive safely, in this way, at a form of pluralism that was enunciated within the unity of science movement—Carnap's plurality of linguistic frameworks. Of course, if our goal is to enunciate a form of pluralism, we might resist making the final Carnapian move here: we might wish to insist that there really are things knowable in only one way and other things knowable in only some other way. Such talk of the really real transcending the frameworks themselves is exactly the sort of talk Carnap eschewed—his pluralism of frameworks has no metaphysical lessons to teach. So, if we want a metaphysical pluralism, we cannot be Carnapians.

There is another reason why a pluralist might resist Carnap's linguistic frameworks (the case of Hacking's styles is more complicated and troubling). This reason is more practical than metaphysical. One might wish to resist the rather totalizing element found in Carnap's project. It is impossible actually to adopt and use two Carnapian languages at once; one's constitutive a priori principles cannot be put on and taken off as needed. (This is the complication in Hacking's styles: they coexist in the sense that we as a culture and we each individually may use more than one of them, but can we actually use more than one of them in any one bit of reasoning?) The point is perhaps best made in Carnapian language: Carnap was happy to say that the choice of linguistic framework was a pragmatic choice and one that was revisable, but for him it was revisable in toto or not at all (that's how semantic holism works). A pluralist might hanker after a more practical point of view, suggesting that Carnap's practical decisions were rather like a carpenter being asked to choose exactly one tool at the start of a job and to stick with it. Pluralist carpenters would rather have saws, hammers, screwdrivers, and the rest, as needed—and sometimes one needs to use both a screwdriver and pliers for one and the same job.

Social and Metaphysical Disorder and Philosophical Anxiety

There are worries that philosophers are prone to when faced with this sort of epistemic eclecticism. The standard worry is that such eclecticism is ultimately a pernicious form of relativism. I think that relativism is the current successor to skepticism in the realm of what C. S. Peirce ([1885] 1992) very properly called philosophical hypochondria. Philosophers have theorized knowledge in such a way that significant shifts in belief or in representational technology lead, *in theory,* to a sort of anxious nightmare of people wholly incomprehensible to one another. But this is not an epistemic phenomenon we are presented with except in the rarest of circumstances—incomprehensibility is (almost) always local and (almost) always subject to diagnosis even if not to reasoned resolution.[7]

Suppose "pluralism" names this position: (1) the theoretical forms of representation used in the various sciences vary in ways that make those theories hard to combine; (2) these forms of representation interact in interestingly different ways with the activities of scientists in the lab and the field, where scientists often use other forms of instrumental representation; (3) these facts of daily scientific practice may have epistemological importance and may require new philosophical resources to be properly understood; (4) let's investigate this and seek those new resources. Such a pluralism is important for us as philosophers of science here and now. For an investigation to be shunted aside because there is a three-move technique in the philosopher's game that reduces it to relativism would be a terrible shame. After all, philosophy of science began not with suspicion about science, but with suspicion that proper attention to science would reveal the poverty of the philosopher's game.

It is not as if philosophers' current attention to the variety of representational resources, practices, and technologies in science is "the latest finding" of philosophy of science, in the sense that philosophers now know something about science that no one ever knew before. The importance of this variety is not as a piece of new information but as a problem for philosophy: if we take this variety of representational resources, practices, and technologies seriously, what has to change in our theories of scientific knowledge making? This is a problem for us particularly because we have our own reasons for thinking that the well-honed tools of philosophy of science, such as formal logic, are not sufficient for solving it.

My "pluralism in science as a philosophical problem" does not take sides on what the lessons of pluralism are or whether those lessons are in fact going to turn out to be epistemologically interesting. I am willing to believe that Morrison may be right—whatever the importance of pluralism, it does not preclude the importance of unity—or even that Creath may be right—there is nothing particularly new here at all. Unlike the suggestion in Morrison's work, however, I am not taking the variety of representational technologies, practices, and so on as *evidence* of pluralism; I am taking them to *be* pluralism. They are part of the phenomena of science-as-practiced. And, unlike the suggestion in Creath's work, I am willing to entertain the idea that the logical empiricists both noticed representational and methodological diversity in the sciences and radically underestimated the importance of it.

Helen Longino has raised a different sort of worry about the eclecticism I have been developing: it is not strong enough to ward off what she takes to be the proper negation of pluralism, which is a position she calls "monism." For Longino, monism is the view that "for every natural process there is one and only one correct account of it" (2002, 44). Rickert's version of pluralism, for Longino, is not a version of pluralism at all—since Rickert is committed to there being a uniquely correct account of the expansion of copper when heated and of Caesar crossing the Rubicon; it is just that those accounts will be very different.

Unlike Longino, I find it hard to get very exercised about monism. I suspect that whether one finds Longino's monism problematic reveals something about deep philosophical commitments, commitments that peel the unity issue back well beyond logical empiricism to a previous philosophical era and a more general philosophical issue. Suppose we notice that scientists use a variety of representational technologies—differential equations, computer simulations, mechanical models, visual models, whatever. Monism is the view that for each (sort of) thing, there is some representational technology and some representation in that technology that gives the single best account of that thing. Well, there are philosophers who think that that is pretty well necessary because they think that the ability of symbolic systems of any sort to represent objects at all can only be explained by virtue of a match between some aspect of the symbol system and the nature of the thing being represented. This is representational realism—the view that representation is possible at all only by a match of some sort between the real nature of the object and the representational system itself.

This representation realism is a form of transcendental realism of exactly the sort that Kant argued against. There are three salient ways of resisting transcendental realism. The first way is Kant's way: any empirically significant talk of objects takes place within a representational system, the ability of which to represent cannot be explained by reference to something wholly external (such as *Dinge an sich*) to the representational system. (Kant had no pluralist concerns, of course, because he had no plurality of representational systems.) The second way is Hegel's way: the object is the never given but always pursued endpoint of the dialectic of the systems of representation themselves.[8] The third way is Bergson's way: the object has a fulsome being that always outruns the ability of any symbol system to represent it.

Kant, Hegel, and Bergson might seem to be rather lofty figures to bring to bear on homey facts such as the variety of techniques used to represent brain activity in psychology and medical science or the variety of formalisms in quantum physics. But high philosophy is meant to explain something about our cognitive lives, so we shall not allow our philosophers to hide in their studies. A multiplicity of contemporaneous representational technologies does lend a certain poignancy to disputes between Kant, Hegel, and Bergson. On the face of things, Hegelian absolute idealism seems to be worst off. At the very least, the logic of dialectic involves a succession of representations that seems ill-suited to illuminating a situation in which the representational systems coexist. Longino seems tempted by something like Bergson when she writes:

> Philosophers who advocate strong forms of pluralism, however, are claiming that the complexity of natural entities and processes (either all such or just organic entities or processes) eludes complete representation by any single theoretical or investigative approach. Any given approach will be partial and completeness will be achieved not by a single integrated theory but by a plurality of approaches that are partially overlapping, partially autonomous, and resisting reconciliation. (2002, 93)

Here, the order of explanation goes from the ontological thickness of objects, through the representational thinness of any particular representational idiom, to the need for a plurality of representational idioms.[9]

We could take a page from the neo-Kantians, however, and explain the plurality of representational idioms not by reference to what eludes them but by virtue of the coordination of the representational idioms themselves. On this view, the object "brain activity" is constituted via the coordination of the various representational idioms. The need for the variety comes not from what goes beyond all of them but from the fact that exactly this object is uniquely picked out only by reference to all of them. This neo-Kantian position might not seem to explain why there must be a plurality of representational systems. But then we might ask whether "because the object is so complex" is really an answer to that question, especially when we note that not only our evidence of the complexity of the object but also what we mean by the complexity of the object are that we do have various representations of the object. This is to say that the multiplicity of representational systems is not metaphysical necessity; the necessity is an internal one: given that we have all these representational systems, any given object is picked out by coordinating among them.[10]

Carnap's Unity of Science Circa 1938

In tracing through some ways of continuing on from negating "the unity of the object domain" of science, we have—in our discussion of a variety of representational idioms or technologies in science—actually arrived at a negation of the most prominent account of unity of science in the *Encyclopedia*: the unity of science is the unity of the language of science.[11] This is a view we can associate with Carnap, for whom the question of the unity of science could not be an ontological question (there being no such questions) but only a question of the logic of scientific language. Now, Carnap's actual view of the unity of the language of science is rather subtle. The briefest statement of it is this: "the class of observable thing-predicates is a sufficient reduction basis for the whole of the language of science, including the cognitive part of everyday language" (Carnap 1938, 60). This view is what Carnap also called "physicalism."

Among the subtleties of the position is the fact that "observable thing-predicates" are discovered psychologically but do not refer to private experience. Thus, observable thing-predicates are predicates like "hot" and "heavy" and "blue"—predicates not of experiences but of things, and predicates that can be taught pretty well in isolation from other language. Thus, we can teach someone to assent to or dissent from "hot" as a predicate of things in a way that does not rely on many judgments and inferences using other parts of the language. Of more interest to us, however, is the fact that "sufficient reduction base" does not for Carnap in 1938 mean anything like "conceptual basis adequate for translational reduction." Carnap does not think that the observable thing language is conceptually rich enough to express what *any* part of science actually says; it simply supplies the terms in which scientific terms are introduced via reduction conditionals. Thus, one might begin to introduce the psychological concept of pain with a reduction conditional like "If you poke someone in the eye with a red-hot poker, then if they scream and writhe, they are in pain," where, ideally, the terms of the antecedents (or consequents) that do not involve the introduced term are in the observable thing language.

The unity of the scientific language for Carnap was, even in 1938, very loose. It involved no definability of terms and no derivability of laws. Nor, despite the polemical terms of Quine's attack on Carnap, did it involve any fictitious history of how the terms were actually introduced into scientific language. Carnap was not in the business of telling just-so stories about how scientists introduced "temperature" or "pressure" into their languages; he was trying to enforce a forward-looking criterion of a logically adequate introduction of concepts into science. We assume that the references of behaviorally accessible observational thing-predicates are fixed, and then we introduce new predicates in ways short of translation but which tie the application of the new predicates to states of affairs ascertainable wholly in the language of the observable thing-predicates we have in hand.[12] Perhaps the place where Carnapian strictures on meaningfulness would be best exemplified would be in the *Diagnostic and Statistical Manual* if the diagnostic criteria were explicitly understood as meaning postulates. (If such were the clear role of the *DSM,* we could have a much clearer sense of whether the recent explosion in cases of autism, for example, is a semantic or a pathological phenomenon.)

Rather than continue to expound or rehearse objections to Carnap's unity of the language of science doctrine, I would prefer to focus our attention on the point of the doctrine. What problem was the unity of the language of science meant to solve? Within Carnap's work, the philosophical problem is exemplified in this statement about behaviorism in psychology (Carnap 1938, 58): "Hence there cannot be a term in the psychological language, taken as an intersubjective language for mutual communication, which designates a kind of state or event without any behavioristic symptom." Carnap's concern was ever directed toward the intersubjectivity of science: intersubjective agreement with respect to the observable thing-predicates was a condition for knowing what those predicates meant (because agreement with the community's judgment is the criterion for having learned the predicates), and the reduction conditionals give intersubjectively valid test conditions under which the introduced predicates are applied or denied.[13]

Subjectivity per se is not much of a problem—my preference for Bosc over Anjou pears should not keep anyone up at night, even if they have the opposite preference. I want to suggest that the cool, bloodless style of Carnap's writing rather masks the key concern that leads to an insistence on intersubjectively available test conditions in the language of everyday life within which proper introduction of the scientific concepts occurs. In order to unmask this concern, I will attempt to be more warm-blooded myself and avail myself of more warm-blooded fellow travelers of this part of Carnap, Neurath, and Dewey.

What was the real concern with metaphysics that motivated the logical empiricists? It was not that metaphysicians are idle fools, telling stories full of sound and fury but signifying nothing. It was, rather, that these stories, while signifying nothing, were not idle and not understood to be foolish. Metaphysics served ideological purposes—long, detailed theories of value that no one could understand were nonetheless held to be importantly true and served to prop up improper systems of authority. When my preference for Bosc pears comes to be grounded in a long metaphysical story that announces the superiority of Bosc pears but cannot be understood or evaluated by any who do not agree with it, something epistemically suspicious is going on. Our suspicions should only increase when such stories serve to support great disparities of wealth or cover over the misery of much human life. The concern with metaphysics and with promotion of unity of science was an effort in the 1920s and 1930s to understand and to solve a very material and social problem of ideology: what to do with communities of alleged experts offering theories that were effectively immunized from internal or external criticism but which proffered explanations of social, political, or economic conditions?

The logical empiricists theorized metaphysics not as hapless foolishness but as obfuscation in support of authority and power. The claim, therefore, that proper scientific status involves intersubjectivity and that this, in turn, means explicit connection to the common language *we all* speak and to explicit test conditions specifiable in that language makes sense in this context. If scientific language works like that or can be regimented to work like that, it exhibits the virtues of transparency and intersubjective control that metaphysics and ideology importantly lack. Moreover, if all of scientific language is regimented in this way, it is not only connected internally, one science to another, but also connected to the lifeworld and sphere of activity of the ordinary person. The unity of science shows how science and scientific philosophy do, in fact, "serve life" and why life would "receive it" as Neurath, Carnap, and Hans Hahn famously said in 1929 ([1929] 1973). This is why Cartwright et al. (1996) insist on stressing Neurath's concern with unity "on an earthly plane" and "at the point of action."

This last point is worth emphasizing, since it is the main theme of two of the six introductory essays in the *Encyclopedia,* the essays by Neurath and Dewey, while also being mentioned with some prominence in the introductory essays by Carnap and Charles Morris. Here is Neurath, as he reaches the crescendo of his essay:

> The empiricalization of daily life is increasing in all countries. . . . A meteorologist trained in Denmark may become a useful collaborator to a Canadian polar expedition; English economists can discuss a Russian analysis of American business cycles; and Russian economists may object to or accept the opinions of English economists about the effect of rural collectivization in the Soviet Union. . . . One can state . . . scientific prognostications in terms of everyday language. . . . Unified science is therefore supported, in general, by the scientific attitude which is based on the internationality of the use of the language of everyday life and on the internationality of the use of scientific language. (1938, 22–23)

Here is Dewey:

> There is also a human, a cultural, meaning of the unity of science. There is, for instance, the question of unifying the efforts of all those who exercise in their own affairs the scientific method so that these efforts may gain the force which comes from united effort. Even when an individual is or tries to be intelligent in the conduct of his own life-affairs, his efforts are hampered, often times defeated, by obstructions due not merely to ignorance but to active opposition to the scientific attitude on the part of those influenced by prejudice, dogma, class interest, external authority, nationalistic and racial sentiment, and similar powerful agencies. Viewed in this light, the problem of the unity of science constitutes a fundamentally important social problem. (1938, 32–33)

For good measure, here is the very end of Carnap's essay:

> For very many decisions, both in individual and in social life, we need . . . a prediction based on the combined knowledge of concrete facts and general laws belonging to different branches of science. If now the terms of different branches had no logical relation between one another, such as is supplied by the homogeneous reduction basis, but were of fundamentally different character, as some philosophers believe, then it would not be possible to connect singular statements and laws of different fields in such a way as to derive predictions from them. Therefore, the unity of the language of science is the basis for the practical application of theoretical knowledge. (1938, 62)

There are characteristic differences in these three passages, but the leading ideas are clear enough. The unity of science unites all the sciences together so that they can yield predictions and support action in the world of everyday life. The institution of science and of scientifically minded people has fostered an activist international community of people united linguistically and in their open and critical habits of mind. Opposed to this trend in science and culture is a trend toward divisive, obscurantist, often nationalistic metaphysical ideologies that condemn human life to arbitrary authority, dogma, and ignorance. Philosophy of science could do good work by fostering the former and fighting against the latter. This required a theoretical understanding of what the unity of science consisted of; Carnap's account was one such theoretical understanding, based on the methods of logical analysis so useful in foundations of mathematics and elsewhere. This is the philosophy of science promoted by exiled Austrians and Germans and progressivist Americans circa 1938.[14]

Projects in a Pluralism of Science Movement

Let us place pluralism in this context for a moment. Let us fully pry apart the motivating pretheoretical claim that science is an internationalist and practical discipline promoting positive change in the world and, as such, is opposed to fragmenting and dangerous stories in ideology and metaphysics on the one hand, and, on the other, the particular tools the logical empiricists circa 1938 had for theorizing this claim.

Most of what we as philosophers of science have been taught has been on the latter, more internal and theoretical, side of the question. Pluralism, indeed, received one of its prime impetuses from work by Thomas Kuhn and others pointing to the variety of representational schemes and media and the variety of persuasive and pedagogical techniques actually employed in science. I think the proper attitude toward such work is not that it "refuted" logical empiricism as much as it seriously undermined the theoretical resources and methods for discussing science employed by the logical empiricists. Here, as usual, Carl Hempel seems to have gotten the tone exactly right in his 1973 essay critiquing the "standard empiricist construal" of the meaning of theoretical terms:

> It might seem that a theory lacking an interpretation in terms that are clearly understood must be deemed, by strict analytic standards, not to be objectively intelligible, to lack objective significance. But the standard here invoked, which has been very influential in logical empiricism and, indeed, in much of analytic philosophy, is much too restrictive. New concepts can become intelligible, the use of new expressions can be learned, by means other than explicit linguistic interpretation; and, as the history of scientific theorizing illustrates, the new linguistic apparatus thus introduced can come to be employed with high interpersonal agreement. (2001, 216)

Hempel concludes from this that the entire demand for an account of meaning of theoretical terms "rests on a mistaken presupposition and thus requires no solution."

Since the use of meaning postulates and bridge principles to interpret theoretical terms was a central aspect of the semantic unity of science project, Hempel's remarks bear some reflection. One can read him as saying that the problem of the unity of science becomes a problem of the specification of the meaning of theoretical terms, and this problem is to be solved through logical techniques of definition, meaning postulates, reduction conditionals, only after we have, as philosophers of science, already understood a logic-based semantics to be the proper tool for formulating our problems. Hempel (1993) thought Kuhn and others had decisively shown this to be the fundamental mistake of logical empiricist philosophy of science: meaningfulness in science was a natural phenomenon regardless of the difficulties logic had in helping us to understand it.

This sort of Kuhnian descriptive pluralism of how scientists actually represent the world and how they teach and transmit their representations may loosen the hold on the philosophical imagination of the explicit tools and resources of logical empiricism. It does not, of course, yet speak to the issue of whether the motivational phenomenon Neurath and Carnap wished to theorize actually obtains. Do various branches of science fit together to make predictions and suggest ameliorative actions in the world of our everyday life, and if they do, do they thereby differ from metaphysics, dogma, ideology? I think much of the question of pluralism in science importantly revolves around these questions and that these questions are not very easy to answer.

An example will illustrate the problems I see for answering this question. Can biology, climate science, economics, anthropology, and whatever other sciences that might be of use combine in such a way as to yield proper guidelines for managing the salmon fishery in the Pacific Northwest?[15] One might have thought that the answer, ascertainable by simple inspection, is NO, but, of course, this is not sufficient. Consider two strategies for answering YES instead:

First, there is the "in principle" YES that says, of course, those sciences could yield such guidelines if they were all sufficiently far advanced to have decently articulated theories or models that could interact with one another. This response yields the "inverse instrumental argument for more science funding" since it is exactly the failure of instrumental success in climate or fisheries science models that indicates that they need more funding.

Second, there is the "immunizing" YES that says, of course, those sciences could yield such guidelines if they were not met throughout by improper special interests, social pieties and taboos, and other contaminants. Advocates of this strategy abound—they typically "know what the science says" but blame the politicians, the bureaucrats, the treaties that grant rights to native peoples, the greed of fish farmers, and other non-scientific factors for the mess. (That no two of them agree on "what the science says" only reinforces each of them in their belief that the others are in the grip of special interests.) If science could proceed untrammeled, it could speak to the general human interest. Of course, one person's contamination is another person's contribution to the debate. Thus, one person might argue that it is multiculturalism gone berserk to expect that fisheries scientists should have to listen to native claims of knowledge of salmon breeding habits—this is the functional equivalent of insisting creationists have a voice in biological debates. But another might argue that the accounts of the native fishing communities in fact do form the best evidence we have of changing salmon populations in the coastal waters over the long term, and after all, it is racist to deny that there is aboriginal expertise that should be heard in the debates. Someone might even argue that the native ways of interacting with the environment are more likely to yield knowledge of what is happening with the fish, a claim that sounds like a sort of romantic fairy tale to others.

I argue that this sort of debate is the crux of the matter regarding unity and plurality of science. I have argued, in essence, that unity of science is the flip side of the demarcation problem for the logical empiricists, and, as we should all be aware by now, the demarcation problem was for them and, even more explicitly, for Popper the way of dividing legitimate critical voices in the culture from "the enemies of the open society." A pluralism that becomes a sort of epistemic multiculturalism and that thus denies the relevance of the demarcation problem ultimately becomes not a plurality of science position so much as an implicit faith in liberalism: if we all work together, we will solve our problems. Such a position need have no account of science, whether pluralistic or unified, since science has no special place in its envisioned world. The fear is that destructive and divisive dogmatists cannot be removed from the table and that one achieves only a cacophony of opinions and no consensus for action in such a world. The unity of science serves as a sort of credentialing system for those worried in this way: before taking your seat at the table, you must go see Herr Carnap in the Unity of Science Bureau and have your papers stamped "in proper epistemic order."

The trouble, as those like Kuhn who brought the variety of scientific practices into their central position today have pointed out, is that overly restrictive criteria used in the bureau would have ruled out all but a few exact scientists and the members of the bureau itself. No one has ever solved the demarcation problem in a way that allows it to do the job of sorting the cranks from the legitimate voices. A radical pluralism—one that denies monism, for example—may very well be motivated in the end by the desire to show that the politics of demarcationism were themselves illegitimate, that no universal criteria apply in all cases of sorting the wheat from the chaff.[16] Pluralism of this type may ultimately rest on the conviction that no one is inherently an epistemic goat, forever separated from genuine knowledge by recognizable epistemic sin.[17]

I suggest then that our brief historical voyage has offered not merely some possible pluralist positions, but has also provided some topics as ways forward in pluralist philosophy of science. These ways forward do not combine into a single doctrine to be inscribed in the hearts and minds of pluralists. I have been arguing that the logical empiricists were quite right to call "unity of science" a "movement" rather than a doctrine; in the same spirit, I offer pluralism as a movement that can proceed on a variety of fronts. It is much more appropriate for pluralism to be a set of projects than for it to be a thesis.

1. One way is to take seriously the idea not that there is no difference between science and ideology or dogma but that no universal difference that can withstand critical scrutiny has been enunciated. Pending further specification of an epistemologically crucial distinction, one might very well take a page from some sociology of science and propose to treat science as "ordinary expertise" no different in kind from the expertise of politicians, plumbers, or, for that matter, philosophers. This move removes the need for fretting about a demarcation criterion for science and, thus, for finding some crucial thing that binds physics together with chemistry but not to Hegelianism, Islam, or party politics. We may as well see fully what the theoretical consequences of such a view are. Let's explore multicultural pluralism at least in the ideal and imagine a thousand flowers blooming—it might not be unattractive.[18]

2. Another way forward would be to insist that there is an important universal demarcation of science from nonscience and, indeed, an important sense in which the various branches of science can and, at times, do cooperate with one another that does not and cannot extend to cooperation with the nonscientific fields. Here, pluralism of a descriptive kind is definitely helpful, if only as an aid to clarifying the phenomena. Suppose there is a clear sense in which the quantum formalism of physics does not quite match the quantum formalism of chemistry—does this hinder cooperation and sharing of theoretical knowledge? When? How? Does this differ in kind from, say, the difference between forms of representation of meaning in linguistic semantics and in philosophical semantics? Are the meanings of pain, illness, and suffering in "the lifeworld" or in "alternative

medicine" importantly different from and unable to be commensurate with the meanings of these things in biomedicine? Before we seek to *save* the phenomena of unity of science, try finding out what the phenomena are; and, in the situation in which find ourselves as theorists of science now, we will have to explore unity despite diversity.[19]

3. If my diagnosis of the demarcation issue for the logical empiricists is correct, then one issue of relevance to the pluralists is whether the pre-theoretical concerns of contemporary philosophy of science are quite the same as those of the logical empiricists. I have suggested that the concern among the logical empiricists was a demarcation of proper and improper epistemic authority—ideology as a threat to human life and science as the enemy of ideology. The sort of pluralism that seems to raise current political-cum-scientific ire is motivated not so much by an external threat from metaphysics or ideology or politics to science or to life, but, as we argued above, a hegemonic or imperialist threat to some scientific ways forward at the hands of others—let us not forget the social factors in autism in a mad rush to a genetic explanation—or even, at times, a worry that science has hegemonic dominance over other ways of knowing that have not been given a chance to express their insights. This is to say that the global epistemic situation has changed—science is less challenged by nonscientific rivals than it is, at least in part, in danger of losing its humility with respect to other aspects of human life and belief. So we can become explicit in our anxieties about hegemonic stances within science. Moreover, we may wish to borrow not merely from our friends in history and sociology of science but also from our friends in Continental philosophy to help us theorize knowledge as power.

4. I have, at last, used a phrase that I have felt the absence of so far: epistemic situation. One of the authors in the *Encyclopedia,* John Dewey, offered a form of pragmatism that relativized knowledge and methods to situations. Pluralism seems often presented in the spirit of situational pragmatism. We might wish to know the causal pathways internal to the individual that put the individual at risk for breast cancer. We ought not, however, lose sight of other, equally valid, questions: what environmental factors put women at risk for breast cancer, when might the individual and environmental risks properly lead to radical prevention including the psychological and physical trauma of prophylactic mastectomy, and so on. In other words, pluralism is not simply a reminder that there are many representational schemes and many things to know, but an effort to remind ourselves within a scientific culture that there are many different actions we might seek to take, many ways we might wish to intervene in the world. We might wish to object to a theory of truth that says that belief is true if it satisfies us, but doctrines and methods are very often pursued because

they are thought likely to satisfy us. Pluralism prods us to wonder about what ails us and what satisfies or might satisfy us.

5. In the great era of the unity of science, science was theorized as the epistemic face of proper democratic society. Scientific culture, for Popper, was the model open society; scientific habits of mind, for Dewey, were the same habits of mind of liberal democracy; the scientific communities, for Robert Merton, were bound together by an ethos almost indistinguishable from the ethos of a properly constituted democratic society. Two things have happened since the 1930s. First, liberal democracy has itself been put under pressure; in particular, multiculturalism and other points of view have raised serious questions about the proper scope and limits of liberal democratic tolerance. Second, various theorists of science, such as Kuhn, Michael Polanyi, and David Bloor, have argued that scientific culture is better theorized not as a liberal society but as a conservative one, bound together as local communities of practice and pedagogy, as keen to limit as to promote dissent, and so on. The problems of liberal political theory and the problems of proper theories of science are two of the great issues of our day, and, given the background assumption that science and liberalism went together in the theories of science in the early twentieth century, the two issues are deeply interrelated.[20] Steven Shapin and Simon Schaffer (1985) argued that problems of the epistemic order were at the same time problems of the social order. Proper attention to the issues in, motivations for, and contexts of scientific pluralism in contemporary philosophy of science suggests that Shapin and Schaffer were right.

Notes

1. I hasten to note that Morrison does not seek to promote unity at the expense of disunity, but rather seeks a proper understanding of both. Her main targets may even be not the disunity camp but those who tie unification in science to explanation and understanding, such as Friedman and Kitcher.

2. See, for example, Friedman 1999; Richardson 1998; Uebel 1991, 2000; Stadler 2001; and the essays in Giere and Richardson 1996 and Hardcastle and Richardson 2004.

3. I had a historian colleague at the University of British Columbia who told a hilarious story of being very annoyed when Ian Hacking appeared at a meeting of the Canadian Society for History and Philosophy of Science meeting in the late 1970s or early 1980s in order to say that he (Hacking) had "discovered laboratories." The existence of laboratories was not news to historians of science. Of course, Hacking did not think he had discovered something; he was arguing that philosophers had ignored something obvious about science to the peril of philosophy of science. Hacking has been arguing ever since that philosophy of science should pay attention to laboratory science. The pluralists strike me as making a similar claim about variety of different sorts within science. The question really is, what about this variety is philosophically interesting?

4. It is rather explicit in Dupré 1993, 1996; and Longino 2002.

5. See Carnap [1928] 1967, 122, for the reference to Cassirer in this context.

6. I take this up in greater detail in Richardson 1998. Since the question of whether laws expressed relations of universals to particulars was at issue in the debate, one can see quite clearly why this methodological debate about the cultural sciences ultimately feeds into the debate about the nature of logic and why Carnap thought the relational logic of Whitehead and Russell's *Principia Mathematica* was exactly the technical tool needed to make precise Cassirer's point contra Rickert.

7. The trouble is simple, really: Philosophers insist that the phenomenon of not knowing what it would be like actually to believe something stems always from not being able to understand what it is that one is asked to believe. But I can articulate pretty well what Osama bin Laden or George W. Bush believes while being unable to imagine anyone both being me and believing it. When Bush acts as if the attack on the World Trade Center was wholly incomprehensible and when bin Laden acts as if the actions of the United States are incomprehensibly evil, *they* perform the philosophical problem of relativism, but they lose their credibility in so doing.

8. Longino reads Kitcher's use of unification as "monistic" (2002, 67), but one might achieve illumination as well as amusement by conceiving of Kitcher's (1993) project as Hegelian at its core: our final unified picture of the world is not true because we have a guarantee that it matches the structure of the world; rather, it is true because it is our final and unified account of the world, which is what we mean by "the structure of the world." Resistance to unification is the Negation that must be rendered into Nothing at the next step of unification. In the absence of resistance we have both maximal unification and absolute Being. I have read Friedman's recent work as a form of Hegelianism also, although Friedman tends toward a more Kantian version than does Kitcher; see Richardson 2003a.

9. Longino does not follow Bergson in the pluralism Bergson wants to stress, which is not really a pluralism within science but a pluralism between science and metaphysics. For Bergson, the representational poverty of symbol systems, which are necessary for science, comes from a direct metaphysical intuition into the nature of things in metaphysics. It is of interest, perhaps, that one of Carnap's explicit targets in the antimetaphysical final portion of the *Aufbau* (Carnap [1928] 1967, 288–300) is Bergson's "intuitive metaphysics."

10. For more on coordination as a key concept in neo-Kantian epistemology, see Ryckman 1991.

11. Interestingly, pluralism does not differ from unity by saying there are not one but many languages all adequate for the whole of science, but by saying that there are no such globally adequate languages. I suppose "the nullity of science" is not a rhetorically winning slogan for a philosophical project.

12. As Creath (1996, 159) says, Carnap is committed to "the requirement that the applicability of all the concepts that we use in science must at least sometimes be empirically and publicly testable. This is pretty benign."

13. Thus, I quite concur with the way Creath (1996) expresses the point of Carnap's unity of science doctrine, especially his remarks on the *publicity* of science.

14. This context is no longer news. It is evoked prominently in Galison 1996, Hacking 1996, and most tellingly in Reisch 2005.

15. Should one seek to claim that such questions are not of much philosophical interest, I wish to note that in May 2003, the Province of Newfoundland was seeking to renegotiate the Canadian Constitution exactly because of issues of resource management and the disappearance of the Atlantic cod fishery. This is one place where issues of scientific knowledge, proper governance, legitimate authority, etc., meet and really matter to people's lives. If that is not philosophically interesting, so much the worse for philosophy.

16. Here as elsewhere, philosophy of science can seek help from sociology of science. A sociological point of view that stresses the practical need for *groups of scientists* to demarcate themselves from nonscientists without trying to find the criteria by which all science everywhere is demarcated from all nonscience can be found in Gieryn 1999.

17. Some might, uncharitably, say that the internalized standards of analytic philosophy of science do pass muster with the bureau and have thus created a society of self-styled epistemic sheep, and that this is why the whole field is so constipated and irrelevant. We philosophers of science form a radically paranoid society, such critics might proceed to say, that has become void of any content and is concerned wholly and only with the upholding of such standards within the community—a form of structural Puritanism in which the standards are themselves the highest value (the latest issue of *Philosophy of Science* as the latest Word of God) and for which any violation of the standards constitutes dissolution and radical evil (Steve Fuller as the Devil).

18. I hasten to add that sociologists who think science is best understood as "ordinary expertise" do not seek thereby to advocate this sort of Feyerabendian anarchism; Collins and Pinch (1993), for example, use the idea mainly as a resource for understanding how controversial science achieves or fails to achieve consensus.

19. Should it be unclear, allow me to say explicitly that this is an endorsement of, not an objection to, the argumentative strategy of Morrison (2000). Of course, such a project may fail, leading us back to a project more like Gieryn's (1999) in which the boundaries around any specific branch of science do get drawn for a time and a place without "the demarcation problem" ever being solved.

20. I deal with these issues in greater detail in Richardson 2003b.

References

Carnap, R. [1928] 1967. *The Logical Structure of the World*. Berkeley and Los Angeles: University of California Press.

———. 1938. "Logical Foundations of the Unity of Science." In Neurath et al. 1938, 42–62.

Cartwright, N. 1999. *The Dappled World: Studies of the Boundaries of Science*. Cambridge: Cambridge University Press.

Cartwright, N., J. Cat, T. Uebel, and L. Fleck. 1996. *Otto Neurath: Philosophy between Science and Politics*. Cambridge: Cambridge University Press.

Cassirer, E. [1910] 1953. *Substance and Function*. New York: Dover.

Collins, H. M., and T. Pinch. 1993. *The Golem: What Everyone Should Know about Science*. Cambridge: Cambridge University Press.

Creath, R. 1996. "The Unity of Science: Carnap, Neurath, and Beyond." In Galison and Stump 1996, 158–69.

Dewey, J. 1938. "Unity of Science as a Social Problem." In Neurath et al. 1938, 29–38.

Dupré, J. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, Mass.: Harvard University Press.

———. 1996. "Metaphysical Disorder and Scientific Disunity." In Galison and Stump 1996, 101–17.

Fodor, J. 1975. *The Language of Thought*. New York: Crowell.

Friedman, M. 1999. *Reconsidering Logical Positivism*. Cambridge: Cambridge University Press.

———. 2001. *The Dynamics of Reason*. Stanford, Calif.: CSLI Press.

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

———. 1990. "*Aufbau/Bauhaus:* Logical Positivism and Architectural Modernism." *Critical Inquiry* 16: 709–52.
———. 1996. Introduction. In Galison and Stump 1996, 1–33.
Galison, P., and D. Stump. 1996. *The Disunity of Science: Boundaries, Contexts, and Power.* Stanford, Calif.: Stanford University Press.
Giere, R. N., and A. Richardson. 1996. *Origins of Logical Empiricism*. Minneapolis: University of Minnesota Press.
Gieryn, T. 1999. *Cultural Boundaries of Science: Credibility on the Line*. Chicago: University of Chicago Press.
Hacking, I. 1996. "The Disunities of the Sciences." In Galison and Stump 1996, 37–74.
———. 2001. *Historical Ontology.* Cambridge, Mass.: Harvard University Press.
Haraway, D. 1991. *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.
Hardcastle, G., and A. Richardson. 2004. *Logical Empiricism in North America*. Minneapolis: University of Minnesota Press.
Hempel, C. G. 1993. "Thomas Kuhn, Colleague and Friend." In *World Changes,* ed. P. Horwich, 7–8. Cambridge, Mass.: MIT Press.
———. 2001. "The Meaning of Theoretical Terms: A Critique of the Standard Empiricist Construal." In *The Philosophy of Carl G. Hempel,* ed. James H. Fetzer, 207–16. Oxford: Oxford University Press.
Kitcher, P. 1993. *The Advancement of Science*. Oxford: Oxford University Press.
Latour, B. 1999. *Pandora's Hope: Essays on the Reality of Science Studies.* Cambridge, Mass.: Harvard University Press.
Lewontin, R. C. 1991. *Biology as Ideology*. New York: Harper-Collins.
Longino, H. 2002. *The Fate of Knowledge*. Princeton, N.J.: Princeton University Press.
Morrison, M. 2000. *Unifying Scientific Theories: Physical Concepts and Mathematical Structures*. Cambridge: Cambridge University Press.
Neurath, O. 1938. "Unified Science as Encyclopedic Integration." In Neurath et al. 1938, 1–27.
Neurath, O., R. Carnap, and C. Morris. 1938. *The International Encyclopedia of Unified Science*. Chicago: University of Chicago Press.
Neurath, O., R. Carnap, and H. Hahn. [1929] 1973. "Wissenschaftliche Weltauffassung: Der Wiener Kreis." In *Empiricism and Sociology,* ed. Marie Neurath and R. S. Cohen, 299–318. Dordrecht: Reidel.
Peirce, C. S. [1885] 1992. "An American Plato: Review of Royce's *Religious Aspect of Philosophy*." In *The Essential Peirce*, vol. 1., ed. N. Houser and C. Kloesel, 229–241. Bloomington: Indiana University Press.
Reisch, G. 2005. *How the Cold War Transformed Philosophy of Science: To the Icy Slopes of Logic*. Cambridge: Cambridge University Press.
Richardson, A. 1998. *Carnap's Construction of the World*. Cambridge: Cambridge University Press.
———. 2003a. "Narrating the History of Reason Itself: Friedman, Kuhn, and a Constitutive A Priori for the Twenty-First Century." *Perspectives on Science* 10: 253–74.
———. 2003b. "Tolerance, Internationalism, and Scientific Community in Philosophy." In *Philosophy of Science and Politics,* ed. F. Stadler and M. Heidelberger, 65–89. Vienna: Springer Verlag.
Rickert, H. [1929] 1986. *The Limits of Concept Formation in Natural Science*. Cambridge: Cambridge University Press.
Ryckman, T. 1991. "*Conditio sine qua non*? Zuordnungen in the Early Epistemologies of Cassirer and Schlick." *Synthese* 88: 57–95.

Shapin, S., and S. Schaffer, *Leviathan and Airpump: Hobbes, Boyle, and the Experimental Life*. Princeton, N.J.: Princeton University Press.
Stadler, F. 2001. *The Vienna Circle in Context*. Vienna: Springer Verlag.
Uebel, T. E. 1991. *Overcoming Logical Positivism from Within*. Amsterdam: Rodopi.
———. 2000. *Vernunftkritik und Wissenschaft: Otto Neurath und der erste Wiener Kreis*. Vienna: Springer.

2
Perspectival Pluralism

In this chapter I explore the extent to which a perspectival understanding of scientific knowledge supports forms of scientific pluralism. I will not initially attempt to formulate a general characterization of either perspectivism or scientific pluralism. I assume that both are opposed to two extreme views. The one extreme is a (monistic) metaphysical realism according to which there is in principle one true and complete theory of everything. The other extreme is a constructivist relativism according to which scientific claims about any reality beyond that of ordinary experience are merely social conventions.

I begin with an exemplar of perspectival knowledge, namely, that gained through color vision, and go on to consider possible pluralistic implications of this sort of knowledge. I will attempt to extend the lessons from this exemplar to scientific observation generally and then to scientific theorizing. After exploring connections between the complexity of the world and a scientific pluralism, I conclude by locating the whole discussion in a broader framework of a scientific naturalism.

Color Vision

There is now a substantial body of scientific knowledge about the workings of the human visual system.[1] In fact, the normal human has several visual systems. One, utilizing rod-shaped receptors, is sensitive to light across a wide range of wavelengths from roughly 400 nm to 650 nm (one nanometer equals one billionth of a meter), with a peak sensitivity around 500 nm. This system is basically achromatic. A second, chromatic system utilizes three different cone-shaped receptors containing different pigments with peak sensitivities at roughly 450 nm, 530 nm, and 560 nm, although there is considerable overlap in sensitivity. These three pigments are conventionally labeled S(hort), M(edium), and L(ong). What gets transmitted to the brain for color vision are *differences* in the activation of the three types of pigments, and these differences stimulate two opponent systems, a red-green system and a yellow-blue system, which together produce the experience of all color hues. The following is a schematic rendering of the neural code, where L, M, and S now represent the respective total activation levels of the three types of pigments:

$(L - M) > 0 \Rightarrow$ Red	$[(L + M) - S] > 0 \Rightarrow$ Yellow
$(L - M) < 0 \Rightarrow$ Green	$[(L + M) - S] < 0 \Rightarrow$ Blue

This scheme provides a scientific explanation of the fact that no surface can appear to be uniformly reddish green (or greenish red). Because of the nature of this code, the experience must be reddish or greenish or neither, but never both. Similarly for yellow and blue. These are not a priori truths, as some philosophers have argued, but experiences for which there is now a straightforward *scientific* explanation.

The fact that downstream color processing works with *differences* in activation also explains the phenomenon of *color constancy,* that is, the fact that the perception of colors is relatively invariant with respect to the absolute *intensity* of light over a wide range of intensities. Relative color differences, for example, appear roughly the same in shadow as in bright sunlight. Color constancy helps to explain why humans are so strongly inclined to regard colors as inherent properties of objects themselves.

In short, normal humans have a species-specific *colored visual perspective* on the world. The specific characteristics of this perspective are due to contingent evolutionary circumstances. Other species, including some birds, are *tetrachromats,* possessing *four* different color-sensitive visual receptors, the fourth typically being in the near-ultraviolet region of the spectrum. On the other hand, most mammals, including domestic dogs and cats, are only dichromats, which means they have only one opponent chromatic processing system.

Even among humans there are major differences in chromatic experiences. It is estimated, for example, that roughly 8 percent of Caucasian males have some degree of so-called red-green color blindness, so that their chromatic visual system resembles that of a dichromat.[2] More dramatically, for genetic reasons, some humans suffer from rod achromatopsia. Such people, also called rod monochromats, have no cones, only rods, and thus have no chromatic experiences whatsoever. These humans experience the world only from a black-and-white perspective.

Focusing on the difference between normal human trichromats and rod monochromats, we have a clear candidate for a significant case of *perspectival pluralism.* The visual perspective of most humans is chromatic, while for some the visual perspective is solely black and white. What features of this example might we suppose to be general characteristics of a perspectival pluralism?

One feature that stands out is that there is no straightforward way to claim that one perspective, say the colored perspective, is objectively correct, or in some sense uniquely veridical, while the black-and-white perspective is incorrect, or nonveridical. The fact is that normal human trichromatic color vision is in no way uniquely correct. It is merely the system that evolved along the evolutionary path to *Homo sapiens*. Within this normal-colored perspective, one can, of course, distinguish judgments that would generally be regarded as correct from those that would be regarded as incorrect. Grass normally appears green, though sometimes brown; the sky often appears blue, and hardly ever green. One may remember a friend's eyes as being blue when, on inspection, they appear dark gray. So, within the normal colored perspective, color judgments can be quite robust. The same holds, however, for the black-and-white perspective. A rod monochromat might remember a surface as being dark with light spots when, on inspection, it appears as light with dark spots. And there could be quite robust agreement among rod monochromats as to the relative lightness of white pine trees as opposed to Norwegian pines.

On the other hand, there is a clear sense in which the colored perspective is *richer* than the black-and-white perspective. There are red and green surfaces that would be indistinguishable to a monochromat. A trichromatic system, we might say, is capable of extracting more information from the environment. So while it is possible to create a colored image from a black-and-white image by assigning colors to shades of gray, it is clear that a colored image that would be judged correct from the perspective of a normal trichromat cannot be unambiguously extracted from a black-and-white image of the same scene.

To proceed further, it is helpful to introduce an idealization of a type common in scientific studies of color vision. Imagine a subject in a dark room with a neutral surface on one wall. An idealized projector illuminates the surface in such a way as to control the total intensity distribution of visible light reflected from the surface to the eyes of the subject, that is, the relative intensity of the light as a function of wavelength across the whole visible spectrum. This idealization provides a theoretically unique physical description of the light reaching the retina of the subject.

Is it possible to infer how a particular intensity distribution will be experienced by a monochromat from how it is experienced by a normal trichromat? The answer, perhaps surprisingly, is no. The experience of the monochromat is determined by two things: the intensity distribution of the light reaching the retina and the relative sensitivity of the monochromat's pigment as a function of wavelength. Obviously, the response of the trichromatic system provides no information about the relative sensitivity of the monochromat's pigment. But the intensity distribution of the light reaching the retina cannot be inferred from the response of the trichromat for the simple reason that *various* intensity distributions can produce the *same* response in a normal trichromatic visual system. The relationship between intensity distributions and chromatic response is many-one.[3]

For exactly the same reasons, one cannot infer the response of a nonnormal trichromat from that of a normal trichromat, where a nonnormal trichromat is one for whom the responses of the three color-sensitive pigments have different peak sensitivities. Thus, even among different sorts of trichromats, there is a pluralism of perspectives. They respond differently to the same intensity distributions. And each response may provide a specific set of both advantages and disadvantages.

Granting that the colored and black-and-white perspectives are *different,* are they *compatible*? It might seem that they are not. For example, a monochromat might claim to see a rug as being of uniform brightness while a trichromat sees a red pattern on a green background. Regarding these claims as incompatible, however, presumes that colors are objective properties of objects rather than products of an interaction between light reflected from objects and particular types of chromatic visual systems. There is no incompatibility in the fact that the pattern can be detected by a trichromatic visual system but not by a monochromatic system.[4]

Can we not say that the reason why different visual systems will not produce genuine conflicts is because they are all interacting with one and the same environment? In our idealized experimental situation, for example, both monochromats and trichromats may face the same intensity distribution of light. In this case, it seems that the uniqueness of the world experienced guarantees the compatibility of different perspectives. I will postpone inquiring into the status of the claim that the world itself has a unique structure.

The most common objection to pluralism of any form is that it may lead to an undesirable relativity. On a perspectival understanding of color vision, however, while there is relativity to a chromatic system, this relativity seems not to be especially objectionable. The agreement on color judgments among people sharing a chromatic perspective is quite robust. I will also postpone raising the obvious question of from what perspective do I make the preceding claims about the nature of color vision and about relationships among different chromatic perspectives. I want first to extend the analysis of color vision to scientific observation more generally.

Scientific Observation

Today, virtually all scientific observation involves the use of instruments. My examples are from astronomy and astrophysics, although almost any contemporary science would do. Readers should think about how what I say here applies to the scientific instruments with which they are most familiar.

Consider standard black-and-white photographs of some familiar celestial object, such as the Milky Way, produced with ordinary optical telescopes. Such photographs provide us with a black-and-white *visible* light perspective on the Milky Way, regular photographic film being designed to register light with wavelengths visible to humans. But there are many other perspectives.

For several years, the Infrared Space Observatory, a satellite-based facility, produced infrared images of many celestial objects, including the Milky Way. Infrared light provides a very different perspective on the universe, one not directly accessible to humans. The colors one sees in infrared images are *false* colors produced by computer manipulation of the original data. Infrared light is of particular interest to astronomers for several reasons. Because it has the shortest wavelengths, which are not mostly absorbed by intergalactic dust, infrared images provide the greatest possible resolution for very distant objects that are typically obscured by intergalactic dust. Additionally, of course, since the light reaching us from very distant objects will be red-shifted, many can best be seen in the infrared.

To say that scientific observation is perspectival relativizes observations to the perspective of the relevant instrument. There is no such thing, for example, as *the* way the Milky Way looks. There is only the way it looks to each instrument. Moreover, even if it were physically possible to build an instrument sensitive to the whole electromagnetic spectrum emitted by a distant galaxy, it would still be blind to things such as neutrinos, which, we presume, are also emitted. There just is no universal instrument that could record every aspect of any natural object or process.

It follows that, to maximize observational knowledge of the world, a plurality of instruments is required, depending on what interesting phenomena there are reasons to suspect might be detected. Thus, among the first satellite telescopes was the Compton Gamma Ray Observatory, designed to observe such things as supernovas, black holes, and galactic cores, all thought to produce significant quantities of gamma rays.

Not only are all instruments limited to recording only a few aspects of the world, they do so with only limited accuracy. There is no such thing as a perfectly transparent instrument. This means that there will always be a many-one relationship between the inputs and the recorded outputs, a relationship determined not only by the inputs, but also by the nature of the instrument. So we may say that part of the perspective of any instrument is its built-in margin of error.[5]

A significant difference between unaided observation and observation with instruments is that the output of an instrument is a *public* object, now often a computer-generated image, but always something available for public scrutiny. In the end, of course, there is going to be some unaided observation of something. Otherwise the end product could not be *human* knowledge. Part of the design of a good instrument is making the output salient to human observers. Nevertheless, we should not forget that the public object itself is always the product of an *interaction* between an instrument and some aspect of the world. Instruments interact with the world from different perspectives than humans, but never from no perspective.

In the case of human color and black-and-white vision, we concluded that the different perspectives are consistent and even complementary. The perspectives of the various instruments used to measure radiation from parts of the universe are likewise both consistent and complementary. A source of infrared light, for example, may often be identified with an object observed optically. Thus, the plurality of perspectives found in scientific observation does not generate an undesirable relativism. Indeed, observational plurality is compatible both with a restricted (perspectival) realism about the objects of observation and also with ordinary standards of evidence for claims about those objects. As with color vision, the compatibility of different observational perspectives could easily be understood as a consequence of all observations being observations of a unique world. I will again postpone considering the status of this possibility.

Scientific Theorizing

I turn now to the second question earlier postponed. From what perspective have I been talking about the nature of color vision or the capabilities of various instruments to record radiation from astronomical sources? Behind this question lurks the suspicion that the whole discussion so far assumes an unacknowledged monism. The general answer is that my discussion presumes various *theoretical* perspectives. Theoretical perspectives are indeed broader than observational perspectives, but they remain perspectives, not absolutes. To see this we need to look more closely at what constitutes a theoretical perspective.

The metaphysical realist (monist) understanding of scientific knowledge is facilitated by a particular understanding of the nature of theories and theorizing. A key component of this view is that theories consist of sets of laws of nature that should be both true and universal. I won't attempt to give my reasons for rejecting this view here.[6] I share this position with a number of philosophers of science including Nancy Cartwright (1983, 1999) and Bas van Fraassen (1980, 1989). What needs to be clear is that my perspectival view of scientific theories invokes an alternative understanding of theorizing.

I understand theorizing as the construction of models. Sometimes these models are actual physical objects, such as Watson and Crick's original sheet metal and cardboard model of DNA. Mostly, however, they are abstract objects. Often, as in physics, abstract models are characterized using interpreted mathematical expressions, but often not, as in much of biology. How models are characterized is not of fundamental importance. Some theorizing utilizes high-level principles such as the principle of inertia, the principle of relativity, or the principle of natural selection. I understand these not as laws in the sense of empirical generalizations, not even as statements about the world. They function, rather, as descriptions of highly abstract models that serve as templates for the construction of more specific (although still abstract) models. Their connection with the world is thus indirect. The world seems to be such that specific models structured according to the principles can be made to exhibit a close fit with systems in the real world. But much model construction in science, as in ecology and more applied sciences, proceeds without strong principles, relying mainly on mathematical techniques.

Theorizing and observing have both similarities and dissimilarities. One obvious difference is that, once in the proper situation, the world impinges on humans or instruments in causally *direct,* though maybe not fully determining, ways. Theorizing is also constrained by causal interaction with the world, but only *indirectly* through experimentation, which involves comparing a theoretical with an instrumental perspective. I am more interested in the similarities. In particular, theorizing is always partial in the way observing is partial. As an instrument may be able to record either infrared or ultraviolet, but not both, theoretical principles may deal with mechanical forces or with electromagnetic propagation, but not both. Newton's equations, I would say, define a particular mechanical perspective on the world; Maxwell's equations define an electromagnetic perspective.

As Einstein pointed out, the principles of Newtonian mechanics and Maxwellian electrodynamics are *incompatible*. It was partly due to the perceived incompatibility of these two perspectives that Einstein developed the special theory of relativity. But the principles of special relativity also define only a perspective, although a more general one than the two perspectives it replaces. The perspective of special relativity differs, for example, from the perspective of general relativity. It is now well known that the principles of standard quantum theory and those of general relativity are also incompatible, quantum principles presuming a flat space-time as opposed to the curved space-time of general relativity. In my terms, there is no possible abstract model defined by the conjunction of principles from these two theories. For the metaphysical realist (monist), this is an unsatisfactory state of affairs. The one true and complete theory of the world cannot be an inconsistent theory; what are now typically taken to be the laws of general relativity and quantum theory cannot all be true. It is a reflection of their implicit metaphysical realism that many theoretical physicists are currently engaged in a search for a "theory of everything" (Weinberg 1992). What are the implications of a perspectival pluralism for this situation?

I see no way a perspectival pluralist could argue that the search for grand unified principles of fundamental physics cannot succeed. Or that the current pluralism of incompatible principles is in itself a good thing. On the contrary, unifying perspectives, when one can find them, are scientifically desirable on many grounds. One must not forget, however, that the supposed unified perspective would still be a perspective determined by whatever might be the principles of that new perspective. On the other hand, from a perspectival point of view, one need not be too upset with the current situation in theoretical physics. Good theoretical science does not *require* finding genuinely universal principles. Well-fitting models, based on a variety of principles, are good enough. And, indeed, that is all that can be found across most of the sciences.

Complexity

Fundamental physics is a highly atypical part of science, even within physics. The principles of general relativity and quantum theory provide very abstract perspectives on the world. Space-time theory abstracts away even the structure of whole galaxies while foundational studies in quantum theory often focus on isolated individual interactions among elementary particles.[7] For many sciences, the complexity of the subject matter puts a premium on finding even just a few aspects of the subject for which one can construct well-fitting models.

Remaining for the moment within physics, consider the simple case of water.[8] If one is studying diffusion or Brownian motion, one adopts a molecular perspective in which water is regarded as a collection of particles. But the situation is far too complex to adopt a Newtonian perspective for individual particles. Instead, one adopts a statistical perspective in which the primary variables are things like mean free path (the average distance a particle travels between collisions). However, if one's concern is the behavior of water flowing through pipes, the best-fitting models are generated within a perspective that models water as a continuous fluid. Thus, one's theoretical perspective on the nature of water depends on the kind of problem one faces. Employing a plurality of perspectives has a solid *pragmatic* justification. There are different problems to be solved, and neither perspective by itself provides adequate resources for solving all the problems.

Of course, a metaphysical realist will ask, "But what is water, really?" assuming that the answer must be "molecules." But perspectivism yields the desired answer without giving in to monism. Nothing in perspectivism dictates that all perspectives are created equal. Some are better than others in many different respects. In this case, there is a clear asymmetry in favor of a molecular perspective. That is, from within a molecular perspective, one can, in principle, explain how a macroscopic fluid made up of microscopic molecules could be fitted very well within a perspective based on principles regarding continuous fluids. We just don't know how to construct molecular models of macroscopic fluids, and maybe we never will. On the other hand, there is no way to construct models within a continuous fluid perspective to model Brownian motion. So we can say that the world is such that there is, in *principle,* a molecular model for all of the many manifestations of water. In *practice,* there are many manifestations of water that can only be modeled within other perspectives. In this case, while pragmatism dictates a pluralistic attitude toward theoretical perspectives, the intuitions, though not the metaphysics, of the metaphysical realist can be accommodated.

Complexity raises more serious issues in biology. In this respect, biology may provide a better paradigm for the sciences than does fundamental physics. Here the traditional way of dealing with the overall complexity of biological phenomena has been to distinguish *levels of organization:* molecules, cells, organs, whole organisms. The same strategy is used in the social sciences: individuals, small groups, communities, corporations, nation-states. Even a metaphysical materialist who presumes that there is ultimately nothing but elementary particles must agree that there is little hope of finding a usable reductive theory in either biology or the social sciences.[9] So models are constructed at various levels, resulting in a pluralism of perspectives at different levels.

Though some disagree, I think we can speak meaningfully of fully legitimate explanations, even causal explanations, at various levels. We may even develop genuine theoretical principles, such as the principle of natural selection or the law of supply and demand, at levels well above that of elementary particles. It seems an understatement to say that this plurality of perspectives at different levels is just a pragmatic response to complexity. It looks to be a more fundamental and relatively permanent feature of the biological and social sciences.

On the other hand, few would insist that either biological or social levels are completely autonomous, particularly not from lower levels.[10] Molecular genetics provides one of the very best examples of the fruitfulness of looking for mechanisms at a lower level. Waters (chapter 9 in this volume) provides a clear example of conflicts that can arise between groups of scientists working at adjacent levels. The hubris of some molecular biologists in claiming that molecular biology will ultimately be the whole of biology is well known. Waters shows that critics of molecular biology may share the underlying ideology that there should ultimately be a single perspective for all of biology. The two groups thus end up disagreeing mainly on which perspective is ultimately the right one. Waters argues, rightly, I think, that biology, and genetics in particular, is better served by maintaining a plurality of perspectives at different levels.

Longino (chapter 6 in this volume) provides a more complex example of the same phenomenon. She considers four different perspectives on the study of human behavior: behavior genetics, developmental systems theory, neurophysiology, and neuroanatomy. Like Waters, she finds that practitioners in these fields overwhelmingly presume that a single perspective (their own!) provides the best, or the whole, explanation of the behavior in question. She concludes that, in this case, maintaining a plurality of perspectives promotes scientific progress and thereby a better overall understanding of human behavior.

There remain some fundamental questions that are too easily passed over in arguments for the benefits of pluralism. Few of those who argue for a pluralism of *perspectives* presume a plurality of *worlds*.[11] The perspectives in question are typically presumed to be different perspectives on a single world. So what is the status of this presumption? Further, if there is but a single world, presumably it has a unique structure. What, finally, is the status of this presumption? If there is a single world with a unique structure, why should it not be at least a goal of science to discover this structure? In short, why is the metaphysical realist's position not at least an ideal toward which science should strive? But if this is so, what is the basis for a pluralism of perspectives?

Metaphysics and Methodology

The claim that there is but one world with its own unique structure sounds like a bit of metaphysics. It is not as if some scientists had done experiments that indicate that, yes, there is indeed a single world with a unique structure. In an earlier age this might have been proclaimed a "metaphysical presupposition" of science. One can even imagine attempts to construct a Kantian-style "transcendental deduction" of this proposition. Supporting this understanding of the situation is the unspoken assumption that science has a kind of deductive structure. The so-called presuppositions then function as ultimate first premises in all scientific arguments for any empirical conclusion.[12] There is, however, another way to proceed.

A good strategy when confronted with what seems to be a metaphysical presupposition is to reformulate it as a *methodological maxim*. The question is then not whether some general proposition is true, or how it might be justified, but merely whether following the maxim is likely to promote achievement of the goals of a scientific inquiry. This may not be an easy question to answer, but at least it is an empirical question concerning the likely results of following a given maxim. It remains within the general scope of empirical inquiry.[13]

In the present case the maxim would be "Proceed as if there is a single world with a unique structure." As stated, this maxim supports the corresponding metaphysical realists' maxim, "Look for the one true and complete theory of the world, that is, the theory describing the unique structure of the world." If one wishes to recommend a pluralistic stance in some particular area of inquiry, the task is clear. Find reasons why following the metaphysical realist's maxim is not likely to promote scientific progress in the area in question. These should be general empirical reasons.

In the case of fundamental physics, there do not seem to be strong reasons not to pursue a program of unification. Many previous attempts at unification have been successful, Einstein's special theory of relativity being a paradigm case. The only plausible argument for giving up trying to unify general relativity and quantum theory seems to be that the task is just too difficult. If Einstein could not do it, who can? But this argument can also be taken as a challenge to bright people to keep trying. The case of human behavior (Longino, chapter 6 in this volume) seems to go the other way. Human behavior is terribly complex. The four fields Longino examines are quite disparate. It is difficult just to make connections between, say, behavior genetics and neuroanatomy, let alone attempt some sort of unification. And it is plausible, as she argues, that all four fields benefit from mutual criticism. Only the polemics promoting one approach over all others seem dysfunctional.

There is another important implication of the "one world" methodological rule. If there is a clear conflict in a border area between two levels (or even between any two areas of interest), this must be regarded as an anomaly. Of course, the conflict may involve something that is currently not of much concern to anyone, or it may involve something that no one has any idea how further to investigate from either of the conflicting perspectives. So the conflict may legitimately be put on a back burner. But one cannot just ignore the conflict altogether on the grounds that we are all good pluralists. There has to be a good scientific reason not to attempt to resolve it.

We may regard the opinion that it must be possible to unify relativity and quantum theory as an application of this methodological implication regarding conflicts among perspectives. The research into human behavior examined by Longino exhibits differences among perspectives in high-level assumptions about important variables and in methodology. There do not, in this account, seem to be clear-cut examples of empirical results or theoretical claims that are both comparable and conflicting. The perspectives being pursued seem to be largely complementary and nonoverlapping.

Why Not to Expect Any Model Perfectly to Fit the Real World

Even if physicists succeed in formulating principles that consistently unify quantum and gravitational forces, there remains a question as to how well the resulting models fit the actual universe. And here I do not mean merely how well the models fit the *data*. We know actual measurements always include an ineliminable margin of error. So the best that can be expected is agreement within the known margin of error. Suppose that were to be achieved. Could we then justifiably conclude that there is an *exact* fit between the models and the world? I think not.

The question of whether any theoretical claim about the fit of a model to the world could be exactly true is connected to the question of whether any model could be complete in the sense that it encompasses the whole truth about everything. My conclusion is that the only way any particular such claim could be exactly true is if it uses a complete model that fits the world exactly in every respect. The assumption that connects exactness with completeness is that everything is causally connected with everything else by some more or less remote chain of causation. Suppose we have a theory that is not complete. Whatever the subject matter of this theory, there will be some (maybe remote) connections between this subject matter and other things in the universe not part of the subject matter of this incomplete theory. Since there will be some influences on the subject matter not accounted for by our incomplete theory, it cannot be exactly correct. Thus, only a complete theory could generate claims that are exactly true. In the case of fundamental physics, it is not to be expected that the general principles of such a theory could specify such things as the exact distribution of matter in the universe or the distribution of elements, for example, the ratio of hydrogen to everything else.

This statement of the argument invokes what seems like a metaphysical assumption of connectedness in the universe. The argument can be made less metaphysical by assuming only that we do not *know* the extent of connectedness in the universe. It follows that we do not know whether or not any of our theoretical claims about the fit of models to the world are exactly true. This more modest conclusion is sufficient to support a robust (perspectival) pluralism.

A Naturalistic Stance

I hope to have shown that a perspectival understanding of scientific knowledge supports a modest pluralism that avoids the extremes of both metaphysical realism and constructivist relativism. In conclusion, I would like to argue that the resulting pluralism is indeed a scientific pluralism both in the sense that it is a pluralism of scientific knowledge claims and that its justification is itself within a scientific framework.

For me, the most fundamental framework is *naturalism*. Minimally, naturalism implies the rejection of appeals to anything supernatural. More theoretically, it also implies the rejection of appeals to a priori claims of any kind. But already there are problems. First, what can be the naturalist's basis for these claims? Is not the denial of supernatural forces as metaphysical a claim as their affirmation? And can the blanket rejection of claims to a priori knowledge be itself anything less than an a priori claim?

Second, how can one determine the boundary of the natural beyond which lies the supernatural? Here naturalists typically appeal to the findings of modern science, but there are severe problems with this response. Today's natural science cannot determine the boundary of what is natural simply because that boundary keeps moving. Once life itself was considered beyond the province of natural science, requiring a supernatural source. The same claims are even now sometimes made for human consciousness. The naturalist assertion that consciousness will eventually be given a natural scientific explanation begs the question against the supernaturalist.[14]

Both of the above problems can be eliminated by taking naturalism not as a doctrine but as a *methodological stance.*[15] When confronted with a seemingly intractable phenomenon, the naturalist supports research intended to produce a natural scientific explanation. The naturalist hopes, even expects, that this research will eventually be successful. This stance can be justified, to the extent that it can be justified at all, simply by appeal to past successes. We have explained life scientifically. Why not consciousness?

Note that the strategy of replacing metaphysical doctrines by methodological stances is less appealing to a supernaturalist or a priorist. The naturalist can wait until success is achieved. And there are good naturalist standards for when this happens. One typically appeals to supernatural explanations or a priori principles because of a pressing need to resolve some issue. Few theists since Pascal have found anything like methodological theism very satisfying. Moreover, it is debatable whether there are equally good criteria for successful supernaturalist or a priorist projects.

Methodological naturalism is not a weak position. The methodological naturalist is free to criticize arguments for metaphysical and a priorist claims. A naturalist would typically attempt to show that such arguments are question begging, lead to an infinite regress, or are in some other way unsound. This makes methodological naturalism a strong position. It provides a comfortable background for a modest scientific pluralism.[16]

Notes

I wish gratefully to acknowledge the support and hospitality of the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS). I thank also the editors of this volume and two anonymous reviewers for many helpful comments and suggestions regarding earlier drafts of this essay.

1. I have relied heavily on Hurvich's (1981) classic modern text. For a good collection of articles on the science of color see Byrne and Hilbert 1997b.
2. The corresponding percentage for females is less than 1.
3. In color science this phenomenon is known as metamerism.
4. To say that colors are perspectival implies that colors are *relational* properties. For an introduction to debates among color objectivists, subjectivists, and interactionists, see Byrne and Hilbert 1997a.
5. This is true also of the human chromatic visual system, which has a resolution of at best plus or minus five nm depending on both the wavelength and the intensity of the light.
6. But see Giere 1988, 1999, 2006.
7. Among philosophers of physics, Nancy Cartwright stands out as a clear exception to the general preoccupation with highly idealized and abstract systems.
8. This example has been examined in detail by Margaret Morrison (1999) and by Paul Teller (2001, 401).

9. On this, Stephen Wolfram (2002) and other advocates of "digital physics" may be exceptions.

10. Nevertheless, in his many publications Jerry Fodor has maintained, wrongly, I think, that the psychological level is autonomous from the neuronal level.

11. Possible exceptions include Nelson Goodman (1978) and Thomas Kuhn (1962, 110), who suggested that "after a revolution scientists are responding to a different world." Hacking (1999) may also be an exception.

12. This way of thinking about science is not restricted to the Kantian tradition. It appears in the empiricist tradition in the form of J. S. Mill's principle of the uniformity of nature.

13. I remember this maxim from older discussions of determinism. Rather than asking whether determinism is true, one need only ask whether proceeding as if were true in particular cases is likely to promote scientific progress in those cases. I do not recall seeing the more general formulation I give in the text, but I assume it has appeared in print more than once. The maxim regarding determinism is particularly interesting because it ultimately proved not to be fruitful in microphysics. Though many regarded this as a metaphysical crisis, it need not have been regarded as anything more than a limitation on the scope of a methodological maxim. Deterministic models continue to be useful in many sciences.

14. Some naturalists have been tempted by the idea that the boundary of the natural will be determined eventually by the end of scientific advance. But this claim is empty. We do not now know when that will be or what the state of scientific knowledge will be then. How could we even know that we had reached the end of scientific advance?

15. Here I adopt the terminology of Bas van Fraassen's *The Empirical Stance* (2002).

16. My concern in this essay has been with *scientific* pluralism. Many times when a group, scientific or not, advocates pluralism, it is primarily just a strategic move in the game of trying to dominate a field or profession. Those in the minority proclaim the virtues of pluralism in an effort to legitimate their opposition to a dominant point of view. But one can be pretty sure that, if the insurgent group were itself ever to become dominant, talk of pluralism would subside and they would become every bit as monistic as those whom they had replaced. This strategic pluralism has nothing to do with metaphysics or epistemology, and everything to do with professional power and dominance. The case in economics described by Sent (chapter 5 in this volume) may be a case of strategic pluralism.

References

Byrne, A., and D. R. Hilbert. 1997a. *Readings on Color: The Philosophy of Color.* Cambridge, Mass.: MIT Press.

———. 1997b. *Readings on Color: The Science of Color.* Cambridge, Mass.: MIT Press.

Cartwright, N. D. 1983. *How the Laws of Physics Lie.* Oxford: Clarendon Press.

———. 1999. *The Dappled World: A Study of the Boundaries of Science.* Cambridge: Cambridge University Press.

Giere, R. N. 1988. *Explaining Science: A Cognitive Approach.* Chicago: University of Chicago Press.

———. 1999. *Science without Laws.* Chicago: University of Chicago Press.

———. 2006. *Scientific Perspectivism.* Chicago: University of Chicago Press.

Goodman, N. 1978. *Ways of Worldmaking.* Indianapolis: Hackett.

Hacking, I. 1999. *The Social Construction of What?* Cambridge, Mass.: Harvard University Press.
Hurvich, L. M. 1981. *Color Vision*. Sunderland, Mass.: Sinauer Associates.
Kuhn, T. S. 1962. *The Structure of Scientific Revolutions.* 1st ed. Chicago: University of Chicago Press. 2nd ed., 1970.
Morrison, M. 1999. "Models as Autonomous Agents." In *Models as Mediators: Perspectives on Natural and Social Science,* ed. Mary S. Morgan and Margaret Morrison, 38–65. Cambridge: Cambridge University Press.
Teller, P. 2001. "Twilight of the Perfect Model Model." *Erkenntnis* 55: 393–415.
van Fraassen, B. C. 1980. *The Scientific Image*. Oxford: Oxford University Press.
———. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.
———. 2002. *The Empirical Stance*. New Haven, Conn.: Yale University Press.
Weinberg, S. 1992. *Dreams of a Final Theory*. New York: Pantheon Books.
Wolfram, S. 2002. *A New Kind of Science*. Champaign, Ill.: Wolfram Media.

3
Plurality and Complementarity in Quantum Dynamics

The Problem of Plurality

Any discussion of "pluralism" in science immediately faces (at least) two very difficult problems. The first problem is one of definition: what is pluralism; what does pluralism concern; and what, precisely, is one's attitude toward the diversity implied by pluralism? If one is a pluralist about some substantial part of science, then the second problem looms large: *how* can one be a pluralist about science while respecting the (approximate) validity of our best scientific theories?

While other essays in this volume speak to these issues at greater length, it is important for me to indicate my general attitude and approach to pluralism in science, so that what is to follow is understood as intended. Specifically, I shall spell out more carefully what I take the "problem of plurality" to be and my general approach to this problem. I shall discuss some background material for my main topic, covering (in an elementary way) what I take to be the dynamical deficiencies of quantum mechanics as it is normally understood, and then I shall discuss one quite general approach to solving this problem. Finally, I shall discuss the sort of plurality to which this solution gives rise.

Although the *Oxford English Dictionary* provides an explicitly philosophical (though vague and unhelpful) definition for the term "pluralism" ("a theory or system of thought which recognizes more than one ultimate principle"), its political definition is a more apt starting point: "The existence or toleration of diversity of ethnic or cultural groups within a society or state." We need to make only a few modifications to obtain a working definition: *pluralism* is the existence or toleration of a diversity of theories, interpretations, or methodologies within science.

I intend the word "theories" to be taken very liberally, to encompass such things as principles (for example, Le Chatelier's principle in chemistry), models (for example, a model of populations in an ecological system), and presuppositions (for example, a psychologist's presupposition that other human beings are intelligent agents). Similar latitude applies to the term "methodologies," which includes such things as principles of experimental design, methods for determining causal relations, methods for analyzing data, and even techniques for proof in mathematics and theoretical physics. Finally, by "interpretations" I mean both attempts to provide foundations for science, whether mathematical or philosophical (both of which one finds, for example, in contemporary philosophy of quantum theory), and attempts to draw extrascientific conclusions from science (for example, attempts to draw conclusions about theism from evolutionary theory).

"Existence" and "toleration" are independent. One can tolerate an existing plurality, or adopt an attitude to science that would tolerate a plurality, even if plurality does not exist. Conversely, one may insist that an existing plurality must be resolved into unity, or one may revel in the existence of just a single theory, interpretation, or methodology.

Finally, the theories, interpretations, and methodologies can be diverse in a few different ways. Diversity of subject matter leads to probably the weakest form of pluralism, namely, the existence (which is manifest) or toleration of different theories, interpretations, or methodologies in different subject matters of science. More interesting forms of pluralism arise in the context of a single subject matter, and indeed the essays in this volume that deal with specific sciences (including this essay) consider pluralism of this form. Here, I see three broad types of diversity.

The first we might call, with some caution, "anomalous monism," which (in this context)[1] is the view that diverse theories, interpretations, or methodologies can be consistently conjoined into a single theory or methodology, but not in any systematic or lawlike way (so the conjunction is a mere conjunction rather than any sort of union). The view is monistic because, in the face of such diversity, there is still (possibly) a single totally true theory about the world, or a single true interpretation, or a single comprehensive methodology. The view is anomalous (*a-nomos,* from Greek) because the diverse elements of the single truth, interpretation, or methodology can only be conjoined to form the one, and cannot be united into a single theory from which the diverse elements are derived in any lawlike way. Or, putting the matter in a way slightly closer to Davidson's original intention, no one of the diverse theories, interpretations, or methodologies can be reduced to another in a lawlike or systematic way. Their terms, for example, may not be interdefinable.

The second type of diversity we might call, with significant caution and some trepidation, "incommensurability," the idea being that the diverse theories or methodologies are simply incomparable. The difference between incommensurability and anomalous monism is that whereas in the latter case the diverse theories or methodologies can be conjoined, in the former case their conjunction is meaningless because there is no "common language" in which they can be conjoined. They are neither consistent nor contradictory, there being no language in which to make such a determination. An extreme case is the sort of situation countenanced by Carnap, who considered the possibility that the very logic of the languages of distinct theories might simply be incompatible (which is not to say "inconsistent," of course—see Richardson [chapter 1 in this volume] for a discussion of Carnap on pluralism).

The third type of diversity is contradiction, the existence of which is common. Indeed, purportedly (and often in fact) science is in the business of determining which of several competing, that is, mutually exclusive, theories is true. Philosophers are purportedly in the business of evaluating competing interpretations. Methodologies can contradict one another as well; for example, they may make incompatible assumptions about underlying causal structure.

While I do not claim that the taxonomy of pluralisms (in principle, twenty-seven types)[2] arising from the distinctions above is exhaustive, I am not aware of a significant type of pluralism that cannot be seen as an example of one of the types that I have just outlined. At the same time, none of the types of pluralism that I have described should be taken lightly. Any acknowledgment of an existence of diversity—of whatever sort—raises questions about what *actual* science is capable of achieving. Any claim that diversity should be tolerated raises more serious questions still about the nature of science, and in particular about what science ought to try to achieve and how we ought to understand what it has achieved.

My main purpose for making these distinctions is to define more carefully than I might otherwise have done (even if only marginally so) the sort of pluralism that I intend to endorse within the context of quantum theory. Of course, there remains a lot of space within any of the given types of pluralism, but rather than making finer distinctions, I prefer to illustrate this point by the example of the remainder of this essay in which I shall describe my own attitude toward quantum theory as an example of pluralism that acknowledges the existence of, and tolerates, a diversity of contradictory theories. In doing so, I shall address the most obvious and serious objection to such a view (corresponding to the second question in the first paragraph of this essay), namely, that it places the scientifically minded person in the intolerable position of explicitly endorsing contradictions within science (as a matter of principle and not merely as a pragmatic matter). To do so is to reject the scientific enterprise.

Of course, one need not worry about this objection—crude relativists do not, for example—but I do worry about it, because I cannot see how to excise the principle of noncontradiction from science without taking the possibility of rational argument with it. (I claim not that science is in practice always free from contradiction, but that appeal to the principle of noncontradiction is always a legitimate move in a scientific argument.) So I take it as incumbent on anybody who tolerates a diversity of contradictory theories to explain how their toleration does not amount to denying the principle of noncontradiction in science.

Any such explanation seems bound to introduce a mechanism by virtue of which the contradictory theories are never brought into direct comparison. As I have just stated it, this strategy remains at best vague, but I shall introduce one way of making it more precise in the remainder of this essay, where I intend to convince the reader that one can maintain contradictory dynamics in quantum theory without having to face up to the simple assertion of a contradiction.

Of course, I do not suppose that one must approach this brand of pluralism in the particular way that I propose. Indeed, it seems likely that other approaches to this problem could work in other cases. The approach that I outline here is therefore not intended as a generic strategy, but rather as a strategy that is suited perhaps only to the specific case at hand.

Quantum Theory as a Theory of the Moment

A Schematic Picture of the Theory

From a certain point of view (which I believe is the correct point of view), quantum theory as it is frequently presented and practiced is a "theory of the moment." It will become clear precisely what I mean by this phrase soon enough. For now, it means that quantum theory makes explicit predictions not about the dynamical history of a physical system, but only about its state at a given moment in time.

In order to make the claim plausible and clear, I must say what I mean by "quantum theory as it is frequently presented and practiced." I will necessarily be abstract—you will not be able to calculate the energy levels of the hydrogen atom based on what I say—but it does capture what I take to be the (relevant) essentials of quantum theory insofar as physicists and philosophers of physics agree on what those essentials are.

To begin, let us grant ourselves the collection of *observables,* or physical quantities of quantum theory—such things as energy, momentum, and so on. Then the *states* of physical systems are probability measures over the observables. So the state *s* of a system *S* determines, for any observable *O,* the probabilities that *S* has for taking any of the possible values of *O*. A pure state provides maximal information about a system. It is a peculiarity of standard quantum theory that pure states can still assign nontrivial probabilities to values for observables (i.e., probabilities that are not 0 or 1). Indeed, every pure quantum state assigns nontrivial probabilities to at least some observables. (In contrast, in classical mechanics as standardly understood, pure states assign a single definite value to every observable.)

Consider, for example, the total energy of an electron. In general, a quantum state of the electron will assign nontrivial probabilities to the various possible total energies. Some pure states assign just a single definite energy (i.e., probability 1 to a single possible value for energy, and probability 0 to the rest). However, such states necessarily assign nontrivial probabilities to values for other observables of the electron, such as its position. Standard quantum theory is, in this way, inherently probabilistic.

Finally, we would like a way to say how these states (probabilities) change in time, that is, a way to answer questions such as "Given that the electron was at the point *x* at time 0, where will I find it at the later time *t*?" In classical physics one does (in general) have an answer to questions of that form. In quantum theory, though, the question must be rephrased, due to the inherent probabilism of the theory. Rather than asking about the evolution of the values of observables, we must ask about the evolution of states, i.e., probabilities for values of observables: "Given that the state of the system was *s* at time 0, what will its state be at a later time *t*?" Quantum theory provides an answer in the form of an equation of motion for the state of a system. The equation is deterministic—the quantum state evolves deterministically—but note that what is evolving deterministically is not the values of observables but the probabilities. Indeed, even if an observable has a definite value at one time, it need not have a definite value at a later time. The probabilities could be trivial at the earlier time but nontrivial at the later time.

One other concept that will be useful later is that of an expectation value. Quantum states determine expectation values for observables in the usual sense: in the state *s,* the expectation value of an observable *O* is the sum of each of *O*'s possible values multiplied by the probability of that value as assigned by *s*—the (expected) long-run average. To take a trivial (but very common) example, if *O* has two possible values, +1 and –1, and the state assigns probability ½ to each of these values, then the expectation value is 0. As it happens, in the context of quantum theory, everything that can be said in terms of states has an equivalent statement in terms of expectation values, and vice versa. I shall sometimes speak (without further comment) of expectation values rather than of states.

The Nondynamical Nature of the Theory

Quantum theory dictates how probabilities change in time. In this sense, the theory provides a dynamics, but it is a dynamics of statistics. It tells us how the statistics of, or probabilities associated with, a system change over time.

I do not wish to quibble about words, but if forced to choose words—and of course I am forced to choose words—I would say that a prescription for the time-evolution of probabilities is not sufficient to make a theory dynamical in the proper sense. I shall indicate what I mean with an example.

Consider a simple system, such as a radiative atom that is subject to some (time-dependent) probability for decay. We consider a simple observable, D, which at any given time may take one of two possible values: 1 for "decayed" and 0 for "undecayed." If the atom is definitely undecayed, its state, s, should assign the expectation value 0 to D; and if the atom is definitely decayed, then s should assign the expectation value 1 to D. (Note that in this case, the expectation value is also the probability that the atom is decayed.) Suppose that at some initial time, the expectation value of D is 0. Under appropriate (and reasonable, even approximately true) assumptions, the quantum-mechanical law of evolution (of states) will give rise to time-evolved states such that for later times t, and letting $s(t)$ be the state at time t, $s(t)$ assigns probability e^{-Kt} to the value 0 for D, for some appropriate (positive real number) constant K, so that there is an exponentially increasing probability for decay as time goes on.

Of course, in the limit $t \rightarrow \infty$ we have what one would expect: the probability of a decay goes to 1. However, during the intervening times the probability is strictly between 0 and 1. In other words, the theory does not answer the question "Is the atom decayed?" but only the question "What is the probability that the atom has decayed?" The answer to the latter question depends on the time, and thus the theory provides a dynamics for probabilities.

But the theory provides little or no dynamical picture of the state of the atom, if "the state of the atom" is taken to mean "its value for observables like D" (as opposed to the quantum state, i.e., the expectation values for observables like D). Indeed, the theory does not even constrain how many times the atom could have decayed. Nobody *believes* that atoms experience a series of decays and antidecays—indeed, nobody believes that antidecays are physical processes—but nothing in quantum theory as I have described it thus far rules this possibility out.

For example, it is consistent with the constraints imposed by quantum theory that the atom rapidly oscillates between decayed and undecayed, spending more and more time per second in the decayed state as time moves on, so that in essence the expression e^{-Kt} is a kind of time average of the state of the atom. In other words, many histories of the state of the atom itself are compatible with the predictions of quantum theory.

It is worth investigating this point in further detail because one is likely to become skeptical as soon as one considers the point more carefully. For example, you might point out that we could place a detector in front of the atom. Surely it will simply be blank for some period of time, then show a single decay-event, after which no further decay-events occur. And surely quantum theory must predict this fact. Right?

Nobody (of whom I'm aware) doubts that the history of the detector is as described above; nonetheless, the answer to the last question (whether quantum theory predicts this history) is a bit subtle. In order to address the question properly, we must introduce the detector into the system, so that we are now considering a compound system with two component systems—the atom and the detector. (The detector is itself a compound system, but because we are interested only in whether it indicates the presence of a particle, we can treat it as a two-state system. We have already done the same to the atom.) We will make the further simplifying assumption that there is a perfect correlation at all times between the detector and the atom: the detector indicates a decay at time t if and only if the atom is decayed at time t.

Now everything is as it was before: the state for the compound system initially assigns probability 0 to the case where the atom is decayed and the detector indicates a decay, and over time, the probability for the compound (atom-plus-detector) system to be in the state where the atom is decayed and the detector indicates a decay increases as indicated earlier. Moreover, the probability that the atom is decayed and the detector fails to indicate a decay, or vice versa, is always zero: the prescription for time-evolution in quantum theory preserves the perfect correlation between the value that the atom has for D and the indication on the apparatus—more precisely, its value for an indicator-observable I with values 0 and +1 corresponding to indicating no decay and indicating a decay. In other words, the quantum state for the compound system assigns probabilities to joint values for the two observables D and I, and it always assigns probability 0 to joint states in which their values differ.

Therefore, at any given time, the detector indicates a decay if and only if the atom is decayed, and yet it is consistent with the probabilities delivered by quantum theory that the atom fluctuates many times between decayed and not-decayed, and of course the detector fluctuates right along between indicating a decay and not indicating a decay.

"But wait," you are thinking, "can we not set up an apparatus to *count* the number of decays, and in this way mark a real difference between the (realistic) case where the atom decays just once and the (unrealistic) case where it fluctuates haphazardly between the two states? And can't quantum theory describe such an apparatus accurately, at least in principle?"

Yes, we can set up such an apparatus, and yes quantum theory (at least in principle) describes it accurately—that is, quantum theory makes accurate probabilistic predictions about it at any given moment in time. Imagine that the detector makes a distinct mark on a piece of paper each time it detects a decay. Therefore, if the detector ever detected two decays, there would be two marks on the paper, and so on. Then the observable of interest for the detector will indicate the number of decays detected. Let us call the observable N. Its possible values are nonnegative integers.

Under reasonable assumptions, quantum theory can describe the evolution of the entire system. Under this evolution, the probability that there will be more than one mark on the paper at any time is 0, as it should be. In other words, the detector will *never* indicate having detected two decays. And, as we expect, the value of N is always perfectly correlated with the value of D.

And yet, quantum theory still only provides single-time probabilities, which are consistent with the detector's wildly fluctuating between indicating that a single decay has occurred and indicating that no decay has occurred (the fluctuations being perfectly correlated with the atom's fluctuations between being decayed and being undecayed).

It should be clear how this discussion can be extended to any imaginable device for recording the supposed history of values that a system takes for an observable, including whatever devices are involved in human memory. Assuming that your brain, and its memories, can be described quantum-mechanically, you would be described, at every moment, as having remembered exactly zero or one decays, but such a memory at any given time is consistent with any number of actual prior decays (and previous contents of your memory).

All of the above follows immediately from the formalism of quantum theory, together with the fact that quantum theory as it has been described thus far provides only single-time probabilities.[3] One may wonder, however, whether certain central principles of physics—whether quantum or classical—are not in fact violated by the sorts of world history in which atoms haphazardly fluctuate between decayed and undecayed, and so on. To put the point the other way around, one might hope that while the single-time probabilities delivered by quantum theory are insufficient to imply much about the history of a system (as opposed to any *reports* or *memories* of the history that might exist at a single time), other principles of the theory might rule out unphysical sequences of decays and antidecays.

To see how the argument might go, consider how one might try to tell a similar story in the context of classical mechanics. We do not need to make reference to the details of classical physics. Just imagine that the (classical) state of the world at one moment bears no relation whatsoever to its state at any other time, but that at each time the world is *as if* classical physics had been true all along. For example, astronomers' notebooks would show records of planets having moved along elliptical orbits, you and I would remember having seen projectiles following parabolic trajectories, and so on, when in *fact* the history of the world (the actual motions of planets and projectiles) violated the laws of classical motion radically.

Despite Kant's best efforts, such a scenario cannot be ruled out on any a priori grounds that I can imagine (it is essentially a version of the Cartesian dreamworld), but it *is* ruled out by classical physics. A world in which the state at one time bears no lawlike relation to the state at another violates the laws of classical physics, in which earlier states highly constrain, if not fix, later states. We can see the point in a very general way by noting that in the world where past and future are not correlated, there would be, in general, massive violations of a variety of conservation laws, including conservation of momentum and conservation of energy.

So there are two (closely related) reasons to reject the view that haphazard evolution of the world is compatible with classical physics: it would violate the detailed laws of motion of classical physics, and it would violate very general principles, such as conservation principles. While, on the one hand, citing either of these reasons as an argument against the claim that the world evolves haphazardly would be circular if done so in the defense of classical physics, on the other hand, and at least if one begins with an antecedent belief in classical physics, haphazard evolution can be ruled out.

This argument does not work in quantum theory. The quantum theoretic law of evolution is, as we have seen, a law of evolution for probability distributions, and such an evolution is compatible with haphazard evolution of the values that a system takes for observables, provided that the single-time statistics are obeyed—but they can be respected by evolutions in which the past bears almost no correlation to the future. (I shall discuss this point in detail later.)

What about more general principles that might be violated by haphazard evolution? The situation here is much the same: those principles are, in quantum theory, expressed (that is, only derivable and indeed only expressible) in terms of expectation values, or probabilities, rather than in terms of the actual values that systems take for observables.

A consideration of the manner in which quantum theory expresses conservation laws will be sufficient to illustrate the point. The conservation of any quantum-mechanical observable O (for example, energy), is expressible in quantum theory *only* as the time independence of the expectation value of O in every state; but that time independence is compatible with fluctuations in the value of O over time (except in the relatively uncommon case where the probabilities are all 0 or 1 for values of O). In other words, conservation laws in quantum theory have consequences for the statistical predictions generated by the theory, but no further consequences (at least not until we supplement the theory as described here with some other principles that could connect conservation laws to something other than expectation values).

Therefore, quantum theory is completely consistent with a world that evolves in such a way that past values for the observables are not correlated in any interesting way with future values (except whatever correlations might be implied by the single-time probabilities). Such an evolution is consistent with everything that quantum theory as I have described it here tells us about the world. In other words, quantum theory does not, in fact, tell us very much at all about how the values that a system takes for observables change over time.

The absence of a dynamical picture of the world in quantum theory is a problem, although some might not think so. Quantum theory is an empirically successful theory, and those who believe that science is about nothing other than empirical adequacy (narrowly construed) might be tempted to let the matter rest there, simply asserting (as, for example, van Fraassen [1991] seems to have done, at least at one time) that dynamical accounts of the values that systems take for observables are irrelevant to the aims of science. My own view is that we cannot let the matter rest there. Quantum theory needs supplementing with a dynamics. I turn in a moment to the argument for that claim.

But first, I emphasize as strongly as possible (and despite my reference to van Fraassen, above) that I do not view the arguments that I shall adduce as having the slightest thing to do with twentieth-century debates between realists and antirealists. My arguments are neither motivated by any position one might take in those debates, nor intended to imply or suggest a position in those debates. Though there might be some positions that are not compatible with my views, these incompatibilities do not systematically rule out realism or antirealism.[4] I shall not elaborate on the point here, but I do hope that the reader will attempt to consider the arguments in this light.

A Dynamics for Quantum Theory

Motivation

If quantum theory without a dynamics for the values of observables is an empirically adequate theory (and let us suppose that it is), then why do we need a dynamics? My answer here (though I think there is more than one answer) is based primarily on the so-called problem of measurement in quantum theory.

In briefest form, the problem of measurement is this: on the one hand, in typical situations (exemplified by the situation at the end of a purported measurement), the quantum state of a system assigns nontrivial probabilities to observables (such as "the location of the pointer on the measurement device") that evidently have a single definite value; on the other hand, one cannot consistently interpret the probabilities assigned by the quantum state as a measure of ignorance.

The last point is crucial. If we could interpret quantum probabilities as measures of ignorance, then we would say that every observable always has an actual definite value—we just do not (always) know what it is. This interpretation apparently requires that we be able to assign values to every observable simultaneously in a consistent way. The obvious (and some would argue, only reasonable) way to do so leads to a logical contradiction.[5]

So somehow one must assign the values that quantum theory does not. Somehow, one must secure the result that, for example, a pointer at the end of a measurement actually does have a definite position. Most philosophers of physics do not want to secure this result by stipulation; we do not want to say, "Whenever I need an observable to have a definite value, I will stipulate that it does, despite the fact that quantum theory assigns nontrivial probabilities for values of that observable." Instead, we want a rule, or principle, that determines in any given situation which observables have definite values.[6]

Presumably such a rule would help one to recover the familiar facts of our everyday experience, such as the fact that middle-sized solid objects have, at least to a very close approximation, a definite location, and so on. But many of these familiar facts are dynamical. For example, if left undisturbed, the book on your shelf does not move relative to the shelf, planets orbit the sun, and so on. In addition, and perhaps even more important, are the facts to which scientists typically appeal in explanations—the definite-valuedness of various observables, as well as dynamical facts.

As we have seen, one can recover the *appearance* of dynamical facts within a theory that is nondynamical (i.e., assigns only single-time probabilities). However, adopting such a nondynamical theory has two consequences that are hard to swallow. First, it precludes our appeal to dynamical facts in explanations. Second, it is self-defeating, because it requires that we countenance radical error in our memories of perceptions of the world, in which case we apparently have no reason to believe in quantum theory in the first place.[7]

I do not claim that subtle philosophical maneuvers might not provide a way out of this difficulty, but it seems clear that the most straightforward way out is to search for a *dynamics* of the definite values that observables take. That is the way that I prefer.

A Plurality of Dynamics

The lack of completely satisfactory solutions to the quantum-mechanical problem of measurement makes it difficult to make many authoritative claims about the nature of any dynamics that might eventually supplement quantum theory. However, recent work has revealed some general features that just about any dynamics is going to have (or so I claim, though here without argument).[8]

The problem, then, is to provide a dynamics for quantum theory as supplemented by some principle or other that solves the measurement problem. Under a very wide range of such principles, providing a dynamics can be reduced to the problem of providing a dynamics for a single quantum-mechanical observable. I shall consider, for simplicity, the case where that observable is discrete—that is, it has countably many possible values.

Those familiar with the problem—or knowledgeable enough to be able to detect the mathematics behind my words—should note that the reduction of which I speak is purely mathematical. My claim is not that the observable that we use in defining a dynamics will itself always be definite-valued, nor indeed that it bears any easy (or even mathematically tractable) relation to the observable or observables that are definite-valued. The claim is only that defining a dynamics for possessed properties of quantum-mechanical systems can mathematically be reduced to the problem of defining dynamics for some single observable. The reduction is, as they say, nontrivial.[9]

Let us call our observable Q and its possible values q_n where $n = 0, 1, 2, \ldots$. At each time, quantum theory provides a probability measure over the possible values, as determined by the quantum state. Let us write the quantum-mechanical probability that the observable has the value q_n at time t as $p_n(t)$. These $p_n(t)$ are the experimentally verifiable single-time probabilities of quantum theory, and therefore these $p_n(t)$ must be respected by any dynamics.

The most promising route to a complete dynamics is to begin with infinitesimal transition rates, $t_{mn}(t)$, which are, intuitively, the probability that a system will change its value from q_m to q_n over the infinitesimal period from t to $t + dt$.[10] These infinitesimal transition rates are constrained by the single-time probabilities of quantum theory, and the easiest way to connect the two is by means of a probability current.

One can conceive of the probability current quite literally and likely (in the present context) not go wrong; so imagine that the q_n are boxes and that probability is flowing from one box to another. Then there is a current of probability that can be defined between any two boxes q_m and q_n, which is just the amount of probability (per unit time) that passes from q_m to q_n minus the amount of probability (per unit time) that passes from q_n to q_m. Call this current $j_{mn}(t)$—the current from q_m to q_n. It depends on time because at different times more or less probability may be flowing between boxes. In particular, the current depends on the (time-dependent) transition rates and the (time-dependent) single-time probabilities:

$$j_{mn}(t) = \mathrm{t}_{mn}(t)p_m(t) - \mathrm{t}_{nm}(t)p_n(t). \tag{1}$$

This equation just says in mathematics what I said earlier when characterizing the current. Note that it implies

$$j_{mn}(t) = -j_{nm}(t) \tag{2}$$

as it should. Finally, the time-derivative of the single-time probabilities (which is fixed by the quantum-mechanical equation for the time-evolution of the state) is related to the current by the continuity equation:

$$\dot{\mathrm{p}}_n(t) = \sum_m \mathrm{j}_{mn}(t). \tag{3}$$

This equation says that the infinitesimal change in p_n over the time t to t + dt is the sum of all the currents into and out of q_n at time t.

Given the above, we have reduced the problem of finding a dynamics for Q that is consistent with the empirical predictions of quantum theory to the following problem:

- Find a current that satisfies (2) (any antisymmetric matrix will do) and (3).
- Solve (1) for the $\mathrm{t}_{mn}(t)$ (under the additional constraint that $\mathrm{t}_{mn} \geq 0$).
- Recover the finite-time transition probabilities, essentially by integrating the $t_{mn}(t)$.

To see how it goes, imagine that we have some current in hand that satisfies (2) and (3). From (1),

$$t_{mn}(t) = \frac{t_{mn}(t)p_m(t) - j_{mn}(t)}{p_n(t)} \tag{4}$$

The numerator of the right-hand side must be positive (because the $p_n(t)$ are always positive), and so we must have

$$t_{mn}(t) \le \frac{j_{mn}(t)}{p_m(t)} \tag{5}$$

and therefore

$$t_{mn}(t) \le \max\left\{0, \frac{j_{mn}(t)}{p_m(t)}\right\} \tag{6}$$

The most natural choice for a solution to equation (3) seems therefore to be the following: for $m < n$ choose

$$t_{mn}(t) = \max\left\{0, \frac{j_{mn}(t)}{p_m(t)}\right\} \tag{7}$$

One easily checks with (4) that the same solution will then hold for all $m \neq n$, so that t_{mn} is given by (7) for all $m \neq n$.[11]

The Mathematics of Plurality

Plurality enters the picture that I have sketched in two places. First, although I chose (7) as perhaps the most natural solution satisfying (6), there are clearly other solutions—the constraints in (6) vastly underdetermine the $t_{mn}(t)$. In addition, the current is highly underdetermined. There is a standard expression for the current in quantum theory, and it is normally derived from the equation of motion. But one cannot *really* derive the current from the quantum equation of motion. This point is clear from the fact that equations (2) and (3) are the *only* empirical constraints on a current, and they do not determine a current uniquely. To see why, let $j_{mn}(t)$ be some solution to (2) and (3). Then define

$$\tilde{j}_{mn}(t) = j_{mn}(t) + k_{mn}(t)$$

where $k_{mn}(t)$ is any set of functions (of time—or if you prefer, think of it as a time-dependent matrix) satisfying (2) and

$$\sum_m k_{mn} = 0. \tag{8}$$

One can see immediately that because the original solution $j_{mn}(t)$ satisfies (3), the new solution $\tilde{j}_{mn}(t)$ will too, by virtue of (8).[12]

There is an intuitive way to understand the underdetermination in both the $t_{mn}(t)$ and the $j_{mn}(t)$. Imagine, again, that the q_m are boxes, and let the probability associated with a box be represented by a number of balls in the box (the more balls, the higher the probability). (In this analogy, boxes are to be associated with the possible values of our observable Q, one box for each possible value.) Over time, the balls are shifted from one box to another. Our theory fixes, at each time, how many balls are in each box. Strictly speaking, such a theory predicts everything we could ever measure explicitly—the only observation we can make is to examine the number of balls in the boxes at any given time.

A very little imagination should make it clear that the balls could shift around in more than one way and still maintain the correct number of balls in each box at a given moment. As Figure 3.1 illustrates, if a given box's count (the number of balls in the box) goes down (between the earlier and later times), it could deliver its excess balls to more than one place—any box whose count goes up could receive one or more balls from our box. Nothing about the single-time probabilities dictates which of these transitions is correct. The freedom to choose illustrates freedom in the choice of the $j_{mn}(t)$.

Moreover, boxes can exchange balls "unnecessarily." For example, although neither box A nor box B "needs" to give up a ball to maintain the correct single-time probabilities (i.e., the correct number of balls) in the previous example, they could nonetheless exchange a ball, or not. (Note that the current is the same in both cases.) The freedom to choose whether they do exchange a ball illustrates freedom in the choice of the $t_{mn}(t)$.

It should be clear that just by combining these two sorts of freedom, one can easily generate a very large number of possibilities for transition probabilities that underwrite any given single-time probabilities. (In this extremely simple example, there are 384 possibilities.) Indeed, when we are dealing with continuous exchange of probability rather than a discrete exchange of balls, the possibilities are uncountably infinite.

Understanding the Plurality

Principles Leading to Alternative Dynamics

How do we determine which is "right"? I suggest that there is no unique answer to the question. It does not follow, of course, that anything goes.

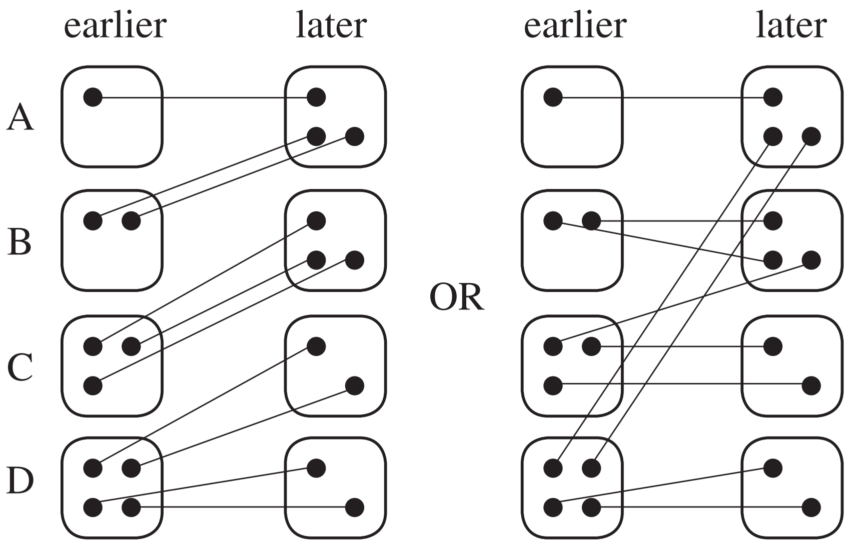

Figure 3.1. The boxes A, B, C, and D start with 1, 2, 3, and 4 balls respectively, and end with 3, 3, 2, and 2 balls. The number of balls represents the amount of probability in the box at the earlier and later times. The boxes on the left show one way that the balls can move among the boxes between the earlier and later times while respecting the single-time probabilities (i.e., the number of balls in the boxes) at each time. On the right is another way, which differs from the first but still respects the same single-time probabilities. Your imagination can produce many other alternatives.

While I do not have preconceived ideas about what does go, I can indicate in general both why a multiplicity of dynamics is not necessarily a bad thing and why this multiplicity is also not so far-reaching as to permit just any dynamical scheme. This subsection considers the first point, and the next subsection considers the second point.

My suspicion is that scientifically minded readers are likely to be very skeptical of the idea that multiple incompatible dynamical schemes may be in some significant sense equally legitimate. If the point of adopting some dynamics is to assert its literal and unqualified truth, then clearly there is a problem allowing multiple dynamics—doing so is a logical contradiction. The point changes little if we adopt a somewhat more antirealist, though still literal and unqualified, way of understanding what it means to adopt a dynamics. If we wish to maintain the simultaneous legitimacy of alternative dynamical schemes, then our options are limited; I see only three.

First, one can assert that $t_{mn}(t)$ does not actually mean what it apparently says. This option is not particularly plausible. (It was tried, in a variety of ways, by some logical positivists.) I shall not consider it further.

Second, one can assert that in fact only a single set of transition rates can ever be asserted. Pursuing this option would require the defense of some principle that rules out all dynamical schemes except one. I am very skeptical that any plausible such proposition can be found. It will not be a purely experimental fact—we have already seen that the experimental facts are consistent with just about any dynamical scheme. It will instead be some other more general principle. But, as I shall discuss later, there are various candidates for such general principles, all equally worthy of attention but giving rise to incompatible dynamical schemes.

I believe we are left with this third option: one can assert that the $t_{mn}(t)$ are, in effect, always indexed by, or conditioned on, something else. My discussion of this option will also illustrate why I think the second option is not attractive.

The basic point is this: in some contexts, it makes sense to adopt one set of transition rates, and in other contexts, it makes sense to adopt others. I find it helpful to characterize these contexts as explanatory contexts, but I am not strongly tied to this characterization. The idea is that a context is characterized by a request for an explanation, which carries with it presuppositions about the explanandum and the explanans.[13] Some such presuppositions are satisfied by some transition rates (or more likely, some families of transition rates), and others are satisfied by other (families of) transition rates. When the families of transition rates satisfying two different explanatory contexts (requests for explanation) do not overlap, then I shall call the contexts incompatible.

Are there important incompatible contexts? I believe there are. Of course, the word "important" is perhaps a bit of a fudge term, but it is important here because there is no doubt that, barring any further constraints, one can specify incompatible contexts: just explicitly demand an explanans that appeals to one or another of two nonoverlapping families of transition rates. By calling an explanatory context important, I mean to suggest, very roughly, that they are scientifically interesting, that questions asked in those different contexts are questions that scientists ought to feel some pressure, or at least desire, to answer, within the boundaries of science.

An example will help to illustrate what I mean. It is well known that in one sense, quantum theory is compatible with Einstein's theory of relativity, and in another sense, their compatibility is suspect. In particular, nothing about the empirical predictions of the quantum theory violates the empirical predictions of relativity. At the same time, the rules that are normally added to quantum theory in order to solve the measurement problem—in particular, the collapse postulate—generally do violate the principles of relativity. (In particular, they violate the principle that there be no preferred inertial frame of reference—not in a way that leads to the possibility of determining the preferred reference frame, but a preferred reference frame exists nonetheless.)

There is very strong evidence that nearly all dynamical schemes of the sort that I have been describing will also violate the principles of relativity theory. This evidence comes primarily in the form of a theorem that says, roughly, that any *stable* dynamics requires a preferred frame of reference (Dickson and Clifton 1998). The condition of stability is essentially a generalized version of the law of inertia: a system that is not interacting (not exchanging energy) with any other system will evolve freely. (This condition is closely related to conservation of momentum and energy.) This principle is supposed to help to secure the sort of everyday dynamical facts that I mentioned earlier, such as the fact that undisturbed books do not jump off shelves.

Both the principle of relativity and the principle of stability have much to recommend them. The principle of relativity is presupposed by many of our explanations of physical matters of fact. Indeed, it is presupposed by many explanations that occur in the context of quantum theory. We are not speaking of one theory respecting the principles of another. There is such a thing as relativistic quantum theory, and the principle of relativity plays a crucial role in that theory, giving rise to the prediction of such significant facts as the existence of antiparticles (for example, the positron). Therefore, there are good theoretical and empirical reasons to adopt the principle of relativity. When those theoretical considerations and empirical facts are important, so is the principle of relativity. When we seek an answer to why-questions in a context where these theoretical considerations and empirical facts must be explicitly acknowledged, we cannot easily give up relativity, on pain of incoherence. (For example, it could be that it makes little sense to ask questions about antiparticles in a context where one is prepared to give up relativity.)

But stability has equally impressive credentials, both theoretically and empirically. We do not, in fact, witness books flying willy-nilly off bookshelves. Nor do we, in theoretical contexts, normally countenance violations of the conservation of momentum and energy. When we seek an answer to why-questions in a context where such empirical facts and theoretical considerations must be explicitly acknowledged, we cannot easily give up stability.[14]

Why are we not in a pickle? Why do we not conclude that physics as we now have it is in a mess, unable to reconcile its interpretational needs (as I see it, the need for a dynamics) with its theoretical principles (for example, relativity and stability)? I certainly do not deny the possibility of a successor theory (or even an acceptable interpretation of quantum theory more clever than I can imagine) in which relativity and stability are reconciled.

However, I do not believe that physics is in need of such a theory in order for it to be satisfactory.[15] I believe that physics can live with complementary dynamical principles.

I use the term "complementary" purposefully, to remind us that in the context of quantum theory the word is often applied to observables that cannot be simultaneously definite-valued. Physics has learned to live with this fact, and at times even thrives on it, to the point where it is becoming better and better understood when to describe a physical situation in terms of one or the other of two complementary concepts (observables). I am suggesting that complementary dynamical principles could, in principle, achieve the same status.

It is crucial for my point to understand that the complementarity of the dynamical principles of which I speak does not extend to the level of empirical predictions, that is, to the level of the dynamics of the quantum state. A magical fact about quantum theory is that it manages (again, at the level of the dynamics for the quantum state) to combine these principles into a single dynamical scheme, as in classical physics. The plurality—that is, the existence of complementary dynamical schemes—becomes apparent only when we insist on a dynamics for the properties of individual systems (rather than the quantum state, which can be understood as a dynamics for the probabilities of a system).

The Limits of Plurality

There are several ways in which the type of plurality that I have been discussing is constrained. I am very far from suggesting that anything goes in the selection of a dynamical scheme in the interpretation of quantum theory (though I do think that there are more dynamical principles than the two I have discussed here). I conclude, therefore, with three warnings, each intended to convey the point that *not* anything goes.

First, some explanatory contexts are probably simply illegitimate. I don't have a general principle here, but not every why-question is worth exploring. I chose the examples of relativity and stability because each of them is clearly important. In fact, while there are other principles relevant to the selection of a dynamical scheme, in general it is difficult to connect acknowledged significant physical principles with the selection of a dynamical scheme.

Second, along the same lines, some why-questions can be ruled out. If there is no dynamical scheme that respects the why-question's presuppositions, then so much the worse for the question.

Third, nothing that I have said suggests that one can introduce dynamical schemes or principles that violate the single-time probabilities of quantum theory. These are the quantities that can be unambiguously measured in the laboratory, and so these quantities must be respected. Dynamical schemes are introduced not for purely empirical reasons, but for interpretational, or ultimately, explanatory reasons. Indeed, it is precisely because they are introduced for these reasons (rather than purely empirical reasons) that multiple dynamical schemes are both allowed and welcomed.

Notes

1. The term "anomalous monism" is lifted from Davidson's (1970) view about the relationship between the mind and the brain, but too much should not be made of this fact.

2. One can believe in anomalous monism, incommensurability, or contradiction for theories, interpretations, or methodologies. In each case, one can believe that a diversity exists, but not tolerate it; that a diversity does not exist, but tolerate one if it did; or that a diversity exists, and tolerate it. Of course, one could also combine types of pluralism (for example, to be an anomalous monist about theories but an incommensurablist about interpretations) to make some very large number of compound types that I have not bothered to count.

3. Readers who are familiar with quantum theory may have other objections to this view. For example, several people have suggested to me that the consistent (or "decoherent") histories approach to quantum theory provides a quantum-theoretical account of histories and not just single-time probabilities. My answer is that many approaches to quantum theory do so, but all of them (including the consistent histories approach) do so by means of introducing interpretive principles or assumptions that go well beyond standard quantum theory. I believe that one *should* go beyond standard quantum theory in a way that introduces histories of some sort, precisely because I believe that we *should* be unsatisfied with quantum theory as a theory of the moment. But it is important to recognize that bare empirical adequacy does not require us to go beyond standard quantum theory, and that doing so involves the addition of interpretive principles to the theory.

4. My interest here in the vast majority of debates between realists and antirealists is effectively zero. I do not mean to dismiss them or those who engage in them; rather, my intent is explicitly *not* to take a position in these debates.

5. The contradiction comes in the form of a theorem known as the Kochen-Specker theorem. The general point is discussed in detail in many places. Dickson (1998b) discusses it, along with various alternatives to the line of reasoning that I am adopting here.

6. An excellent discussion of these matters, advocating this approach in general terms, is given by Bub (1997).

7. A similar point in only a slightly different context was made by Barrett (1996).

8. By "recent work" I have in mind a number of lines of thought in contemporary philosophy of physics, but a good place to start is with Bub 1997.

9. A more or less exhaustive treatment of these issues can be found in Bacciagaluppi and Dickson 1999; fanatics can also consult the references therein.

10. More rigorously,

$$
\begin{aligned}
t_{mn}(t) &= \lim_{\varepsilon\to 0} \frac{p_{mn}(t+\varepsilon, t) - p_{mn}(t,t)}{\varepsilon} \\
&= \lim_{\varepsilon\to 0} \frac{p_{mn}(t+\varepsilon, t)}{\varepsilon},
\end{aligned}
$$

where $p_{mn}(t)$ is the probability that a system in the state q_m at time s will be in the state q_n at time t. A series of theorems from the mid-twentieth century lays out the conditions that the $t_{mn}(t)$ must satisfy in order for one to be able to define the finite-time transition probabilities $p_{mn}(t,s)$ from them. I am ignoring such subtleties here. See Bacciagaluppi and Dickson 1999 for details.

11. This solution is a (very) slightly improved version of the choice made by Bell (1984). Bell's choice is explicitly motivated by the guidance condition in the de Broglie-Bohm theory. Vink (1993) discusses how, in the appropriate sense, the de Broglie-Bohm theory is in fact the continuum limit of a dynamics of this kind, and how different solutions to (4) lead to different kinds of theory, one example being Nelson's (1985) stochastic mechanics.

12. One might suppose that the situation changes when we go to the perhaps more reasonable case of observables with many continuously possible values. It does not. The continuous analogue of (3) is

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\rho(x,t) = \nabla \cdot \mathrm{j}(x,t)$$

where r is the probability density. This expression makes it clear that by adding a divergence-free function to any solution $\mathrm{j}(x,t)$ we obtain another solution.

13. Dickson (1998a) follows through this approach in a little more detail.

14. The fact that relativity and stability both have an empirical side does not imply that they are empirically inconsistent. Strictly speaking, it is compatible with everything we know empirically that relativity is false. The same holds for stability. The point is that the consideration of certain empirical phenomena strongly suggests that we take relativity seriously, while other empirical phenomena do not seem to press the principle on us so strongly. The same holds for stability.

15. I do believe that physics is in need of a successor theory, but for other reasons, reasons with which the vast majority of physicists would agree, I suspect.

References

Bacciagaluppi, G., and M. Dickson. 1999. "Dynamics for Modal Interpretations." *Foundations of Physics* 29: 1165–201.

Barrett, J. 1996. "Empirical Adequacy and the Availability of Reliable Records in Quantum Mechanics." *Philosophy of Science* 63: 49–64.

Bell, J. 1984. "Beables for Quantum Field Theory." In *Speakable and Unspeakable in Quantum Mechanics,* 173–80. Cambridge: Cambridge University Press.

Bub, J. 1997. *Interpreting the Quantum World*. Cambridge: Cambridge University Press.

Davidson, D. 1970. "Mental Events." In *Experience and Theory,* ed. L. Foster and J. Swanson, 79–101. Amherst: University of Massachusetts Press.

Dickson, M. 1998a. "On the Plurality of Dynamics." In *Quantum Measurement: Beyond Paradox,* ed. Richard A. Healey and Geoffrey Hellman, 160–82. Minnesota Studies in Philosophy of Science, vol. XVII. Minneapolis: University of Minnesota Press.

———. 1998b. *Quantum Chance and Non-Locality.* Cambridge: Cambridge University Press.

Dickson, M., and R. Clifton. 1998. "Lorentz-Invariance in the Modal Interpretation." In *The Modal Interpretation of Quantum Mechanics,* ed. Dennis Dieks and P. E. Vermas, 9–47. Dordrecht: Kluwer Academic.

Nelson, E. 1985. *Quantum Fluctuations*. Princeton, N.J.: Princeton University Press.
van Fraassen, B. C. 1991. *Quantum Mechanics: An Empiricist View*. Oxford: Oxford University Press.
Vink, J. 1993. "Quantum Mechanics in Terms of Discrete Beables." *Physical Review A* 48: 1808–18.

Geoffrey Hellman and John L. Bell

4
Pluralism and the Foundations of Mathematics

Contrary to the popular (mis)conception of mathematics as a cut-and-dried body of universally agreed-on truths and methods, as soon as one examines the foundations of mathematics, one encounters divergences of viewpoint and failures of communication that can easily remind one of religious, schismatic controversy. While there is indeed universal agreement on a substantial body of mathematical results, and while classical methods overwhelmingly dominate actual practice, as soon as one asks questions concerning fundamentals—such as "What is mathematics about?" "What makes mathematical truths true?" "What axioms can we accept as unproblematic?" and notoriously, even "What are the acceptable *logical rules* by which mathematical proofs can proceed?"—we find we have entered a minefield of contentiousness. Platonists treat mathematics as an objective study of abstract reality, no more created by human thought than the galaxies, and, accordingly, classical logic and a rich theory of the transfinite are entirely legitimate.[1] Radical constructivists (intuitionists) challenge even the meaningfulness of classical, objectivist thinking in connection with the infinite, and propose a reconstructed mathematics with restricted logic (e.g., no existence proofs by reductio ad absurdum) and different axioms (e.g., the least upper-bound principle is jettisoned). Classicists respond (if they respond at all, which is unusual) by accusing their critics of changing the subject. And between and beyond these camps there is a significant variety of positions or "schools," e.g., predicativism, or "semi-constructivism," which accepts classical logic but only those infinite sets we can actually describe in an acceptable way (which can be spelled out precisely); constructivism of the Bishop school, which, in contrast with intuitionism, adds no new, nonclassical mathematical axioms; constructivism of the Russian school, which lives with Church's thesis identifying constructive functions with the Turing-computable ones; strict finitism; and so on (see Beeson 1985). A plurality or multiplicity of approaches to central questions of *truth* and *proof* is simply an observable fact. What is the nature and significance of this multiplicity? Is it reasonable to think it can be transcended, or is it a permanent fact of life? What lessons, if any, does it hold for general questions concerning pluralism?

A second locus of pluralism in mathematics is ontology. Even within the classical framework, one may ask whether there is a single, all-embracing universe of discourse for mathematics, as set-theoretic reductionism on a customary reading would have it, that is to say, *the* cumulative hierarchy of sets, or should we think of a plurality of universes? Although ordinary mathematics—all that is required in typical graduate programs in the subject—can indeed be developed within set theory, specifically in the favored system known as ZF (Zermelo-Fraenkel), when one considers set theory itself (a branch of "extraordinary mathematics"), one in fact encounters a multiplicity of theories. Usually the Axiom of Choice is added (giving ZFC), but we know that its negation is a consistent option (relative to the consistency of ZF itself). We also know that we need not insist on well-foundedness (sets can be allowed to contain themselves; there can be infinitely descending membership chains). And then there is the whole subject of large cardinal extensions of ZFC, many of which are very natural from a set-theoretic standpoint but that cannot even be proved relatively consistent (a phenomenon known as "Gödel's curse"). Does it make sense to think of unique, determinate answers to all such questions, as talk of the "cumulative hierarchy" implies? Or should we rather think of "many worlds"? Furthermore, there is a different foundational approach with claims to universality, namely, *category theory,* more specifically *topos theory,* which generalizes on set theory in certain ways. Originating in algebraic geometry, toposes are categories in which certain key set-theoretic operations are generalized, notably, the formation of Cartesian products, function classes by exponentiation, and extensions of predicates. They have been proposed as universes of discourse for mathematics, introducing even more options. Thus topos relativity (unlike set-theoretic relativity with regard to large cardinals, for example) prima facie flies in the face of ordinary talk of "the real numbers," "the complex numbers," "the continuous functions of reals," and so on, where uniqueness is presupposed. This suggests a structuralist (re)interpretation of such talk, and even of set theory itself (better, set theories themselves), contrary to the single, fixed-universe view.

Let us elaborate on these two main topics in turn.

Constructivism versus (?) Classicism

The various forms of constructivism (apart from predicativism) have at their common core adherence to intuitionistic logic, usually described as "rejecting the law of excluded middle (LEM)," in the sense of not allowing it in proofs (*not* in the sense of affirming its negation—in fact, the double negation of LEM is a theorem of intuitionistic [propositional] logic). Similarly, allied principles are rejected, such as the law of double negation, proof of existence by reductio ad absurdum, and so on. Formally, intuitionistic logic appears simply to be a *proper part* of classical logic; if you restore LEM to the intuitionistic rules, you recover classical logic. So formally there is no inconsistency between the two.[2] But intuitionists are famous for holding that LEM and allied principles are "not correct"; it seems that we have a genuine disagreement over certain laws of logic! Is that really so? Can even pluralists tolerate such a disagreement? Is even propositional logic up for grabs? Whatever one thinks about the analytic/synthetic distinction in general, don't the (truth-functional) meanings of 'or' and 'not' guarantee that, within the intended domain of determinate propositions such as those of arithmetic, LEM has to be correct?

Indeed, if one looks more closely at intuitionistic usage—even as its proponents have explained it—it is abundantly clear that the key logical words are being used with very different meanings from the classical ones. The very idea of giving *truth-conditions* for logically complex statements is abandoned in favor of *proof-conditions* in which one explains when a (mathematical) *construction* counts as a *proof* of a complex statement. So, for example, intuitionistic 'or' is explained by a condition such as

> *c proves 'p or q' iff c proves p or c proves q.*

(Here the 'or' on the right is supposed to be neutral or pretheoretic, somehow shared by all parties.) The conditional is explained by

> *c proves '$p \rightarrow q$' iff c is an operation on constructions tranforming any proof of p into a proof of q.*

And intuitionistic *negation* is then explained via

> *c proves '$\neg p$' iff c proves '$p \rightarrow 0 = 1$'.*

(Here "0 = 1" may be replaced by any other absurdity.) Given these meanings, no classicist would wish to affirm *'p or $\neg p$'* as a general logical principle, for, when spelled out, it asserts that every (mathematical) proposition is decidable! Similarly, considering that existential quantification is explained as a generalization of *'or,'* so that a proof of '$\exists x\varphi$' provides a method of *finding* an instance together with a (constructive) proof that it satisfies φ, no classicist would apply the method of *reductio* to establish such "existence." Clearly, to avoid confusion, all the connectives should carry subscripts indicating "intuitionistic" or "classical" readings. And then, we have not a single-law "LEM" but two radically distinct ones, the intuitionistic instance of which no one accepts, and similarly for the allied laws. So the controversy seems not to be over the correctness of any logical laws after all!

At this stage, constructivist positions split apart. The mere decision to eschew certain classical forms of proof can be made for a variety of reasons and does not by itself indicate any genuine disagreement with classical mathematics. Specifically, we must distinguish a *radical constructivist* view, which insists that mathematical reasoning must be intuitionistic and that classical reasoning is illegitimate or incoherent (views expressed in different ways by intuitionists from Brouwer to Dummett), from a *liberal* view, which, without challenging the meaningfulness or correctness of nonconstructive classical mathematics, prefers to pursue constructive mathematics for its own intrinsic interest and virtues. So here we have a stark contrast, within constructivism, between *hegemonists* and *pluralists.*

The hegemonist position, as Dummett (1977) has articulated it, rests on a verificationist view of meaning. Platonist or realist truth conditions pertaining to the infinite are in general incommunicable, as terminating procedures for testing them are not available. Rather than taking this (in addition to all the criticisms of verificationism developed by Quine, Sellars, Smart, Putnam, and others over many decades) as indicating a deficiency in the view of meaning, the hegemonist view leads to an extreme stance that Shapiro (1997, 6, passim) has dubbed "philosophy first," namely, that of rejecting mathematics itself for philosophical reasons. David Lewis's reaction (originally to a version of nominalism, but equally applicable here) is germane:

> I'm moved to laughter at the thought of how *presumptuous* it would be to reject mathematics for philosophical reasons. How would *you* like the job of telling the mathematicians that they must change their ways. . . . Can you tell them, with a straight face, to follow philosophical argument wherever it may lead? If they challenge your credentials, will you boast of philosophy's other great discoveries: that motion is impossible, that a Being than which no greater can be conceived cannot be conceived not to exist, that it is unthinkable that anything exists outside the mind, that time is unreal, that no theory has ever been made at all probable by evidence (but on the other hand that an empirically ideal theory cannot possibly be false), that it is a wide-open scientific question whether anyone has ever believed anything, and so on, and on, *ad nauseam?*
>
> Not me! (Lewis 1991, 59; italics in original)[3]

That Dummett's reasoning can also be invoked to challenge the determinateness of, for example, ordinary claims about the past (e.g., four years ago to this day, there were exactly twenty-seven paper clips on my desk) has not been a deterrent.

Thus far, we have seen that the dispute between radical constructivism and classicism is not really over logical laws per se; rather, it is over the meaningfulness of talk presupposing truth-determinate sentences or propositions of infinitistic mathematics. The classical logical connectives and quantifiers, however intelligible they may be in other contexts, are alleged to be unintelligible here in mathematics (except in its constructive part, on which the classical and intuitionistic theorems coincide), ironically the very domain in which the idealization of genuine bivalence built into classical logic has its clearest illustration, and for which it was originally developed!

Fortunately, not all constructivists are radicals. If you want to keep track of computational content in mathematics, requiring reasoning to obey intuitionistic logic makes eminent sense. It is an excellent bookkeeping device. So long as your starting points are constructively justifiable, your conclusions will also be. But if you try to recover standard mathematics along such lines, you will encounter many problems. As soon as you come to the real numbers (as convergent rational sequences), for example, you will realize that you cannot assert that they are totally linearly ordered. You will have to make do (and often can) with a weaker condition: if you know that $x < y$, then you will also be able to show, for any z, that either $x < z$ or $z < y$. You will also not be able to prove fundamental facts, such as the intermediate-value theorem (that every continuous function on [0,1] negative at 0 and positive at 1 has a 0 for some x, $0 < x < 1$), but you will be able to prove something very close to that by tinkering with the statement, strengthening the hypothesis of the theorem or weakening the conclusion (getting within ε of 0). Indeed, Errett Bishop (1967) took constructive analysis far beyond anything previously thought possible by the persistent and clever use of such methods, conquering even such apparently nonconstructive territory as measure theory. It is nontrivial to find genuine examples of scientifically applicable mathematics that cannot be recovered constructively in this sense, although there do appear to be some limitations.[4]

An important lesson we can learn from all this is that there are, indeed—as Carnap recognized through his principle of tolerance—*multiple logics,* legitimate for their own purposes. The notion of "*the* correct logic" is simply a mistake, one which fails to take account of the purpose-relativity and language-relativity of logic. Classical logic is designed for truth-preservation in an idealized setting in which we are dealing with bivalent propositions. The classical connectives (and quantifiers) are introduced as idealizations or simplifications of ordinary language expressions with simply statable, bivalent truth conditions, and classical logical principles and rules pertain to reasoning with *these* connectives. If a rule is *sound* (truth preserving), that is sufficient justification for it, regardless of (lack of) constructivity. But whether classical systems are applicable in a given domain or context is not a matter of logic, but a matter of usage and goals. *Logic does not proclaim its own applicability to particular situations.* Independently, however, it is clear that classical reasoning *is* especially useful in scientific, as well as purely mathematical, contexts in which we are interested in what holds or would hold in a certain situation or model, given certain assumptions, as an objective matter regardless of computability. Then there can be no objection to use of $\mathrm{LEM}_{classical}$ or $reductio_{classical}$, and indeed, forswearing their use would seem like tying a hand behind one's back. However, if computability or constructivity is our goal, then obviously it will not be achieved unless we modify our rules, and we may even introduce a new language (also rooted in ordinary language), as intuitionism does. For these new connectives, some but not all of the classical *forms* will be correct. Intuitionistic formal systems codify correct forms of reasoning from this standpoint, and no one can quarrel with that.[5] Classicists as well as constructivists can see all of this. Moreover, as both purposes—truth-preservation *simpliciter* and constructive interpretability—are worthy and important, we should certainly have peaceful coexistence and even cooperation.[6]

This brings us to a second main lesson. Mathematics as practiced is clearly very rich and diverse in its content and in the interests and purposes it supports. As just indicated, both classical and constructive purposes are encompassed; moreover, often they may be intertwined and not neatly separated by branch or subfield. The situation was well summed up by Feferman over twenty years ago:

> Since neither the realist nor constructivist point of view encompasses the other, there cannot be any present claim to a *universal foundation* for mathematics, unless one takes the line of rejecting all that lies outside the favored scheme. Indeed, *multiple foundations* in this sense may be necessary, in analogy to the use of both wave and particle conceptions in physics. Moreover, it is conceivable that still other kinds of theories [of operations and collections] will be developed as a result of further experience and reflection. (Feferman 1977, 151; italics in original)

This accords with the general hypothesis that the complexity and richness of scientific subject matter and practice may actually *require* a pluralistic approach, that any single one that we have contrived, or perhaps can contrive, will simply not do justice to an important aspect of the subject. The classicism-constructivism duality in mathematics is, we submit, an excellent illustration.

Many Worlds

At the end of a landmark paper, credited with the discovery of large cardinals in set theory, Zermelo wrote of "two polar opposite tendencies of the thinking mind, the idea of *creative progress* and that of *all-embracing completeness*" (italics in original). These, he continued:

> find their symbolic expression and resolution in the concept of the well-ordered transfinite number-series, whose unrestricted progress comes to no real conclusion, but only to relative stopping points, the "boundary numbers" [inaccessible cardinals] that divide the lower from the higher models. And so the "antinomies" of set theory, properly understood, lead not to a restriction and mutilation, but rather to a further, as yet unsurveyable, unfolding and enrichment of mathematical science. (Zermelo 1930, 47, trans. the author)

The central problem calling forth these "two polar opposite tendencies," in a nutshell, is this: over what totality do the unrestricted quantifiers of set theory range? We know that, on pain of contradiction, it cannot be taken to be a set, but if we take it as a collection of some higher type, we face the conundrum that we can apply set-like operations to it, leading to collections of higher and higher type, behaving just like sets, so that our effort to speak of absolutely all sets seems indistinguishable from speaking of all sets below a certain inaccessible level (one of Zermelo's "boundary numbers").[7] Indeed, whatever totality of collections we recognize—whatever we call it—can be properly extended, indeed, by the very operations that gave rise to set theory in the first place (forming singletons, power sets, etc.) The standard set-theoretic "way out" of remaining within a first-order language, officially recognizing *no* totality of all sets, while consistent and useful for mathematics in practice, does not really solve the problem, for the very *possibility* of considering new totalities and proper extensions is intrinsic to mathematics. As Mac Lane has put it:

> Understanding Mathematical operations leads repeatedly to the formation of totalities: the collection of all prime numbers, the set of all points on an ellipse . . . the set of all subsets of a set . . . , or the category of all topological spaces. There are no upper limits; it is useful to consider the "universe" of all sets (as a class) or the category *Cat* of all small categories as well as CAT, the category of all big categories. After each careful delimitation, bigger totalities appear. No set theory and no category theory can encompass them all—and they are needed to grasp what Mathematics does. (1986, 390; italics in original)

Still, we do seem to have unrestricted quantifiers in our language, allowing us to speak of *anything* and *everything*. But then we should be able to speak of *anything* and *everything mathematical,* among which would be all collections or set-like objects, totalities which would violate the general principle of *extendability* articulated by Mac Lane (generalizing Zermelo's own versions, independently arrived at also by Putnam).

It counts as a strike against set-theoretic foundations that it seems to be incapable of resolving this problem. Zermelo's resolution of recognizing an unending, ascending series of models of set theory, each of greater and greater ordinal characteristic (strongly inaccessible cardinal), is a major advance over the fixed universe view, but, as already indicated, it is only a partial resolution, for we still seem capable of speaking without contradiction of "all inaccessible cardinals," or "all full models of ZFC (characterized by Zermelo)," and so on, leading right back to our puzzle. (Indeed, if we formalize Zermelo's logic, which would be a fragment of second-order logic, the standard comprehension scheme leads to classes of all sets, all ordinals, all inaccessibles, all models, etc., after all, conflicting with general extendability, as already described.)

This naturally leads us to consider alternatives to set theory, and indeed *category theory* (CT) stands ready and waiting to step in. Its proponents have been maintaining for decades that it provides an autonomous, alternative foundational scheme that in fact is superior in a number of ways to set-theoretic foundations. Not only is it claimed to be more closely in contact with the actual content of advanced mathematics (e.g., algebraic topology and geometry problems, which it helps solve), it is also claimed to capture better certain key *structuralist* ideas, such as the interdependence of structures through various kinds of mappings and, in particular, the idea of a multiplicity of universes of discourse for mathematics in contrast with the fixed universe view of set theory. Unlike set theory, in which the content of a mathematical concept is fixed by referring it once and for all to a fixed absolute universe of sets, in "category theory" any mathematical concept acquires a *plural* reference through varying the category of discourse to which it is referred. This is well illustrated by the group concept. As a set-theoretical object, a group is a set equipped with a couple of operations satisfying certain elementary axioms expressed in terms of the elements of the set. By contrast, in category theory the group concept is given an "arrows only" formulation, in which it becomes a "group object" capable of living in virtually any category. In the category of topological spaces, for example, a group object is nothing other than a *topological group;* in the category of differentiable manifolds it is a *Lie group;* and in a category of sheaves it is a *sheaf of groups.*

Mac Lane, Bell, and others have proposed developing mathematics within suitable toposes, categories with a rich hierarchical structure generalizing certain key features of sets, roughly, those features that persist when sets are allowed to *vary* in some way. (As we have said, these features include the formation of Cartesian products, function classes by exponentiation, and extensions of predicates.) Any topos may be conceived as a possible universe of discourse in which mathematical arguments can be pursued and mathematical constructions carried out. A topos has its own internal language that describes it, and its own internal logic, which, in general, is not classical but intuitionistic. But classical logic emerges if certain further mathematical conditions, for example, the Axiom of Choice, are imposed. Thus, topos theory already accommodates both classical and constructive mathematics, allowing different universes for them built on a common core.

The plurality of reference already conferred on mathematical concepts by category theory is carried a stage further in topos theory. Take, for example, the concept *real-valued continuous function on a topological space X*. Any such function may be regarded as a real number, or quantity, *varying continuously* over *X*. Now consider the topos **Sh**(*X*) of sheaves on *X*. Here a sheaf on *X* may be conceived as a set undergoing continuous variation, in a suitable sense, over (the open subsets of) *X*. In that case, **Sh**(*X*) may be viewed as a universe in which everything is undergoing continuous variation over *X*, "co-moving," as it were, with the variation over *X* of any given varying real number. This causes the variation of the latter to be "unnoticed" in **Sh**(*X*); it is accordingly regarded there as being a *constant* real number. In other words, the concept *real number,* interpreted in **Sh**(*X*), corresponds to the concept *real-valued continuous function on X*. This shows that, from the standpoint of topos theory, a mathematical concept may be assigned a fixed *sense,* but may nevertheless have a plural *reference*. Indeed, we may take the sense of the concept *real number* as being fixed by a suitable definition in the common internal language of toposes, while its reference will depend on the topos of interpretation. In **Sh**(*X*), that reference will be, as we have seen, not the usual *real number* concept but *real-valued continuous function on X*. That is, reference is determined only *relative* to a topos of interpretation.

Another instance of the relativity of mathematical concepts, one familiar to all set-theorists, is the phenomenon of *cardinal collapse*. Here, given an uncountable set *I,* we can produce a "universe of sets"—actually a Boolean extension of the universe of sets—in which *I* is *countable*. This means that the cardinality of an infinite set is not an absolute or intrinsic feature of the set but is determined only in relation to the mathematical framework with respect to which that cardinality is "measured."

This shows that topos theory is *pluralistic*. But it is at the same time objective in that (certain) toposes may be seen as depicting, in an idealized way, objective aspects of the world, only no unique topos describes that world in its totality. For example, the *smooth topos* provides an idealized description of the geometric structure of the world, idealized through the assumption that all objects and maps are continuous and smooth. At the other extreme, the *topos of sets* presents the world as an entirely discrete structure in which objects are given purely in terms of their cardinality. Still another example is the *effective topos,* in which the world is viewed in terms of computability, and requires all functions to have algorithms. The evident pluralism we again see here arises not because we are dealing with competing theories, but because the alternatives are suited to different purposes. So it is not meaningful to ask whether it is "really" the case that all functions from the real line to itself are differentiable, or whether it is really "true" that all functions from the natural numbers to themselves are recursive, let alone whether any solid sphere is "really" decomposable into five pieces that can be fitted together to make two solid spheres of the same size. (This is an instance of the *Banach-Tarski paradox,* a consequence of the Axiom of Choice, which is generally assumed to hold in a topos of sets.) Instead, one recognizes such features as being tied to the relevant idealization, as being, if you like, "objective" features of that idealization, but not embodying any sort of claim about the (mathematical or physical) world *tout court.*

All this has led Bell (1986), for example, to propose that mathematics should be seen as local, or relative to a choice of background topos. Theorems common to all the suitable toposes form the constructively provable common core. Beyond that, objectivity requires relativization to particular toposes, in analogy with relativistic physics. (Whether famous examples of undecidables of set theory, such as the Continuum Hypothesis, can be thought of as "objective" even in such a relative sense—i.e., in the case of CH, relative to a topos in which power objects are maximal—is a separate, debatable matter.)

This is an attractive view, as far as it goes. Category theory does provide a mathematically interesting generalization of set theory and does offer insights into "mathematical structure," revealing, for instance, how mathematical content is often only "up to isomorphism." However, it does not go far enough, or, better, it does not start early enough, or—more accurately still—it is not clear just where it starts. The problem can be brought out by attending to the term "category theory" itself. It is ambiguous, along with the term "axiom." On the one hand, there are first-order axioms defining what a category is, and various additions to these defining various types of topos (elementary, free, well-pointed, etc.). These are axioms only in the sense of *defining conditions,* telling us what these structures are, as in abstract algebra, where one has axioms for groups, rings, fields, and so on. As components of definitions, these so-called axioms assert nothing, and so are not proper axioms in the traditional, Fregean sense—evident truths, in an absolute sense, or at any rate *assertions* with a determinate truth-value, apart from being evident. (To be sure, this is compatible with such axioms decisively capturing a prior, well-motivated *conception* of a domain or type of object, as in the cases of the axioms for toposes mentioned above, for smoothness or discreteness or recursivity. Here the axioms are akin to the postulates of Euclidean geometry, if we read those as being true of our conception of space, rather than applying literally to actual physical space.) In fact, some category theorists have gone even further, explicitly reading their defining conditions in a Hilbertian, structuralist way: any objects whatever bearing a relation formally behaving like composition of functions (as spelled out in the CT axioms) constitute a category. In other words, the primitives of the language of CT are not even given a definite interpretation, but are treated as placeholders or variables. On the other hand, "category theory" as practiced by mathematicians involves substantive, even deep theorems, and surely these are assertory. But in what framework are these results proved? Not simply in the systems of definitions, as is clear from cases in which various categories or toposes are brought into functorial relations with one another. As Feferman (1977) pointed out, notions of *collection* and *operation* are presupposed just in saying what a category or a topos is as well as in relating them. And indeed, the typical text in the subject, which of course is presented as informal mathematics, makes reference early on to a given, background universe of sets, that is, category theory is not being presented as an autonomous foundational framework at all; rather, set theory is presupposed in the background as is standard in other branches of abstract mathematics (algebra, topology, etc.). As pure mathematics, this is fine; but clearly the CT foundationalist who would transcend the single-universe set-theoretic hierarchy must put on another hat and articulate an alternative framework. At a minimum, a background logic must be specified, including (asserted) axioms governing operations or relations and, presumably, governing the mathematical existence of categories and toposes.

Efforts to create such an alternative framework by explicitly axiomatizing the metacategory of all categories were in fact initiated by Lawvere (1966) and extended by Blanc and Donnadieu (1976). But there are difficulties with the claim that these axiomatizations could constitute an autonomous foundation for mathematics. Primitives such as "category" and "functor" must be taken as having definite, understood meanings, yet they are in practice treated algebraically or structurally, which leads one to consider interpretations of such axiom systems, that is, their semantics. But such semantics, as of first-order theories generally, rest on the set concept: a model of a first-order theory is, after all, a set. The foundational status of first-order axiomatizations of the metacategory of categories is thus still somewhat unclear.

The bearing of all this on the issue of pluralism should be evident: if indeed CT is dependent on a background universe of sets, then the plurality of universes of discourse for mathematics ultimately reduces to the plurality of universes of discourse for set theory. The plurality of toposes in which much mathematics can be developed may still be quite interesting in its own right, but all toposes would be seen as living inside models of set theory. On the level of theories, set theory would have to be seen as more fundamental, and CT's promise of an alternative, autonomous foundational approach would not be fulfilled. This is especially disappointing when we recall that if we ask, "What plurality of intended universes for set theory is there?" the standard answer is "None, there is just *the* cumulative hierarchy," although within this there may also be many (less than exhaustive) models. And it seems we are also stuck with set theory as a massive exception to a structuralist interpretation of mathematics.

It turns out, however, that there is a way out of this impasse, but at a price. If we introduce modality and tolerate talk of the possibility of large domains of discourse—essentially just large numbers of objects—then we have a natural way of recognizing a plurality of models of set theory and toposes, living side by side within these domains, of which there also can be many, but without ever allowing for any totality of *all* such domains. In this view, it does not even make sense to speak of collections or wholes of actual things combined with what merely *might* have existed! One makes sense of collecting, forming wholes, and so on, only within a world, so to speak, not across worlds. (Officially, worlds are not recognized; all this is spelled out with modal operators, ultimately with just one: "it is mathematically possible that . . .") In fact, surprisingly, second-order logical machinery is available to describe not only large domains, in the sense of having inaccessible cardinality, but also structures for set theory and category theory, without ever officially quantifying over classes or relations as objects. Clever combinations of mereology and plural quantification suffice. One must be able to speak of arbitrary wholes of enough pairwise nonoverlapping things (about whose nature we can remain neutral), and we must allow plural locutions, such as "Any things whatever that φ also ψ," as achieving the expressive power of quantification over arbitrary subcollections of the (given, hypothetical) domain of things.[8] For example, the second-order least upper-bound principle takes this form: "Any reals whatever which are all $\leq$ some real are all $\leq$ a least such." This, together with the usual axioms for a complete ordered field, characterizes the real number system up to isomorphism. Similar methods yield characterizations of other key mathematical structures such as the natural numbers, full models of set theory, and various toposes, and so on, again, without ever countenancing classes or relations as objects.

The upshot is that we do have at least one way of consistently combining set theory, category theory, and an open-ended plurality of universes of discourse for mathematics, in accordance with structuralist insights. The assertory axioms of the proposed framework are those of the background logic (essentially second-order logic with mereology) together with axioms asserting the possibility of large domains and guaranteeing extendability, that is, the possibility of ever larger ones. The axioms of set theories proper can then be interpreted structurally as defining conditions on certain kinds of structures. And category theory can be carried out relative to background domains without thereby becoming a (late-ish) chapter of set theory. (In effect, the Grothendieck method of universes has been recovered nominalistically.)

Unlike the first kind of pluralism discussed above, this pluralism in ontology seems distinctively attractive, even necessary, for mathematics, as compared with the natural sciences. After all, we live in a unique world, don't we? Pure mathematics is content to deal with mere conceptual possibilities, but the natural sciences aim to describe and explain reality.[9] Surely there is no analogue of the principle of extendability, articulating the "creative progress" that Zermelo found inherent in mathematics. Short of this, there may well be other multiplicities involving ontology. Of course, on the plane of metascience and perhaps in physics, there are multiple ways of conceiving even the material world, with or without properties, with or without space-time points as objects, with or without particles (e.g., with only quantum fields), and so on. Are these cases of genuine equivalence, and hence (?) only apparent choices, or undecidable questions? Or are we driven to a kind of ontological relativity favored by Carnap (1956, Suppl. A, 205–21)? If so, then in the natural sciences, as well as in mathematics, absolutist talk of "reality" or even the more humble sounding "everything," should really be given up.[10]

For a scientific example of "many worlds" in a very different sense, there is, of course, the notorious "many worlds" interpretation of quantum mechanics (the de Witt version of the Everett interpretation, with actual splitting practically whenever "anything definite happens"), but the objections that have been raised against this seem to us decisive. More promising, perhaps, cosmologists now explore ideas about a *multiverse* instead of the universe, multiple real cosmoses arising from quantum mechanical processes, including inflation. Certain seemingly intractable questions are then blocked, for example, "Why does the actual cosmos satisfy the very special conditions of the constants of nature permitting the formation of galaxies, let alone life?" (Answer: Bad question, for there are many actual cosmoses in which galaxies never form. The reformulated question, "Why does *this* cosmos—the one *we* experience—satisfy those special conditions?" seems like a nonsense question, something like "Why am I me and not you?" which nevertheless kept us from getting to sleep sometimes as children.) And, of course, there is the whole issue of emergence (versus reduction), still not entirely resolved, even at the level of chemistry vis-à-vis quantum physics. Should we recognize multiple categories of *properties* and *relations (attributes),* corresponding to different, irreducible levels of scientific inquiry? But these questions cannot be addressed within the scope of this essay (which we hereby guarantee by stopping).

Notes

1. To be sure, classical practice itself does not imply endorsement of Platonism, as many mainstream mathematicians, if pressed, fall back on some kind of formalism or fictionalism. "Platonism" designates a reflective view, based on a literal, face-value reading of mathematical discourse, which would justify the practice. It may well not be the only, or the best, justification, however.

2. While it is true that no consistent intuitionistic *propositional* theory can be in formal contradiction with classical logic, this is far from being the case for intuitionistic first or higher-order theories. For example, the sentence $\neg\forall x\forall y\ (x = y \lor x \neq y)$ is consistent in intuitionistic, but not classical logic. Indeed, such striking "conflicts" with classical mathematics—famously Brouwer's continuity theorem—arise in intuitionistic analysis, where nonclassical axioms of continuity governing choice sequences are available. The Bishop framework abandons any such nonclassical axioms and so generates no such conflicts. However, even in the intuitionistic case, these formal conflicts are only apparent, not real, turning on ambiguity of the logical notation, as will be explained later. (For a fuller discussion bringing out certain expressive limitations of intuitionism, see Hellman 1989.)

3. For a sustained critique of Dummett's case, see Burgess 1984.

4. These arise in connection with quantum mechanics and general relativity, but need not concern us here. See, e.g., Hellman 1993 and 1998.

5. To be sure, one can raise questions concerning the "universe of constructions" to which constructive proof-conditions appeal. But at least on a rough-and-ready, ordinary understanding of those conditions, anyone can see that the intuitionistic rules are correct and why certain classical principles and rules must be dropped.

6. According to anecdote, even the intuitionist Heyting seems to have shared this perspective, as he liked to teach classical recursion theory. He said he found it interesting.

7. On the iterative conception, sets are arranged in a hierarchy of stages corresponding to (finite and transfinite) ordinals. These "go on and on" in virtue of two main operations, passing from a set to its power set (set of all subsets of the given set), and taking the limit of any ordinal sequence of sets (or taking as a set the range of any function on a given set or ordinal (the content of the Axiom of Replacement). A stage so large that it cannot be reached from below by either of these operations is called "(strongly) inaccessible."

As Zermelo (1930) proved, an inaccessible stage provides a model for the ZF axioms, and so, by Gödel's second incompleteness theorem, the existence of inaccessibles cannot be proved within ZF. Nevertheless, they are regarded as quite legitimate by set theorists.

8. A famous example of Geach, "Some critics admire only one another," illustrates that the logic of plurals goes well beyond first order. On the usual, "singularist" view, there is hidden quantification over *classes* (of critics, in this case), but Boolos (1985) proposed turning this on its head, taking plural quantification as already understood and interpreting class quantification through it. This idea has been applied by Burgess, Hazen, and Lewis (Appendix to Lewis 1991), to get the effect of ordered pairing of arbitrary individuals without any set-theoretic machinery. Lewis (ibid.) has also argued, persuasively in our view, that we do have an independent grasp of plural quantifiers. These ingredients have played an important role in recent developments of a modal-structuralist approach to mathematics (e.g., Hellman 1996, 2003). A systematic, more ambitious treatment of the logic of plurals is given by Yi (2005).

9. Notoriously, Nelson Goodman (1978) challenged this assumption, where the "real world" literally gives way to multiple world versions (even apart from mathematics). We think, along with Scheffler, however, that here Goodman goes too far. See, e.g., Scheffler 1980.

10. This is quite compatible with an open-ended, context-relative understanding of quantifier phrases, which, arguably is all that is needed for ordinary expression and reasoning.

References

Beeson, M. J. 1985. *Foundations of Constructive Mathematics.* Berlin: Springer.

Bell, J. L. 1986. "From Absolute to Local Mathematics." *Synthese* 69: 409–26.

———. 1988. *Toposes and Local Set Theories.* Oxford: Oxford University Press.

Bishop, E. 1967. *Foundations of Constructive Analysis.* New York: McGraw-Hill.

Blanc, G., and M. R. Donnadieu. 1976. "Axiomatization de la categorie des categories." *Cahiers Topologie Geom. Differentielle* 17: 135–70.

Boolos, G. 1985. "Nominalist Platonism." *Philosophical Review* 94: 327–44.

Burgess, J. 1984. "Dummett's Case for Intuitionism." *History and Philosophy of Logic* 5: 177–94.

Carnap, R. 1956. "Empiricism, Semantics, and Ontology." In *Meaning and Necessity.* Chicago: University of Chicago Press.

Dummett, M. 1977. *Elements of Intuitionism.* Oxford: Oxford University Press.

Feferman, S. 1977. "Categorical Foundations and Foundations of Category Theory." In *Logic, Foundations of Mathematics, and Computability Theory,* ed. R. E. Butts and J. Hintikka, 149–69. Dordrecht: Reidel.

Goodman, N. 1978. *Ways of Worldmaking.* Indianapolis: Hackett.

Hellman, G. 1989. "Never Say 'Never'! On the Communication Problem between Intuitionism and Classicism." *Philosophical Topics* 17: 47–67.

———. 1993. "Constructive Mathematics and Quantum Mechanics: Unbounded Operators and the Spectral Theorem." *Journal of Philosophical Logic* 22: 221–48.

———. 1996. "Structuralism without Structures." *Philosophia Mathematica* 4, no. 2: 100–123.

———. 1998. "Mathematical Constructivism in Spacetime." *British Journal for Philosophy of Science* 49: 425–50.

———. 2003. “Does Category Theory Provide a Framework for Mathematical Structuralism?” *Philosophia Mathematica* 11, no. 2: 129–57.

Lawvere, F. W. 1966. “The Category of Categories as a Foundation for Mathematics.” In *Proceedings of the Conference on Categorical Algebra, La Jolla 1965*, ed. S. Eilenberg et al. Berlin: Springer.

Lewis, D. 1991. *Parts of Classes.* Oxford: Blackwell.

Mac Lane, S. 1986. *Mathematics: Form and Function*. New York: Springer.

Putnam, H. 1967. “Mathematics without Foundations.” *Journal of Philosophy* 64: 5–22.

Scheffler, I. 1980. “The Wonderful Worlds of Goodman.” *Synthese* 45: 201–9.

Shapiro, S. 1997. *Philosophy of Mathematics: Structure and Ontology.* New York: Oxford University Press.

Yi, B. 2005. “The Logic and Meaning of Plurals. Part I.” *Journal of Philosophical Logic* 34: 459–506.

Zermelo, E. 1930. “Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre.” *Fundamenta Mathematicae* 16: 29–47.

5
Pluralisms in Economics

In 1992, a group of economists issued a "Plea for a Pluralistic and Rigorous Economics" in an advertisement in the *American Economic Review,*[1] calling for "a new spirit of pluralism in economics, involving critical conversation and tolerant communication between different approaches. Such pluralism should not undermine the standards of rigor; an economics that requires itself to face all the arguments will be a more, not a less, rigorous science." The announcement had been organized by Geoffrey Hodgson, Uskali Mäki, and D. McCloskey, and signed by forty-four illustrious names including Nobel laureates Franco Modigliani, Paul Samuelson, Herbert Simon, and Jan Tinbergen.

In 1993, the International Confederation of Associations for Pluralism in Economics (ICAPE) was founded as a "consortium of over 30 groups in economics" that "seeks to foster intellectual pluralism and a sense of collective purpose and strength."[2] Its 1997 resource list contained thirty professional associations, thirty-two academic and policy journals, eleven publishers, sixteen departments, sixteen centers, and nine special projects, not all of which were formally affiliated with ICAPE. The consortium's statement of purpose suggests: "There is a need for greater diversity in theory and method in economic science. A new spirit of pluralism will foster a more critical and constructive conversation among practitioners of different approaches. Such pluralism will strengthen standards of scientific inquiry in the crucible of competitive exchange." ICAPE's first conference, "The Future of Heterodox Economics," was held during the summer of 2003.

In 2000, a group of economics students in France, under the banner "autisme-économie," published a petition on the Web in favor of a pluralism of approaches in economics.[3] The students wrote, "We want a pluralism of approaches, adapted to the complexity of the objects and to the uncertainty surrounding most of the big questions in economics." Their plea was supported by a petition from the hands of some economics teachers in France, who also stressed the need for a plurality of approaches adapted to the complexity of objects analyzed.[4] They noted: "Pluralism is not just a matter of ideology, that is, of different prejudices or visions to which one is committed to expressing. Instead the existence of different theories is also explained by the nature of the assumed hypotheses, by the questions asked, by the choice of theoretical spectrum, by the boundaries of problems studied, and, not least, by the institutional and historical context." The teachers concluded: "Pluralism must be part of the basic culture of the economist. People in their research should be free to develop the type and direction of thinking to which their convictions and field of interest lead them. In a rapidly evolving and ever more complex world, it is impossible to avoid and dangerous to discourage alternative representations."

In 2001, twenty-seven economics Ph.D. students at Cambridge University in England, who have come to be known as the "Cambridge 27," issued a petition titled "Opening Up Economics."[5] They ended their proposal for reforming economics as follows: "We are not arguing against mainstream methods, but believe in a pluralism of methods and approaches justified by debate. Pluralism as a default implies that alternative economic work is not simply tolerated, but that the material and social conditions for its flourishing are met, to the same extent as is currently the case for mainstream economics. That is what we mean when we refer to an 'opening up' of economics."

Implicit in all these appeals is the observation that economics lacks pluralism. The pleas are defended by means of an assortment of arguments, such as discussions of the complexity of the economy, evaluations of the restrictions inherent in modeling, and assessments of the cognitive limitations on the part of economists. The advertisement in the *American Economic Review* also employs a reflexive strategy: "Economists today enforce a monopoly of method or core assumptions, often defended on no better ground [than] that it constitutes the 'mainstream.' Economists will advocate free competition, but will not practice it in the marketplace of ideas."[6]

Since pluralism itself is a reflexive doctrine, this chapter develops an understanding of various forms of pluralism, or lack thereof, in economics.[7] In particular, it argues that pluralism in economics is recurring, but often denied. Instead of locating the source in epistemology, metaphysics, and the like, the analysis in the subsequent sections proposes that the lack of success of the monist movement in economics strengthens the case for pluralism and therefore suggests that pluralism is contingently true. The following section offers an overview of movements toward monism about theories, showing that repeated efforts at securing a single theory have failed. These developments have extended toward the level of economies, as suggested by the subsequent section, which shows that attempts to treat economic agents monistically have failed. The lack of success of these efforts to achieve monism has paved the way for a (full-fledged) return to pluralism, as elaborated in the final section.

Monism about Theories

As evidenced by the pleas organized by Hodgson, Mäki, and McCloskey, the French students, the French faculty members, and the Cambridge 27, economics is currently characterized by efforts to achieve monism at the theoretical level. However, this has not always been the case. During the period before World War I and the interwar period, pluralism was the dominant force in economics (Morgan and Rutherford 1998a).[8]

Before World War I, the social gospel movement exerted an extensive influence on economics (Bateman 1998). Since it was compatible with several types of economic analysis, it served the function of sanctioning pluralism, provided the focus was on social justice. As a result, it supported several approaches in economics, including institutionalism and neoclassicism, also known as marginalism. In Bradley Bateman's (1998, 39) words, "Institutionalists and marginalists could coexist . . . as long as the issue was reform rather than revolution and as long as ethical concerns informed their work." When the progressive movement declined after World War I to make room for a focus on "realism," both institutionalism and neoclassicism continued to flourish.[9]

During the interwar period, pluralism characterized economics on many levels. Whereas institutionalism and neoclassicism coexisted, they were individually also highly pluralistic. Institutionalism was a nonexclusive, broad movement and neoclassical economics was highly diverse as well. In addition, individual members of these groups adopted a variety of theoretical stances. Mary Morgan and Malcolm Rutherford (1998b, 8) describe the situation as follows: "Economists of the early twentieth century shared a kind of scientific economics (more often concrete than abstract), a moral commitment to ensure standards of scientific inquiry, and an evenhanded objectivity combined with advocacy. Pluralism was supported, not compromised by these standards." Institutionalist economists started coming under attack in the 1930s, partly because they were unable to provide a set of policy recommendations that were considered to be successful against the Great Depression (Bateman 1998). However, it took a watershed event like World War II for these to have the effect desired by the neoclassical economists.

World War II stimulated the move in economics toward monism about beliefs, ideology, theories, models, and policy advice, with the formalism of neoclassical economics pushing out her institutionalist sister. During the war, heavy demands had been placed on economists to develop tools for solving policy problems. Sharing in the glory of the subsequent victory, economists emerged with a firm belief in the formalism that characterized neoclassical economics. While economics became associated with a certain tool kit as opposed to a particular area of study, the formalism further supported economists' efforts to gain identity as a national science, to achieve professional status. As Morgan and Rutherford (1998b, 19) note: "[T]he transformation into formal economics involved changes in language, form, and tools. This new style became a set of mores that reduced in itself the possibility of pluralism in economics."

To fully understand the transformation from pluralism and monism, one must not only appreciate the changing nature of mathematics and mathematical economics,[10] but also the multiple dimensions of the process that strengthened neoclassicism and weakened institutionalism. While there had been a focus on personal qualities and attitudes of economists during the interwar period, objectivity came to be associated with a particular set of methods, namely, mathematics and statistics, after World War II. At the same time, economists gradually moved away from advocacy. The success of the new set of methods with which neoclassical economists came out of World War II instilled in them a belief in the ideas behind them. Simultaneously, American society moved from a desire for economic intervention toward support for free markets and open competition, thereby further strengthening the neoclassical belief system.[11]

Institutionalism was at odds with the new scientific styles demanded by the patrons of economics[12] and further weakened by the turn away from planning and regulation towards the market and competition as instruments of control (Balisciano 1998). During the Cold War period, the technical turn in economics was intensified as a result of a continued narrowing in the range of beliefs, an additional tightening of acceptable ways of expressing them, and open prosecution during the McCarthy period.[13] In the process, the possibilities of pluralism in economics persistently waned as the language, form, and tools of economics continued to narrow. Morgan and Rutherford (1998b, 24) conclude that the decline of pluralism in American economics took place "within structures involving patrons and hierarchies operating within the context of a political and economic society that supported calls for economic intervention in the interwar period and for free markets in the postwar period."

Complicating our admittedly simplified description here and foreshadowing our claim that pluralism in economics is recurring, though often denied, some have suggested that neoclassical economics owes its strength to its persistent inability to enforce any monolithic orthodoxy. For instance, Wade Hands and Philip Mirowski (1998; Mirowksi and Hands 1998) outline three approaches to neoclassical demand theory, associated with the University of Chicago Economics Department (in particular Milton Friedman and George Stigler), the Cowles Commission at the University of Chicago (especially Kenneth Arrow and Gerard Debreu), and the Massachusetts Institute of Technology (most notably Paul Samuelson).[14] And Perry Mehrling (1998, 295) suggests that "although the neoclassical language might have become hegemonic, what economists wanted to say with that language remained as pluralist as in the interwar years."

In addition, our focus has been almost exclusively on developments in the United States, which is justified by "the United States' predominant influence on the expansion and internationalization of economics during the past half century" (Coats 1996b, 4). As a result, the trends outlined here are spreading, with some lag, to Europe and Japan. In Europe, this has occurred more rapidly in the United Kingdom than on the Continent against the background of the growth of new universities, the imposition of the research assessment exercise, and an expansion of student numbers along with a reduction of resources (Backhouse 2000). At the same time, "the process of internationalization has by no means obliterated national differences" (Coats 1996b, 4). This may explain why several of the pleas outlined in the introduction originated from Europe, perhaps as opposition to the type of economics emanating from the United States.[15]

It should also be acknowledged that our focus has so far been on microeconomics, which concentrates on the decisions of people and businesses. We will learn in the remainder of this section that pluralism reemerged during efforts to reduce other fields to microeconomics. To start, microeconomics has come under attack for not having a notion of "the social" other than summing "the individual" (Hands 1994, 1995, 1997a). As Hands (1997a, S112–13; original emphasis) explains: "Since the social is merely the sum of the individuals, economists cannot accommodate any concept of the social that is *qualitatively different* from that which is possessed by the individual economic agents." Briefly, neoclassical economists rely on two notions of social efficiency, the Pareto criterion and the compensation principle. According to the Pareto criterion, an allocation of resources is Pareto efficient if it is not possible to make one person better off without making another person worse off. Hence, assessments of social efficiency are based on individual well-being. In other words, there is no qualitative transformation involved. According to the compensation principle, an efficiency-improving reallocation of resources requires the gains to the winners to be greater than the losses to the losers, which would allow the winners to compensate the losers and still be better off. Again, assessments of social efficiency are established by adding up over individuals.

In other words, again the social is not different from the individual, not something with unique or emergent properties.

Even if one accepts the exclusive focus on the individual in economics, problems occur. As neoclassical economists themselves have acknowledged (e.g., Arrow 1959), competitive markets require something beyond an individualistic explanation. Basically, each individual agent in a competitive market takes prices as given in her individual choice problem. This raises the question of where these prices come from (Hands 1995). If they arise from something other than the individual agents, then one no longer offers a consistently individualistic explanation. Hands (617), therefore, concludes, "The result is that the 'competitive market model,' ostensibly the paragon of successful individualistic social science, is dependent on something outside of (or above, or prior to) the individual agents for its primary explanandum (competitive prices)."

Accepting the stress on the individual and ignoring some of its limitations, the efforts to achieve monism about theories in microeconomics inspired efforts to reduce other fields to it, as suggested by McCloskey (1982, 7): "Although its Greek meaning is 'small housekeeping,' microeconomics is not the little or trivial portion of economics. On the contrary, it comes close to being the whole. Not all fields of economics are based on microeconomics, but all strive to be. Most of the lasting advances in economic thinking over the past century or so have consisted of reducing one or another piece of economic behavior to microeconomics."[16] These endeavors have extended to macroeconomics, which studies the national and global economy. In particular, they have focused on establishing microfoundations of macroeconomics, as Lawrence Boland (1982, 80) confirms: "[T]he demonstration of the existence of microfoundations for macrotheories is considered essential by many leading economists. The reason . . . is easy to find. Demonstrating the dependence of all macroeconomics on microeconomic principles is essential for the fulfillment of the (methodological) individualist requirements of neoclassical economics."[17]

These attempts to develop neoclassical microfoundations for macroeconomics date back to the years just after World War II, as evidenced by the observation by Lawrence Klein (1946, 93): "[T]hese aggregative theories [i.e., macroeconomic theories] have often been criticized on the grounds that they mislead us by taking attention away from basic individual behavior. The problem of bridging the gap between the traditional theories based on individual behavior and the theories based on community or class behavior is, to a large extent, a problem of proper measurement" (also see Janssen 1993; Nelson 1984; Weintraub 1977, 1979). To be more precise, the problem of aggregation consists of two components (Deaton and Muellbauer 1980; Green 1964; Theil 1954): first, whether there exist functional relationships among macroquantities obtained by aggregating relevant microquantities; and second, whether the functions obtained by aggregating microfunctions are the same as the macrofunctions derived independently. In the process, neoclassical economics was modified in a variety of ways to provide a conceptual base for the formulation of macroeconomic concerns.[18] And "by say 1960 the microfoundations problem appeared, on the surface, to be 'settled'" (Weintraub 1977, 4).

Matters changed during the 1960s, when non-neoclassical economists uncovered difficulties with aggregating from the individual (Harcourt 1969, 1972; Kurz and Salvadori 1995; Robinson 1953), as evidenced by what has come to be known as the "Cambridge controversies in the theory of capital," indicating the critics in Cambridge, England, and the defenders in Cambridge, Massachusetts. The target of attack was the aggregate production function, which refers to a neoclassical construct (a macroeconomic version of a firm's production function) in which inputs or capital and labor are considered to have a technical (i.e., engineering) relation to aggregate production. In the course of investigating the meaning of this production function for total output, Joan Robinson (1953) found that this construct is incoherent because of the fuzzy nature of the capital variable. In particular, the British side of the controversy outlined two problems with the aggregate production function: reswitching and reverse capital deepening. For reswitching to occur, one set of techniques must be chosen for at least two different ranges of the interest rate, with other sets of techniques selected at intermediate ranges. Consequently, there is no unambiguous relationship between changes in input proportions and changes of the so-called factor prices, which is a central element of the neoclassical explanation of distribution in terms of supply and demand. For reverse capital deepening to arise, the relationship between the value of capital (per capita) and the rate of profits must be increasing. As a result, a higher interest rate may be associated with a switch to a more capital-intensive technique, implying that the interest rate is not a scarcity index for capital, which is a core component of the neoclassical approach.

Neoclassical economist Joseph Stiglitz (1974, 898) drew the following conclusion from these insights: "[T]he restrictions embodied in neoclassical macroeconomic models do not necessarily follow from the microeconomic (disaggregative) models from which they should be derived." For non-neoclassicals, who continued to remain outsiders as a result of the forces outlined before, this implied the need to create alternative microeconomic models.[19] One response of neoclassical economists was to retreat to microeconomic theory. Another was to refer to the aggregate production function as a useful parable and to dismiss the possibilities explored by the British as a curiosum, a perversity, not a serious economic problem, or a red herring. The former answer started drawing support only after the discovery of additional hurdles, as we will learn shortly, but the first one was the favored response for most neoclassical economists, including Stiglitz (1974, 899): “I believe that, under most circumstances and for most problems, the errors introduced as a consequence of aggregation of the kind involved in standard macroanalysis are none too important.”[20]

In the late 1960s, the rise of rational expectations economics at the macro level gave new impetus to the microfoundations project and the associated efforts to achieve monism about theories (Sent 1998).[21] In particular, rational expectations economists argued that the suboptimal use of available information under adaptive expectations was hard to reconcile with the idea of optimization that was the foundation of neoclassical economic analysis. Instead, rational expectations economists claimed that since agents were posited as optimizers, it was only natural to presume that they would also form their expectations optimally. In other words, the rational expectations hypothesis was a direct derivation from the neoclassical optimization principle extended to the problem of expectations of future events. In particular, optimizing over perceptions implied that agents did the best they could and formed their views of the future by taking account of all available information, including their understanding of how the economy works. If perceptions were not optimally chosen, there would exist unexploited utility or profit-generating possibilities within the system. Hence, rational expectations economists insisted on the disappearance of all such unexploited possibilities.[22]

Rational expectations economists contrasted their approach with Keynesian analyses. They argued that economics had to account for the decisions of firms and people in ways that were consistent with the idea of optimizing behavior, because ad hoc assumptions about the behavior of firms and people did not sit well with the microfoundations of economic theory. At the same time, they criticized the typical Keynesian assumptions that markets did not clear and that economic agents did not always pursue optimizing strategies, because both implied ad hoc departures from the axiom of rational behavior. Hence, rational expectations economics may be viewed as replacing earlier ad hoc treatments with an approach squarely based on the microfoundations of incentives, information, and optimization.

In the 1970s, textbook author John Beare (1978, 7) felt justified to celebrate the inclusion of macroeconomics in the efforts toward monism about theories when he wrote: “Macroeconomics deals with relationships between aggregate variables, the rigorous derivation of which now tends to be based on relationships implied by microeconomic theory." However, the author celebrated too soon, as illustrated by the so-called Sonnenschein-Debreu-Mantel result (Sonnenschein 1972; Debreu 1974; Mantel 1976; Kirman 1989, 1992; Rizvi 1994a).[23] In 1972, Hugo Sonnenschein considered the restrictions imposed on the structure of aggregate demand functions; in 1974, Gerard Debreu continued this line of work. They found that under standard neoclassical assumptions on the individual consumers, such as strict convexity and monotinicity of preferences, so that each agent is characterized by textbook indifference curves and a positive bundle of endowments of all goods, one can derive an excess demand curve for each individual. Summing over all individuals, of whom it is assumed that there are only a finite number, gives the excess demand curve for society as a whole. Under certain not-very-restrictive conditions, three properties will carry over from the individual's excess demand curve to the aggregate demand curve: continuity, a value of total excess demand that must equal 0 at all prices, and excess demand that is homogeneous of degree 0.

In addition, Sonnenschein and Debreu established that these three properties are the only ones that carry over from the individual to the aggregate demand function. In particular, the weak axiom of revealed preference (WARP) may not be satisfied at the aggregate level. Yet, if we are to obtain uniqueness and stability of equilibria, some such restrictions must be imposed. Hence, if WARP is imposed on aggregate excess demands, the economy is presumed to act as if it were just one big consumer. This line of work did not remain isolated, and research by Rolf Mantel showed that the same situation obtains even if the class of admissible preferences is restricted even further. Hands (1995, 617) succinctly summarizes the problem: "In other words, the standard micro model has almost no implications for macrobehavior."

These difficulties in achieving monism about theories may have inspired neoclassical economists such as Frank Hahn to endorse pluralism about theories. For instance, Hahn (1984, 7–8) wrote: "The most strongly held of my views . . . is that neither is there a single best way for understanding in economics nor is it possible to hold any conclusions, other than purely logical deductions, with certainty. I have since my earliest days in the subject been astonished that this view is not widely shared. Indeed, we are encompassed by passionately held beliefs. . . . In fact all these 'certainties' and all the 'schools' which they spawn are a sure sign of our ignorance . . . we do not possess much certain knowledge about the economic world and . . . our best chance of gaining more is to try all sorts of directions and by all sorts of means. This will not be furthered by strident commitments of faith."[24] In fact, the present situation in (mainstream) economics may be characterized as one of moderate pluralism. Sheila Dow (2002,

7) explains: "There is in particular a bifurcation between theoretical and applied mainstream economics. Both theoretical and applied models, in turn, are often partial." Recent years have witnessed, for instance, efforts to incorporate bounded rationality approaches, behavioral insights, chaos theory, complexity approaches, and experimental methods, some of which will be discussed in the final section of this chapter.

Before I conclude, a discussion of further responses to the failed attempts at monism about theories will lead us to a section on monism about economies.

Monism about Economies

Whereas pluralism about theories is a familiar concept, pluralism with economies as the object is perhaps less so. It concerns an economy in which people (or groups) value things differently and in which this diversity is valued (Hargreaves Heap 1997). There is not just a plurality, but also a political commitment to pluralism. As this section shows, economics does not respect a diversity of views concerning the agents who populate its models. Much like the previous one, it illustrates the failure of efforts to establish monism, though now at the level of economies rather than theories. As we will learn, this is one of the consequences of the Sonnenschein-Debreu-Mantel result, but let us first provide two illustrations that economics has always had difficulties dealing with distinctly different agents (Sent 1998).

First, consider the cloning argument Francis Ysidro Edgeworth (1881) developed in the course of analyzing exchange. He started with the idea that exchange between single traders is, to some extent, indeterminate, whereas exchange among numerous buyers and sellers in a competitive market is determinate. Edgeworth, following Antoine Augustin Cournot's lead, proposed to begin with bilateral monopoly and work his way toward perfect competition.[25] This was his famous "recontracting" process, which is based on the suspicion that the core, which is the set of possible outcomes,[26] might shrink as the economy grows. However, since the core is a subset of the allocations space, its dimension keeps changing as the economy grows. Generally, if we allow the economy to grow by increasing the number of agents, we will have more possible coalitions and hence more possibilities for improvement. This led Edgeworth to limit himself to a particularly simple kind of growth in which the number of types of agents stays constant, in other words, in which restrictions are placed on the heterogeneity of the agents. Thus, large economies just have more agents of each type.

Second, consider the fact that general equilibrium theory does not successfully apply to an economy that is fully specialized and in which the possibility of self-sufficiency is the exception rather than the rule (Rizvi 1991). When not every individual in the economy is endowed with sufficient quantities of all commodities required for subsistence, exchange is a necessity for participants' survival. Since the level of equilibrium prices cannot be prejudged, subsistence might not be possible for all agents. The approach taken in existence proofs of general equilibrium before 1975 was basically to remove those agents who are specialized and who need the market to trade into their consumption sets from further consideration, and that means that the economy is not specialized.[27] Nevertheless, even for an economy of self-subsistent individuals, existence could not be shown without further assumptions because the possibility of zero prices precluded a successful demonstration of continuous demand. The continuity problem was remedied by one of two assumptions that further reduce the differences among agents: The interiority assumption increases the endowments of all goods to levels exceeding even those minimally required for self-subsistence; the irreducibility assumption is aimed at securing the use of more realistic, but still self-subsistent, endowments.

Likewise, one response to the Sonnenschein-Debreu-Mantel result has been to reduce differences among economic agents, with macroeconomic models assuming "that the choices of all the diverse agents in one sector . . . can be considered as the choices of one 'representative' standard utility maximizing individual whose choices coincide with the aggregate choices of the heterogeneous individuals" (Kirman 1992, 117). If the behavior of the economy could be represented as that of a representative agent or a number of identical agents, then the situation might be saved, since textbook individual excess demand functions do have unique and stable equilibria.[28] With one representative agent, there clearly can be no difference of opinion, which we call a situation of monism in economies.

Much like the efforts to achieve monism about theories, the developments under review in this section on economies encountered major stumbling blocks. Some of these problems concern the relationship between the representative individual and the group she supposedly embodies. In particular, Alan Kirman identifies several of those difficulties.[29] First, there is no direct relation between individual and collective behavior, because well-behaved individuals need not produce a well-behaved representative agent. Second, the reaction of the representative agent to change need not reflect how individuals of the economy would respond to change. Third, the preferences of the representative individual cannot be used to decide on the desirability of economic situations because they may be diametrically opposed to those of society as a whole. Kirman (1992, 125), therefore, concludes that "the assumption of a representative individual is far from innocent; it is the fiction by which macroeconomics can justify equilibrium analysis and provide pseudo-microfoundations. I refer to these as pseudo-foundations, since the very restrictions placed on the behavior of the aggregate system are those which are obtained in the individual case and, as we have seen, there is no formal justification for this."

Besides the troubled connection between the individual and the collective, representative agent analysis has encountered many other problems (Sent 1998). First, a representative agent is ill-suited to studying macroeconomic problems that are coordination failures, such as unemployment. Second, a representative individual cannot exhibit the complicated dynamics witnessed at the macroeconomic level. Third, how can there be trade among one representative agent? Or suppose there are several representative agents who are alike in several dimensions, how can there be trade among these? One suggestion, following a line of research started by Robert Lucas (1972), is to introduce a certain amount of pluralism in the sense that equilibrium probability beliefs differ and that agents actually trade on the basis of different information. However, a whole series of no-trade theorems overrule this commonsense intuition (see Hakansson, Kunkel, and Ohlson 1982; Milgrom and Stokey 1982; Rubinstein 1975; Tirole 1982; Varian 1987). Briefly, when it is common knowledge that traders are risk-averse, are rational, and have the same priors, and that the market clears, then it is also common knowledge that a trader's expected monetary gain given her information must be positive for her to be willing to trade at the current asset price. In such a situation, other agents would be unwilling to trade with her, because they realize that she must have superior information. The equilibrium market price fully reveals everybody's private information at zero trades for all traders.

One solution to these no-trade theorems has been to return to pluralism about economies. For instance, agents may have different prior beliefs. Now, if differences in prior beliefs can generate trade, then these differences in belief cannot be due to information as such, but rather can be only pure differences in opinion.[30] In other words, they reflect pluralism. Overall, the response to the Sonnenschein-Debreu-Mantel results and the problems associated with the resulting embrace of representative agent analysis has been for neoclassical microeconomics to move toward game theory. As Kirman (1992, 131) explains, "An alternative and attractive approach is offered by game theory, where the interaction between heterogeneous individuals with conflicting interests is seriously taken into account" (also see Rizvi 1994b). Yet, much like its predecessors, game theory does not accommodate a diversity of views concerning the agents who populate its models.

Briefly, game theory relies on a whole range of common knowledge assumptions, thereby reducing pluralism in the sense of diversity of view.[31] Principally, common knowledge is the limit of a potentially infinite chain of reasoning about knowledge.[32] Yet, much like the efforts to achieve monism about economies through representative agent analysis, the common knowledge assumption encountered major hurdles. First, according to the so-called agreement theorem, common knowledge of actions negates asymmetric information about events. In other words, agents cannot agree to disagree. As a result, whenever economic agents come to common knowledge of actions, the joint outcome does not in any way use the differential information about events they each possess. In addition, agents with identical priors must have the same opinion, even with different information, if those opinions are common knowledge. Second, and especially relevant for our narrative, according to the so-called nonspeculation theorem, agents cannot bet and speculation is banished. If it is common knowledge that the agents want to trade, as occurs when agents bet against each other, then the agreement theorem implies that trades must be zero. This is reminiscent of the no-trade theorems mentioned earlier.

Similar to the reasons monism about theories failed, monism about economies encountered significant stumbling blocks, most importantly the no-trade theorems for representative agent analysis as well as game theory. This observation brings us to the concluding section.

Conclusion

In light of the efforts to establish monism on the part of neoclassical economists outlined in the previous sections, it comes as no surprise that outsiders to the mainstream appear to be supporting pluralisms and criticizing monisms, as evidenced by the pleas with which this essay began.[33] However, upon closer scrutiny, heterodox economists frequently are monists about theories. In the opinion of John Davis (1997, 209; original emphasis), the motivation of heterodox economists "is not that their own theoretical approaches are *also* correct—a theoretical pluralist view—but rather that neoclassical economics is mistaken and misguided in its most basic assumptions, and that their own approaches remedy the deficiencies of neoclassicism—a theoretical monist view." This motivation is evidenced, for example, by the observation that the first conference of the International Confederation of Associations for Pluralism in Economics (ICAPE) is on the future of heterodox economics, while orthodox economics is considered to be "vapid, exclusionist, and detached from its social and political milieu."[34] The French students write about neoclassical economics: "We no longer want to have this autistic science imposed on us."[35] And their teachers concur: "Neoclassicalism's fiction of a 'rational' representative agent, its reliance on the notion of equilibrium, and its insistence that prices constitute the main (if not unique) determinant of market behavior are at odds with our own beliefs."[36] Employing the categorizations developed by Ronald Giere (chapter 2 in this volume), the appeals to pluralism on the part of heterodox economics may be seen as an instance of strategic pluralism. Though their advocacy of pluralism may be couched in metaphysical or epistemological terms, it is primarily inspired by efforts to achieve professional power and dominance.

Despite the apparent acceptance of monism, this chapter has illustrated the failure to achieve monism on the part of mainstream economics. It has shown that pluralism is recurring, though often denied. Monism about theories required an evaluation of the individual as well as the social. However, on the one hand, mainstream economics has no notion of the social other than the summing up over individuals. On the other hand, it cannot maintain a unique focus on the individual because this would preclude complete explanation of competitive markets. At the same time, microeconomic findings concerning the individual were shown not to carry over to the social level, as illustrated by the Sonnenschein-Debreu-Mantel result. For monism about economies, these findings resulted in an effort to populate economies with one representative agent. This effort to reduce differences of opinion resulted in major stumbling blocks, including a problematic connection between the representer and the represented as well as a lack of trade, which, supposedly, is one of the main foci of economy analyses. These difficulties resulted in a move toward game theory, which laid bare new problems with monism about economies. In particular, agents cannot agree to disagree, they cannot bet, and speculation is banished.

The breakdown of the microfoundations project suggests that phenomena at the micro and macro levels in economics are so complex that one theoretical approach, such as microeconomics, does not have the resources to provide a complete explanation or description of them. For economists, these failures have led them in the direction of exploring cognitive limitations on the part of the agents who populate their models. For macroeconomists, incorporating bounded rationality could modify or take the edge off the very sharp no-trade theorems (Sargent 1993, 15; Sent 1997). For game theorists, absence of a fully rational treatment of knowledge may circumvent no-trade theorems by allowing speculative trade (A. Rubinstein 1998, 56–60; Sent 2004).

Observing these developments, Abu Rizvi (1994, 19n) noted that "[i]t is interesting that Simon's ideas were not used by mainstream theorists for years but have recently been 'discovered.'" And Herbert Simon (1992, 266) observed: "Readers would not be deceived by the claim that economists flocked to the banner of satisficing man with his bounded rationality. The 'flocking' was for a long time a trickle that is now swelling into a respectable stream."[37] These connections with Simon's insights strengthen the suggestion that some parts of the world are so complex that they cannot be fully accounted for from the perspective of a single representational idiom because Simon's research agenda focused on analyzing complex, hierarchical systems (Sent 2001). Simon's (1996, 184) interpretation of these systems implied that "the whole is more than the sum of the parts" and that "it is not a trivial matter to infer the properties of the whole."

Ironically, when economists made the agents in their models more bounded in their rationality, they had to be smarter because these models became larger and more demanding econometrically. As macroeconomist Thomas Sargent (1993, 168) explains: "Within a specific economic model, an econometric consequence of replacing rational agents with boundedly rational ones is to add a number of parameters" because we "face innumerable decisions about how to represent decision-making processes and the ways that they are updated." This, in turn, gives additional plausibility to the suspicion that pluralism further results from cognitive limitations on the part of human inquirers. The main focus of this essay, however, has been to strengthen the case for pluralism by offering an overview of the lack of success of several monist movements in economics.

Notes

The author wishes to thank participants of the Minnesota Center for Philosophy of Science "Workshop on Scientific Pluralism" in 2002 for helpful feedback, in particular Ronald Giere, Helen Longino, and Michael Root. She is also grateful for the comments she received from the participants in the Seminar of the Amsterdam History and Methodology of Economics Group, especially John Davis. Finally, she acknowledges the generous support and kind hospitality of the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS).

1. The advertisement appeared in the *American Economic Review* 82, no. 2 (1992): xxv.

2. Information on ICAPE can be found at http://www.econ.tcu.edu/icare/home.html.

3. The text of the French students' petition is available at http://www.btinternet.com/~pae_news/texts/a-e-petition.htm.

4. The text of the professors' petition circulated in France can be found at http://www.btinternet.com/~pae_news/texts/Fr-t-petition.htm.

5. The open letter of the twenty-seven Cambridge University Ph.D. students may be accessed at http://www.btinternet.com/~pae_news/Camproposal.htm.

6. One of the organizers of the plea, Uskali Mäki (1999), clarifies that some economists who are supporters of free market (object-) economics refused to sign, whereas

some economists who are less enthusiastic about free market (object-) economics did sign. He conjectures that "when economists talk about the 'free market' of ideas, they do not use the expression in the sense in which it appears in their theories of the goods market" (504). This enables consistency, but eliminates full self-referentiality.

7. Unfortunately, space limitations allow us to consider only two forms and not others such as pluralism about methodologies, methods, and the like.

8. What follows is a very crude characterization of the transition from pluralism during the interwar period to monism after World War II, focusing mostly on the United States. The reader is referred to Morgan and Rutherford (1998a) and the contributions therein for much more detailed descriptions. We will briefly consider the developments in Europe later on in this section.

9. To be more precise, institutionalism was strengthened during the interwar period as a result of the embrace of "realism," as explained by Bateman (1998, 45): "In this new 'realistic' world of efficiency and scientific management, the institutionalists made a much bigger initial impact than the neoclassicists."

10. In particular, Roy Weintraub (1998, 228) warns: "[A]ny narrative in the history of economics of the twentieth century that employs the idea of 'increasing' mathematization' should be read with skepticism."

11. In Bateman's (1998, 48) words: "Now, instead of an ethical economics that sought to reform the nation, America had a scientific economics that sought to make the nation more efficient and to control its economy."

12. Craufurd Goodwin (1998) offers an insightful, detailed study of the influence of the demands stemming from higher education, the government, business, and foundations on the content of economics.

13. Goodwin (1998, 57) explains, "The attacks on radical economists in the 1940s and 1950s were motivated in part by reasoned fear of 'planning' by those who were scheduled to be planned and in part by unreasoned public paranoia about conspiracies of various kinds."

14. Our narrative focuses mostly on the Arrow-Debreu version, since this has come to be considered the most prestigious one. For instance, Roger Backhouse (2003) notes, "In the 1950s, however, the Arrow-Debreu model . . . came to be regarded as the definitive statement of the most rigorous version of neoclassical price theory."

15. The reader is referred to the volumes edited by Bob Coats (1996a, 2000a) for an international perspective on the developments outlined in this essay and to the contributions by Roger Backhouse (1996, 2000) and Roger Middleton (1998) for a focus on the United Kingdom. Comparisons between Europe and the United States are the focus of Bruno Frey and René Frey (1995), Bruno Frey and Reiner Eichenberger (1993), Serge-Christophe Kolm (1988), and Richard Portes (1987).

16. This observation is echoed by Gary Becker (1976, 5): "The combined assumptions of maximizing behavior, market equilibrium, and stable preferences, used relentlessly and unflinchingly, form the heart of the economic approach as I see it."

17. In fact, the earlier lack of connection had come under heavy attack from Arthur Okun (1980, 818): "Keynes . . . departed from classical microeconomics only by modifying the labor supply function to include a wage floor. But this bridge between micro and macro was defective; none of the explanations flowed directly from the implications of optimization by economic agents."

18. The reader is referred to Roy Weintraub's (1977, 1979) contributions for insightful, detailed accounts of the search for microfoundations of macroeconomics.

19. For philosopher Alan Nelson (1984), it implied that a distinction ought to be made between the problem of aggregation and the question of reduction. In particular, he

suggests that there are three possible aggregation procedures. Crudely, first, given microeconomics and aggregation principles, macroeconomics may be derived. Second, given microeconomics and macroeconomics, aggregation principles may be derived. Third, the solution favored by Nelson, given macroeconomics and aggregation principles, microeconomics may be derived. In this case, there is aggregation, but not reduction of macroeconomics to microeconomics.

20. See Heinz Kurz and Neri Salvadori (1995) for a detailed discussion of the various responses. They conclude: "While in that controversy it was conclusively shown that the view long-period neoclassical theory takes of the relationship between input use (per unit of output) and the price of the input cannot generally be sustained, surprisingly that view has not been jettisoned. . . . The disquieting fact remains that in economics propositions that have been proved wrong are still used by many (the majority?) of its practitioners" (251–52).

21. The reader is reminded that space constraints prohibit the author from covering all intricate details of these developments and is referred to the references list for further particulars.

22. According to some, it is not at all clear that the hypothesis of rational expectations is derivable from general assumptions of rationality. Frank Hahn (1986, 281) points out that to jump from "the respectable proposition that an agent will not persist in expectations which are systematically disappointed" to the proposition that "agents have expectations which are not systematically disappointed [is a] non sequitur of a rather obvious kind." And Maarten Janssen (1993, 142) shows that the rational expectations hypothesis "is an aggregate hypothesis that cannot unconditionally be regarded as being based on [methodological individualism]."

23. For simplicity, we restrict our attention to an exchange economy. However, matters get worse, not better, by the introduction of production (Kirman 1992).

24. Also see Hicks (1983, 4–5): "Our theories, regarded as tools of analysis, are blinkers. . . . Or it may be politer to say that they are rays of light, which illuminate a part of the target, leaving the rest in the dark. As we use them, we avert our eyes from things that may be relevant, in order that we should see more clearly what we do see. It is entirely proper that we should do this, since otherwise we should see very little. But it is obvious that a theory which is to perform this function satisfactorily must be well chosen; otherwise it will illumine the wrong things. Further, since it is a changing world that we are studying, a theory which illumines the right things may illumine the wrong things another time. This may happen because of changes in the world (the things neglected may have grown relatively to the things considered) or because of changes in our sources of information (the sorts of facts that are readily accessible to us may have changed) or because of changes in ourselves (the things in which we are interested may have changed). There is, there can be, no economic theory which will do for us everything that we want all the time."

25. Edgeworth's "recontracting" process is not just an alternative rationalization of perfect competition. His primary interest was not in the limiting case of perfect competition but in the indeterminacy of imperfect competition.

26. To be more precise, Edgeworth called this the "available portion" and it became known as the "core" in the era of game theory.

27. Some work after 1975 gave existence proofs for economies that are to a certain extent specialized. However, rather than supposing self-subsistence, it assumes that goods produced by firms are included in what agents may supply. This is clearly not legitimate, because rights to receive a share in the profits of a firm are not the same as the right to

dispose of the same share of the firm's physical plant and inventory. Furthermore, the existence proof now requires a stronger irreducibility assumption, as the links among individuals must not only be present, but also be strong enough to allow for high enough prices.

28. Kirman (1992, 122) observes: "[S]ince [macroeconomists] wish to provide rigorous microfoundations and they wish to use the uniqueness and stability of equilibrium and are aware of the Sonnenschein-Mantel-Debreu result, they see this as the only way out."

29. The reader is referred to Kirman (1992) for a detailed discussion.

30. In general, in order to examine models with different equilibrium beliefs and nonzero trading volume, the only solution is to consider models that lack one of the necessary hypotheses for the no-trade theorems. Other solutions have been offered that do not necessarily involve a move back to pluralism, but these have been problematic. For instance, there may be some risk-loving or irrational traders. The problem with pursuing this approach lies in deciding what kinds of irrational behavior are plausible. Or, insurance and diversification considerations may play a significant role. However, after a single round of trading based on hedging and insurance considerations, there is no further reason to trade when new information arrives because in a market of rational individuals there would be no one with whom to trade.

31. (Brandenburger 1992; Geneakoplos 1992; Rizvi 1994b). To be more precise, a distinction needs to be made between the hypotheses that events are common knowledge, that actions are common knowledge, that optimization is common knowledge, and that rationality is common knowledge. Some are required for certain results, others for different situations. Again, the reader is referred to these publications for detailed discussions.

32. In practice, it may not be possible to reach common knowledge. With an infinite state space, opinions will converge, but common knowledge of actions may never be reached.

33. For instance, Hands (1997b, 194) comments: "The plea for pluralism in economics has been a frequent refrain throughout the history of modern economic thought. This refrain has usually been voiced by those who were outside, or critical of, the mainstream in modern economics."

34. This information is available at http://www.econ.tcu.edu/icare/home.html.

35. See http://www.btinternet.com/~pae_news/texts/a-e-petition.htm.

36. See http://www.btinternet.com/~pae_news/texts/Fr-t-petition.htm.

37. In fact, Simon (1991, 385) had earlier lamented: "My economist friends have long since given up on me, consigning me to psychology or some other distant wasteland."

References

Arrow, K. J. 1959. "Toward a Theory of Price Adjustment." In *The Allocation of Economic Resources,* ed. Moses Abramovitz, 41–51. Stanford, Calif.: Stanford University Press.

Arrow, K. J., and G. Debreu. 1954. "Existence of an Equilibrium for a Competitive Economy." *Econometrica* 22, no. 3: 265–90.

Backhouse, R. E. 1996. "The Changing Character of British Economics." In Coats 1996a, 33–60.

———. 2000. "Economics in Mid-Atlantic: British Economics, 1945–99." In Coats 2000a, 20–41.

———. 2003. "The Stabilization of Price Theory, 1920–1955." In *Blackwell Companion* *to the History of Economic Thought and Methodology,* ed. J. E. Biddle, J. B. Davis, and W. J. Samuels, 308–24. Oxford: Blackwell.

Balisciano, M. L. 1998. "Hope for America: American Notions of Economic Planning between Pluralism and Neoclassicism, 1930–1950." In Morgan and Rutherford 1998a, 153–78.

Bateman, B. W. 1998. "Clearing the Ground: The Demise of the Social Gospel Movement and the Rise of Neoclassicism in American Economics." In Morgan and Rutherford 1998a, 29–52.

Beare, J. 1978. *Macroeconomics.* New York: Macmillan.

Becker, G. 1976. *The Economic Approach to Human Behavior.* Chicago: University of Chicago Press.

Blaug, M. 1992. *The Methodology of Economics, or How Economists Explain*. 2nd ed. Cambridge: Cambridge University Press.

Boland, L. 1982. *The Foundations of Economic Method*. Boston: Allen and Unwin.

Brandenburger, A. 1992. "Knowledge and Equilibrium in Games." *Journal of Economic Perspectives* 6, no. 4: 83–101.

Caldwell, B. J. 1988. "The Case for Pluralism." In *The Popperian Legacy in Economics,* ed. N. De Marchi, 231–44. Cambridge: Cambridge University Press.

Coats, A. W., ed. 1996a. "The Post-1945 Internationalization of Economics." *History of Political Economy* 28. Annual Supplement.

———. 1996b. Introduction to Coats 1996a, 3–11.

———, ed. 2000a. *The Development of Economics in Western Europe since 1945*. London: Routledge.

———. 2000b. Introduction to Coats 2000a, 1–19.

Davis, J. B. 1997. "Comment." In Salanti and Screpanti 1997, 207–11.

———. 1999. "Postmodernism and Identity Conditions for Discourses." In *What Do Economists Know? New Economics of Knowledge,* ed. R. F. Garnett Jr., 155–68. London: Routledge.

Deaton, A., and J. Muellbauer. 1980. *Economics and Consumer Behavior.* Cambridge: Cambridge University Press.

Debreu, G. 1974. "Excess Demand Functions." *Journal of Mathematical Economics* 1, no. 1: 15–23.

Dow, S. C. 1997. "Methodological Pluralism and Pluralism of Method." In Salanti and Screpanti 1997, 89–99.

———. 2002. "Pluralism in Economics." Paper presented at the annual conference of the Association of Institutional and Political Economics, November 29, 2002.

Edgeworth, F. Y. 1881. *Mathematical Physics: An Essay on the Application of Mathematics to the Moral Sciences*. London: C. Kegan Paul.

Frey, B. S., and R. Eichenberger. 1993. "American and European Economics and Economists." *Journal of Economic Perspectives* 7, no. 4: 185–93.

Frey, B. S., and R. L. Frey, eds. 1995. "Is There a European Economics?" *Kyklos* 48, no. 2: 185–311.

Geanakoplos, J. 1989. "Arrow-Debreu Model of General Equilibrium." In *The New Palgrave: General Equilibrium,* ed. J. Eatwell, M. Milgate, and P. Newman, 43–61. New York: W. W. Norton.

———. 1992. "Common Knowledge." *Journal of Economic Perspectives* 6, no. 4: 53–82.

Goodwin, C. D. 1998. "The Patrons of Economics in a Time of Transformation." In Morgan and Rutherford 1998a, 53–81.

Green, H. A. J. 1964. *Aggregation in Economic Analysis*. Princeton, N.J.: Princeton University Press.

Hahn, F. H. 1984. *Equilibrium and Macroeconomics*. Boston: Basil Blackwell.
———. 1986. “Review of Arjo Klamer, *Conversations with Economists*.” *Economics and Philosophy* 2, no. 2: 275–82.
Hakansson, N. H., J. G. Kunkel, and J. A. Ohlson. 1982. “Sufficient and Necessary Conditions for Information to Have Social Value in Pure Exchange.” *Journal of Finance* 37, no. 5: 1169–81.
Hands, D. W. 1994. “Blurred Boundaries: Recent Changes in the Relationship between Economics and the Philosophy of Natural Science.” *Studies in History and Philosophy of Science* 25, no. 5: 751–72.
———. 1995. “Social Epistemology Meets the Invisible Hand: Kitcher on the Advancement of Science.” *Dialogue* 34: 605–21.
———. 1997a. “Caveat Emptor: Economics and Contemporary Philosophy of Science.” *Philosophy of Science* 64 (Proceedings): S107–16.
———. 1997b. “Frank Knight’s Pluralism.” In Salanti and Screpanti 1997, 194–206.
———. 2001. *Reflection without Rules: Economic Methodology and Contemporary Science Theory*. Cambridge: Cambridge University Press.
Hands, D. W., and P. Mirowski. 1998. “Harold Hotelling and the Neoclassical Dream.” In *Economics and Methodology: Crossing Boundaries,* ed. R. Backhouse, D. Hausman, U. Mäki, and A. Salanti, 322–97. London: Macmillan.
Harcourt, G. C. 1969. “Some Cambridge Controversies in the Theory of Capital.” *Journal of Economic Literature* 7, no. 2: 369–405.
———. 1972. *Some Cambridge Controversies in the Theory of Capital*. Cambridge: Cambridge University Press.
Hargreaves Heap, S. 1997. “The Economic Consequences of Pluralism.” In Salanti and Screpanti 1997, 282–94.
Hicks, J. R. 1983. *Classics and Moderns*. Collected Essays in Economic Theory, vol. 3. Cambridge, Mass.: Harvard University Press.
Hodgson, G. M. 1998. “The Approach of Institutional Economics.” *Journal of Economic Literature* 36, no. 1: 166–92.
Janssen, M. C. W. 1993. *Microfoundations: A Critical Inquiry*. London: Routledge.
Kirman, A. P. 1989. “The Intrinsic Limits of Modern Economic Theory: The Emperor Has No Clothes.” *Economic Journal* 99, no. 395: 126–39.
———. 1992. “Whom or What Does the Representative Individual Represent?” *Journal of Economic Perspectives* 6, no. 2: 117–36.
Klein, L. 1946. “Macroeconomics and the Theory of Rational Behavior.” *Econometrics* 14, no. 2: 93–108.
Kolm, S.-C. 1988. “Economics in Europe and the United States.” *European Economic Review* 32, no. 1: 207–12.
Kurz, H. D., and N. Salvadori. 1995. *Theory of Production: A Long-Period Analysis*. Cambridge: Cambridge University Press.
Lucas, R. E. 1972. “Expectations and the Neutrality of Money.” *Journal of Economic Theory* 4, no. 2: 103–24.
Mäki, U. 1999. “Science as a Free Market: A Reflexivity Test in an Economics of Economics.” *Perspectives on Science* 7, no. 4: 486–509.
Mantel, R. R. 1976. “Homothetic Preferences and Community Excess Demand Functions.” *Journal of Economic Theory* 12, no. 2: 197–201.
McCloskey, D. 1982. “The Applied Theory of Price.” New York: Macmillan.
Mehrling, P. 1998. “The Money Muddle: The Transformation of American Monetary Thought, 1920–1970.” In Morgan and Rutherford 1998a, 293–306.

Middleton, R. 1998. *Charlatans or Saviors? Economists and the British Economy from Marshall to Meade*. Cheltenham, UK: Edward Elgar.
Milgrom, P., and N. Stokey. 1982. "Information, Trade, and Common Knowledge." *Journal of Economic Theory* 26, no. 1: 17–27.
Mirowski, P. 1989. *More Heat than Light*. Cambridge: Cambridge University Press.
Mirowski, P., and D. W. Hands. 1998. "A Paradox of Budgets: The Postwar Stabilization of American Neoclassical Demand Theory." In Morgan and Rutherford 1998a, 260–92.
Morgan, M. S., and M. Morrison, eds. 2001. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.
Morgan, M. S., and M. Rutherford, eds. 1998a. "From Interwar Pluralism to Postwar Neoclassicism." *History of Political Economy* 30. Annual Supplement.
———. 1998b. "American Economics: The Character of the Transformation." In Morgan and Rutherford 1998a, 1–26.
Nelson, A. 1984. "Some Issues Surrounding the Reduction of Macroeconomics to Microeconomics." *Philosophy of Science* 51, no. 4: 573–94.
Okun, A. 1980. "Rational-Expectations-with-Misperceptions as a Theory of the Business Cycle." *Journal of Money, Credit, and Banking* 12, no. 4: 817–25.
Portes, R. 1987. "Economics in Europe." *European Economic Review* 31, no. 6: 1329–40.
Rizvi, S. A. T. 1991. "Specialization and the Existence Problem in General Equilibrium Theory." *Contributions to Political Economy* 10: 1–20.
———. 1994a. "The Microfoundations Project in General Equilibrium Theory." *Cambridge Journal of Economics* 18, no. 4: 357–77.
———. 1994b. "Game Theory to the Rescue?" *Contributions to Political Economy* 13: 1–28.
Robinson, J. V. 1953. "The Production Function and the Theory of Capital." *Review of Economic Studies* 21: 81–106.
Rubinstein, A. 1998. *Modeling Bounded Rationality*. Cambridge, Mass.: MIT Press.
Rubinstein, M. 1975. "Security Market Efficiency in an Arrow-Debreu Economy." *American Economic Review* 65, no. 5: 812–24.
Salanti, A., and E. Screpanti, eds. 1997. *Pluralism in Economics: New Perspectives in History and Methodology*. Cheltenham, UK: Edward Elgar.
Sargent, T. J. 1993. *Bounded Rationality in Macroeconomics*. Oxford: Oxford University Press.
Sent, E.-M. 1997. "Sargent versus Simon: Bounded Rationality Unbound." *Cambridge Journal of Economics* 21, no. 3: 323–38.
———. 1998. *The Evolving Rationality of Rational Expectations: An Assessment of Thomas Sargent's Achievements*. Cambridge: Cambridge University Press.
———. 2001. "Sent Simulating Simon Simulating Scientists." *Studies in History and Philosophy of Science* 32, no. 3: 479–500.
———. 2004. "The Legacy of Herbert Simon in Game Theory." *Journal of Economic Behavior and Organization* 53, no. 3: 303–17.
Simon, H. A. 1991. *Models of My Life*. New York: Basic Books.
———. 1992. "Living in Interdisciplinary Space." In *Eminent Economists: Their Life Philosophies*, ed. Michael Szenberg, 261–69. Cambridge: Cambridge University Press.
———. 1996. *The Sciences of the Artificial*. 3rd ed. Cambridge, Mass.: MIT Press.
Sonnenschein, H. 1972. "Market Excess Demand Functions." *Econometrica* 40, no. 3: 549–63.
Stiglitz, J. E. 1974. "The Cambridge-Cambridge Controversy in the Theory of Capital: A

View from New Haven; A Review Article." *Journal of Political Economy* 82, no. 4: 893–903.
Theil, H. 1954. *Linear Aggregation of Economic Relations.* Amsterdam: North-Holland.
Tirole, J. 1982. "On the Possibility of Speculation under Rational Expectations." *Econometrica* 50, no. 5: 1163–82.
Varian, H. R. 1987. "Differences of Opinion in Financial Markets." In *Financial Risk: Theory, Evidence, and Implications, Proceedings of the 11th Annual Economic Policy Conference of the Federal Reserve Bank of St. Louis,* 3–37.
Weintraub, E. R. 1977. "The Microfoundations of Macroeconomics." *Journal of Economic Literature* 15, no. 1: 1–23.
———. 1979. *Microfoundations: The Compatibility of Microeconomics and Macroeconomics.* Cambridge: Cambridge University Press.
———. 1998. "From Rigor to Axiomatics: The Marginalization of Griffith C. Evans." In Morgan and Rutherford 1998a, 227–59.

6
Theoretical Pluralism and the Scientific Study of Behavior

> With the greatest of hubris, quantitative behavior genetics strives to traverse the molecular and psychological levels in one grand inferential leap. (Wahlsten and Gottlieb 1997, 166)

> Complex developmental processes, composed of millions of moments extending over many years, are not amenable to any microanalysis we currently know how to conduct. . . . Thus, mechanistic science (i.e., developmental systems science) is unlikely to yield useful information about complex behavioral problems, at least in the foreseeable future. (Scarr 1995, 155)

The social account of scientific knowledge holds that critical interaction is crucial to the epistemic acceptability of scientific content. In this essay I focus on methodological differences in the scientific study of behavior and criticisms that proponents of different approaches direct toward one another. My analysis is based on scrutiny of dozens of papers from four approaches to the study of behavior: behavior genetics, developmental systems theory, neurophysiology and anatomy, and, to the degree it is set up as an alternative to behavior genetics, research on social/environmental factors. I focused on research on two kinds of behavior—aggression and sexual orientation—reviewing first the kinds of empirical studies being done and then examining the theoretical and polemical writing to provide a guide to the intellectual contexts of these studies.[1]

The critical comments directed by researchers at practitioners of different approaches to the study of behavior belie their professed commitment to interactionist explanations of behavior. Everybody says that *both* nature and nurture contribute to/cause behavior. But most seem also to believe that only their approach successfully articulates the nature of the interaction and hence produces genuine knowledge. Such comments, and others scattered through this literature, express the belief that only one approach to studying behavior can be correct, that of the writer.[2] I analyze this polemical work not to participate in the debate over who is correct, but to help identify the assumptions and standards framing the different approaches. As these become clear, the debates seem less to require a resolution than to underscore the advantages of adopting a pluralist philosophical view of this body of research.

While the four programs of research have in common the aim of understanding behavior, they approach the study of behavior differently. In what follows, I describe the methodologies employed and review the critical interaction among proponents of the different approaches. I then extract common and distinctive assumptions informing the research thus analyzed. The most interesting assumptions differentiating the approaches concern not the causes of behavior directly, but the structure of the causal spaces within which they frame their investigative strategies. The noncongruence of the causal spaces both puts integration or unification of these programs out of reach and undermines any strategy of evaluation designed to identify the correct approach.

Methods of Investigation

One of the critical methodological differences at issue is the identification and definition of behaviors, as well as of the causal factors contributing to their expression. A second concern is how to measure and establish the relatedness of the phenomena under investigation. I have discussed the first of these with particular attention to aggression in Longino 2001. Here I will briefly review the variety of observational and experimental strategies used by researchers in each of the four approaches.

Behavior Genetics

Behavior genetics has traditionally been the application to behavior of population genetics. Granting that both the genes and the environment of an organism causally influence its behavior, the behavior geneticist asks what the genetic contribution to a given behavior is. Classical behavior genetics has been able to specify how much of the variance in expression of a trait in a population in a given environment is associable with genetic variance. Molecular behavior genetics seeks to associate behaviors with particular allelic configurations. Behavior geneticists have operationalized aggression as criminality, as noncompliant or oppositional behavior, and as verbal abusiveness and physical aggression (hitting others). These characteristics are ascribed to individuals on the basis of court records, parent or teacher interviews, responses to questionnaires, or self-report.

Classical human behavior geneticists use traditional methods of ascertaining heritability: twin studies and adoption studies. These studies measure and compare the frequencies of some behavior B in twins reared apart, in monozygotic (MZ) and dizygotic (DZ) twins (and full siblings), in adoptees and biological and adoptive parents, and in adoptive and biological siblings. These measurements enable estimates of comparative frequencies of aggressivity or of a given sexual orientation in populations of different degrees of biological relatedness. The presumption that different families constitute different (effective) environments supports treating correlations of biological relatedness with similar behavior as evidence of the heritability of the latter.

More powerful experimental methods, such as breeding, as used in classical fruit fly genetics, are not appropriate—for reasons both of ethics and of scale—for human studies. While carried out under the rubric of genetics, twin and adoption studies can at best show the extent of heritability of a trait in a given population in a given common environment, that is, how much of the difference in expression of a trait in a given environment can be ascribed to (or perhaps only associated with) genetic difference. Multivariate analysis can also assist in the decomposition of complex traits. But these studies are not capable of discriminating between polygenic and monogenic traits, nor are they able to distinguish between intrauterine and genetic effects, nor, obviously, are they able to identify genes. It should not be surprising that even a sympathetic review of behavior genetics research on aggression concludes only that there is *some* support for the claim that heredity plays a significant role in the development of (severe) antisocial behavior (Mason and Frick 1994). Notwithstanding this favorable conclusion, the review notes methodological deficiencies, some of which might lead to underestimates, others of which might lead to overestimates of the actual figure. What genetic studies can show, when methodological limitations are overcome, are the limits of postnatally induced variability in a specified range of environments. When extended by longitudinal analysis they can also show periods of relative flexibility and inflexibility, the duration of postnatal influences, and so on, but always within the specified range of environments.

The difference between classical and molecular genetics lies primarily in the means for identifying the heritable or genetic variation and assessing its contribution to phenotypic variation. So-called reverse genetics introduces specific mutations at specific sites of the genome. Applied to the behavior of fruit flies, this research strategy makes possible the identification of allelic variants that are strongly associated with behavioral variations (Hall 1994). While such manipulation of human genes is not possible, retrospective identification, through linkage analysis, of loci associated with a phenotypic trait is possible. Thus, allelic variation at the Q28 locus was associated with homosexuality in a group of men with homosexual relatives on the mother's side (Hamer et al. 1993).

Social-Environmental Approaches

Social- or environmental-oriented studies seek to establish the role in the establishment and expression of behavioral dispositions of socialization patterns, familial environments, and/or parental attitudes and interactions with their children. The social-environmental approaches use measures of aggression similar to those used by the behavior geneticists, except that some studies include a continuum of social behaviors from the antisocial to the prosocial rather than limiting themselves to the antisocial. The prosocial (or sociability) behaviors include offers of assistance, participating in cordial verbal exchange, and display or exchange of affectionate gestures. Aggression is measured by physical behavior (hitting and starting fights) and verbal behavior (lashing out in anger) as ascertained by self- and other-reports, by delinquency as ascertained in court records, by psychological classifications like Antisocial Personality Disorder and Childhood Conduct Disorder as ascertained through psychiatric diagnosis, or by hostile, confrontational interactions ascertained through direct observation.

The social-environmental approaches in psychology seek to associate distributions of one or another of these measures of aggression with variation in some environmental factor, either parental behavior, educational experience, or peer relations. Most of the studies using questionnaires or interviews report using more than one measurement strategy, such as both self- and other-report, or both self-report and court records, to enhance the reliability of behavioral ascriptions. The methods employed are both retrospective and prospective.

Retrospective methods are employed with populations whose relevant parameters are determined via interview and questionnaire or via direct observation. They include comparing the distribution of one or more postulated causal factors (abuse and neglect, foster home placement, socioeconomic status, disciplinary practices in the home) in the target population with the distribution in a control population. Some studies employ path analysis to ascertain the strength of links among multiple factors.

Prospective methods involve introducing some change into some portion of the target population and determining its effect and include, for example, coaching parents in alternate forms of discipline and observing changes in parent-child interactions or in some aggression measure in affected children.

Neurobiology (Physiology and Anatomy)

Neurophysiological and neuroanatomical approaches to behavior seek to identify and characterize aspects of the neural substrate of behavior. Many of the neurobiological studies on aggression have worked with human populations already clinically identified as antisocial. However, when studying the behavioral effects of psychoactive agents, they also use categories of physical or verbal aggressiveness similar to those employed by the other approaches. In the case of sexual orientation, self-report or Kinsey measures based on questionnaire or interview responses are used. The study methods employed are retrospective, concurrent, and prospective.

Retrospective methods include the use both of autopsies to identify neurostructural correlates of behavioral patterns attributed to the individual and of correlational studies of prison and hospital records to identify associations between brain injuries or other trauma (e.g., birth complications) and later criminal behavior. Concurrent methods include brain imaging to identify areas of brain activity related to certain thoughts or sensory stimuli. Prospective methods include animal experimentation (and follow-up clinical trials in humans) to identify the effects on behavior of organizational or activational exposure to bioactive and psychoactive substances.

Developmental Systems

The developmental systems approach has its theoretical base in embryology and developmental biology. Here the central question is how the organism develops from a single fertilized cell into a mature individual characterized by multiple and specialized organs, tissues, and behaviors. For the systems approach, genetic and environmental contributions to development are not separable. So-called developmental systems theorists argue that the entire organism must be studied as a complexly interacting nondecomposable whole. The relations between the constitutive parts of the whole are nonadditive and nonlinear. Developmental systems theorists seem not to study particular behaviors in humans, although individual researchers claim that this approach is the appropriate one to adopt either for human behavior generally or for some specific behavior (as Byne and Parsons [1993] do for sexual orientation). There is then no issue yet about defining or measuring behaviors. As in the other approaches discussed here, indeed even more so, the focus of attention is (or would be) behavior considered as a trait or property of individuals. Given the holism that characterizes this approach, it is not clear how detachable or separable from the functioning of the entire human organism specific behaviors (like “starts fights without provocation,” “hits without provocation,” etc.) would be or what strategies researchers would use to individuate and measure behaviors. In animals, by contrast, various behaviors amenable to experimental intervention have been studied. The main form of argumentation, however, seems to be either conceptual or addressed to the shortcomings of other approaches and reanalyses of their data. The direct experimental methods involve intervening in a developmental system to show that a given factor is (contrary to what might be expected from the perspective of another approach) essential to normal development of a trait or behavior. One such experiment involved preventing ducklings in utero from hearing their own or siblings' vocalizations (Gottlieb 1991). These ducklings, once hatched, failed to display species-specific responsiveness to maternal vocalizations. This is claimed to show that experience contributes to the canalization of species-specific behavior (and thus that such canalization is not, or not always, genetically controlled). An experiment studying the suckling behavior of rat pups of hypertensive mother rats is offered as demonstration of the coactional character of development—the necessity of both prenatally disposing factors and subsequent inputs from the environment (Gottlieb 1995). The developmental systems approach, unlike the other approaches, which are characterized by the performance of many similar studies reinforcing their central principles, is characterized by a few totemic experiments, treated as demonstrative of some principle of the approach and thus validating the reinterpretations offered of the studies in other approaches, especially the behavior genetic approach.

Criticism

In the critical interactions of proponents of these different approaches with each other, several crucial issues emerge in addition to problems inherent to specific methodologies: the characterization of the causal milieu, the nature of causal action/interaction, and the questions held to be important. Three of the approaches are linked in critical exchange with each other; the fourth, the neuroanatomical/neurophysiological, is often represented as naturally allied with the behavior genetic approach, but this link is more rhetorical than actual.

The behavior geneticists' twin and adoption studies are in general criticized for failing to take gene-environment interactions into account in their calculations of heritability and consequent inferences to genetic influence.[3] The particular degree of parent-offspring similarity in a trait ascertained in one environment, critics claim, may not hold in another. And if a trait is polygenic (involves a multitude of genes), studies on MZ twins will give an overestimate of the heritability of that trait in the general population. Since MZ twins share all their genes, they will a fortiori share the full set of genes sufficient for expression of a polygenic trait in a given environment. Unless the genes are linked, the probability of their co-occurrence in non-MZ siblings is much less than that of the occurrence of any single gene in the set.

Some observers have criticized studies of twins reared apart for failing to observe proper methodological precautions when testing twin pairs after they have been reunited. In this case, it cancels the assumption that the study in question is detecting spontaneous, unplanned similarities. But critics have a more principled objection to the methodology of twin and adoption studies because, they claim, it must rely on two problematic assumptions about environments (Billings, Beckwith, and Alper 1992; Lewontin 1991; Haynes 1995). The assumption that twins reared apart are reared in significantly *dis*similar environments is affirmed both in the absence of agreed-on measures of environmental similarity and dissimilarity and in the face of placement practices that give priority to finding similar homes for twins and homes for adoptees that resemble what were or would have been their natal home environments. This assumption also ignores the possibility that the degree of environmental influence may be a function of age at adoption and glosses over the similarity of uterine and early postnatal environments. The second, correlative assumption is that adoptive and biological siblings are reared in similar environments. This treats environment as gross (or shared) characteristics of the home setting, rather than individualized (or nonshared) aspects. Developmentalists also like to point out that twin studies and adoption studies have yet to identify a single gene associated with behavior. They claim that the behavior geneticists can at best show how much of a behavior is heritable in a population, but nothing about how it comes to be expressed. This complaint, however, may be more of a clue to the assumptions of developmentalists than to those of the behavior geneticists, and it is rebutted by showing the link between behavior genetics and molecular genetics, whose aim *is* such identification.

Molecular genetics is criticized less for its approach than for the assumption/expectation that anything particularly illuminating has emerged so far. Linkage analysis, claim Billings, Beckwith, and Alper (1992), has proven useful only for traits with simple Mendelian or X-linked modes of inheritance (e.g., XQ-28 and homosexuality among male relatives on an individual's maternal side, perhaps only in a limited population). Most behavioral traits will be multifactorial or polygenic, so the association of any single gene with a behavioral phenomenon will be difficult to confirm. Initial reports of genetic associations for bipolar disease, alcoholism, and schizophrenia have not been replicated, and replications may in general be difficult to obtain given the small samples generally used. The small samples are further problematic in that one or two misidentified alleles can affect the significance (positively or negatively) of a candidate association. Thus, even if there is a genetic component, it may be too fragile or too complex for current tools to isolate it.

Environmentalists are accused of valuing political correctness over knowledge and of permitting their concerns for social justice to interfere with their science.[4] This leads them, it is said, to select possible causes for investigation based on their perceived manipulability and to ignore genetic issues out of fear of their social meaning. They also fail to extract or to include genetically relevant information in their samples, as they use only biological families or do not distinguish between biological and adoptive families (DiLalla and Gottesman 1991). This strategy permits a confounding of genetic transmission with socialization. Where biological causes are separable from social ones, the behavior geneticist claims against the environmentalist that biology is a better predictor of similarity and dissimilarity in behavior than social factors. (This, of course, presumes that the causes can be separated.)

A number of authors reanalyze social-environmental studies to demonstrate this point. McGue (1994) rebuts the hypothesis that parental divorce is a major environmental factor in the divorce rate in offspring by comparing divorce rates for MZ twins in the Minnesota Twin Registry with those for DZ twins. These data, he claims, suggest that the similarity between parents and offspring is a function of genotype, not of parents' failure to model stable marital relationships. Scarr (1997) reanalyzes several studies of relation between parental rearing styles and children's school achievement to show that the factor most highly correlated with children's performance is parental IQ and education. The clear implication of both these writers, and of others,[5] is that other socialization studies will show the same defect: restriction of study variables to environmental factors fails to reveal the most likely causal factor—genes.

Developmentalists are, like the environmentalists, accused of being nonscientific. Their alleged failing, however, is not social or political correctness but the reintroduction of vitalism (Scarr 1993). The more substantive criticisms concern (1) the possibility of acquiring knowledge at all within the developmental approach, and (2) the multilevel antireductionism of the developmentalists. In leveling the first of these charges, Scarr (1997) argues that it is not possible to design studies that will identify causes of phenotypic outcomes if one is committed to treating all causes as mutually modifying one another and hence to not singling out any particular one for study. She goes on to claim that the *how* question of the developmental systems approach will produce no or little knowledge in comparison with the *how much* question of the behavior geneticists. Regarding antireductionism, Burgess and Molenaar (1995) argue that reductionism is not incompatible with acknowledging multiple levels of organization. They recognize multiple levels, but nevertheless claim that, since the most general and hence most explanatorily basic propositions will be found at the lowest level of organization, all explanation must ultimately invoke genes, and is therefore reductionistic. Clearly, different senses of reductionism are at play here. Some of the critical response to the developmentalists is defensive in character, addressed to the objections they raise to the behavior geneticist program. Scarr (1995) claims that one doesn't need mechanistic knowledge, that is, knowledge of the process of development (or even of gene action), to identify causes of a phenomenon. The *how much* question can generate causal knowledge of distributions of behavior. She also suggests that developmentalists are closet determinists. This is an odd accusation given their insistence on the contingencies of interaction. Scarr may think that an experimental understanding of the mechanisms, the how, of development involves the specification of necessary and sufficient conditions, and that the developmentalists reject behavior genetics because it provides only probabilities. But this seems just to misread the developmentalists or to misunderstand mechanisms. More temperate critics, such as Turkheimer and Gottesman (1991), stress that behavior geneticists and developmentalists are asking different questions: developmentalists are seeking to understand proximate mechanisms, while behavior geneticists want to know about the variation in a population.[6] Furthermore, by focusing their criticisms only on single variable genetic analysis, developmentalists underestimate the kinds of information that behavior geneticists can provide.

Two features of these disputes seem most to animate the polemics. One is the shared idea (*pace* the plea of Turkheimer and Gottesman) that the approaches are all asking the same question. The other is the associated idea that there is one correct way to represent the domain, the causal landscape, in and about which the question is asked. At some very general level, they *are* asking the same question: what causes behavior? But this question is both too broad and too vague to admit of any single answer. By "behavior," for example, do we mean tendencies in a population, particular episodes in the history of an individual, patterns of behavior, or dispositions to respond to situations in one way rather than another? And to get a grip on this as a causal question, behaviors must be distinguished from one another and assigned criteria of identification and strategies for determining when these are satisfied. This requirement presupposes that behaviors, at least those susceptible to causal explanation, constitute, if not natural kinds, phenomena at least stable enough to permit reidentification. Causal questions may concern a population or species (why do Xs Φ?) or an individual (why does/did X Φ?), and may be evolutionary (how did X come to be a Φing organism?) or dispositional/episodic (why do Xs Φ when S?) or mechanical (how do Xs Φ?). Particular sequences of movements are integrated differently into behavioral or physiological processes in different species. Thus, it is only meaningful to ask such questions relative to a particular behavior (or disposition) of a particular species, or relative to an individual as a member of a species. As I noted earlier, the question of causation has moved far beyond the simple nature/nurture dichotomy, as it is recognized that nature *and* nurture are causally implicated both in the evolution of particular behaviors and in their expression. And while the approaches differ on the precise implications for research of this, they do agree that, for any causal factor, one can think only in terms of its contribution to behavior relative to that of others. The disputes, therefore, are about the weight or strength of one type of factor vis-à-vis that of others, about the degree to which certain types of factors can be ignored, and about the relative value of different kinds of knowledge.

Local Epistemologies

The prominence of the disputes, however useful in bringing out substantial and methodological assumptions of the research programs, also masks the specificity and locality of the epistemologies that structure those programs. That is, the disputes about the relative importance of different causal factors suggest that there is a single question about whose correct answer there are contradictory claims. While there is a sense in which this is true, each approach is also characterized by distinctive questions, methods, assumptions, and contextual links. Rather than one element being basic or primary to the others, I see all four as developing simultaneously and as mutually co-constituting, or reinforcing. For example, behavior geneticists, in asking how much of the variance in aggressive behavior is accounted for genetically, are committed to developing methods that will tease out the heritable from the environmentally dependent. Environmentalists, in asking which of a variety of possible environmental factors are responsible for antisocial or aggressive behavior, are committed to developing methods to distinguish distinct environmental factors from one another. In spite of the charges and countercharges listed above, the research methodologies of each group are not relevant to the questions the other is asking. As each approach refines its questions and its methods for answering those questions, it constructs a domain those methods are fit to explore. This construction can be followed in the assumptions at work in the different research programs as well as in their explanatory goals.

Behavior Genetics

Perhaps the main distinctive assumptions of the behavior genetic approach are as follows:

- The effects of genes are separable from the effects of other factors.
- The causal contributions of genes to a behavior are separable from other causal influences on that behavior.
- Heritability is an appropriate measure of genetic contribution.

The less visible assumptions are those that support the use of behavior genetic methods to answer the questions posed by the approach and that are used to reject the claims of other approaches. What must the classical quantitative behavior genetic approach assume in taking twin and adoption studies as answering the questions attributed to behavior genetics? These assumptions are partially revealed by the criticisms from other approaches. One common misunderstanding of the behavior genetic approach is that it purports to identify (genetic) causes of behavior in individuals rather than the basis of variation in a trait in a population.[7] This misunderstanding is promulgated primarily in popularizations, but also sometimes by researchers themselves in discussion of the potential value of the research. Thus, one common criticism consists in pointing out the difference between causes of traits in individuals and concomitants of variation in populations, an objection directed against the assumption that identifying causes of variation in a particular environment is the same as identifying causes of traits. Another assumption concerns the nature of the other influences, and thus the nature of the causal space. The kinds of influence with which the behavior geneticists' genes compete and interact are familial and social environmental, since that is what is purportedly varied or held constant in adoption and twin studies. Since these are postnatal environmental factors, heritability (and genes) end up encompassing any prenatal influence, such as uterine environment, maternal health, and so on. Researchers sometimes discount such prenatal (but nongenetic and nonheritable) factors by including them in a noise factor, which also encompasses measurement error, trait instability, and anything else that might affect the observed variation. The behavior genetic approach thus divides the effective causal space into two main areas—prenatal and postnatal—with a little space for noise. (As noted above, the use of DZ and MZ comparisons also means that the full genome rather than single genes must be meant since MZs will share all epistatic interactions as well as all genes. This, and the fact that MZ twins also share the same uterine environment, accounts for the criticisms that this method overestimates heritability.)

The use of twin and adoption studies also implies assumptions about how the (postnatal) environment is understood. Environment turns out to mean *shared* family characteristics. One such characteristic is socioeconomic status (SES), which includes education and income levels of parents; urban, rural, or suburban location of rearing; and other such factors. Another is professed attitudes of parents to education, discipline, and other factors relevant to childhood experience. Nonshared environmental factors like age and birth order effects, as well as the variation in interaction of individual personalities in different combinations, pose a challenge to this partition of the causal space.

Behavior geneticists differ in their responses to this challenge. Some treat such phenomena as noise that can be neglected. To some degree this is the effect of Scarr's (1987) proposal of the "average expectable environment" as the background against which studies should be interpreted. Scarr distinguishes between two kinds of environment: the damaging environment and the nondamaging (or "average expectable") environment. Damaging environments are those that can interfere with the individual's genetically determined development and thus are causally influential. Whatever differences there may be in the average expectable (or nondamaging) environment are fluctuations that cancel each other out and permit human individuality (which is a function of genetic individuality) to be expressed. Scarr (1992) goes even further in erasing nonshared factors from the causal landscape. She proposes that individual variations within the "average expectable environment" are actually the effects of the genetic makeup of the individuals whose behavior is the object of study. Individuals create their own nonshared environments because their genetically determined differences evoke different responses.[8] Robert Plomin (Plomin, Owen, and McGuffin 1994), on the other hand, sees the gap left between genes and the macroenvironmental factors controlled for in twin and adoption studies as causally effective nonshared environment, and proposes that one of the values of behavior genetic methods is that they show both the limits of genetic influence and the extent of nonshared environmental influence. The presumption is that macro-level factors have been ruled out by the design of the experiment; thus, the variance not accounted for genetically must be ascribed to the micro- or nonshared environment.

The application of molecular genetics methods to behavior also relies on assumptions, some testable, others not. The significance of any association found via linkage analysis depends on assumptions about the base rate of the trait in the general population. A reliable estimate of base rate requires clear criteria of identification. While there are estimates of base rates for many of the physical and mental disorders to which molecular methods are usually applied, disorders such as susceptibilities to specific cancers or schizophrenia, aggression is not yet precisely enough defined to obtain a base rate, and there are ongoing debates about both the modes of sexual orientation and the base rates for those modes.[9] In addition, while techniques for identifying genes and their allelic variants are improving all the time, establishing the nature of their contribution to any given phenotype is more complicated. This is especially so in the case of behavior where allelic variation can be associated with only a small portion of the variance in a population. The assumptions regarding base rate and adequacy of sample size have proven the undoing of many purported associations—for example, of given alleles with alcoholism and with bipolar disorder—which have failed to be replicated in subsequent studies. The interpretation of molecular (or traditional) approaches using experimental animals such as rodents or drosophila fruit flies as relevant to humans must also assume that the behaviors they study are sufficiently analogous to human behaviors to warrant inferring from one species to the other.[10]

Both the classical and molecular genetics approaches must assume that they are addressing well-defined traits. Molecular genetics is prompting some conceptual refinement through the "one gene, one disorder" approach, which breaks down a complex (or composite) trait like mental retardation into distinct (but similar) traits individuated by distinct etiologies.[11] "One gene, one disorder" has its complement in the multivariate analysis of twin studies that can be used to identify co-occurring traits that may be constituents in whole or in part of some more grossly identified trait. These approaches lie behind the identification of impulsivity as a lower-level disposition that may be a heritable and genetically based component of aggressivity. Thus, the assumption that there is a genetic influence on behavior motivates both traditional and molecular researchers to refine their concepts to develop constructs that can be more reliably associated with genes. That is, the general assumption drives a search for ways to realize the subordinate assumption regarding the individuation and identification of traits.

Social-Environmental Research

While behavior genetic researchers tend to see their work as refuting or blocking claims regarding the role of gross SES factors or of factors common to all members of a household, many of the socially oriented studies investigate more fine-grained features of the environment, the nonshared environment. Their research design is intended to discriminate among the potential causal factors within the environment, not to discriminate between genetic and environmental contributions to behavior. Researchers studying and intervening in social and familial interactions may, but need not, assume that these factors are causally independent from the subjects for whom they constitute an environment. Since the interventions tend to focus on changing parental behavior, nonshared environment is effectively treated as independent. Diana Baumrind (1991, 1993) defends this focus on the grounds of the greater power of the adults in parent-child relationships. This, of course, assumes that the adults are not unconsciously responding to features of the children's behavior or personalities, an assumption directly contradicting Sandra Scarr's proposal that nonshared environment, that is, those differential parental behaviors, can be treated as a genetic effect. Just as genetically oriented researchers must assume uniformities of effective environment (e.g., the "average expectable environment"), environmentally oriented researchers must discount genetic variation in their subjects, assuming that subjects are sufficiently endogenously uniform or that genetic variation in their subjects averages out and does not interact systematically with the environmental/experiential variables being studied.[12] Large samples sorted for sex and SES and matched control populations help to eliminate some confounding variables, but these tend to be the same sorts of causal influence, such as family SES, for which the genetically oriented researchers control. Neither genetically oriented nor environmentally oriented approaches have the resources to control for the influences the other is studying.

Neurobiology

The methods employed in the neurobiological approaches involve a number of assumptions. Neuroimaging techniques assume that the brain areas showing greater glucose metabolism during a particular thought process are causally involved in that thought process (and not epiphenomenally or incidentally affected by it). The assumption is that if a brain area is involved in the thought process it will show glucose activity. If additional conclusions (e.g., that the activity causes the process) are drawn about the involvement of that area in behaviors related to the thought process, then further assumptions are also being made. These would include assumptions about the modality of the relationship (causal or merely associational) between the process and behavior and, via the prior assumption, about the relationship between the region of brain activity and the behavior in question. Thus, for example, studies that correlate areas of brain activity with hostile thoughts and, through those thoughts, with aggressive behavior may be described as finding associations or causal relations, depending on what assumptions are used in presenting the findings. The proposal that neural structures identified through autopsy are causally related to a behavior rests on the assumptions that the anatomical correlates have a function related to the behavior with which they are correlated and that the development of these anatomical correlates preceded rather than succeeded the relevant behaviors. To the extent that animal experimentation is relied on for support of these assumptions, then both structural and functional cross-species uniformity and analogy of the human and animal behaviors is assumed.[13] The clinical research that attempts to identify both physiological effects and behavioral effects of neuroactive pharmaceuticals need not make the same kinds of causal assumptions that the imaging and autopsy studies do. Their interest is in establishing that the administration of certain chemicals is followed by certain desirable states, and not by undesirable ones.[14] To the extent the work relies on psychiatric classifications, however, it is assuming those categories pick out genuine kinds (Coccaro 1993; Coccaro et al. 1994). Some of the researchers in this field also attempt to establish connections with heritability and genetic research and are thus implicitly committed to many of the assumptions of the behavior genetics program discussed (Coccaro et al. 1997).

Developmental Systems

Developmental systems theorists assume that interactivity of causes means that separation of causes is never possible. Their approach also assumes that humans and nonhumans are sufficiently similar with respect to overall developmental processes that conclusions based on animal experimentation carry over to humans. Finally, their approach implies that to understand development it is essential to understand intraindividual processes. In addition, the polemical writings suggest that the only interesting biological question is a developmental question. Their arguments are not directed at the claims of genetic (or environmental) contribution. Instead, these approaches are dismissed in arguments that rely on premises asserting that such and such a technique cannot illuminate development, that is, the processes whereby an individual of a given species comes to express a particular trait, whether that be eye color or a certain kind of intellectual ability. If the point of such arguments is to delegitimate the use of those methods (such as twin studies) in the study of behavior, then they must assume that the only interesting question is this (narrowly defined) developmental one. But there are a number of questions behavior geneticists can address that are not developmental.[15] Developmental systems theory has still to produce a full repertoire of empirical methods in order to produce data that can be used to support their particular theoretical proposals. Thus, many of the assumptions that would facilitate judgments of evidential relevance are not yet in play. Gottlieb's style (1991, 1995) suggests a reliance on experimental reasoning rather than on the collection of statistical data. The experiments described above are adequate to show the necessity of specific multiple factors in particular cases, but are not powerful enough to support views about the mode of their interaction.

Pluralism

Researchers from the different approaches do not differ in their ways of identifying the behavioral phenomena with which they are concerned. They all treat the behavioral phenomena as expressions of dispositions inherent in individuals.[16] They would all be open to identifying a more basic component or precursor of the behavioral disposition whose etiology they seek to identify.[17] And researchers on sexual orientation from behavior genetics and neurobiology, as well as developmental systems, share a pragmatic aim of normalizing homosexuality through naturalizing it.[18] Nevertheless, some of the assumptions differentiating these approaches from each other involve differences so deep as to make their unification, whether through integration or through the elimination of all but one compatible set of hypotheses and approaches, impossible.

One set of these assumptions has to do with the conception of causal relations employed, the other with the structure of the domain of investigation. While all give lip service to a more complex and interactive notion of causality, three of the approaches design their studies using linear, single-factor conceptions of causality. These studies seek to establish correlations between a specified behavior and a state or feature precedent to and independent of that behavior (or a signifier of such a feature). These correlations (subject to statistical testing) serve as the empirical basis for ascribing causal relevance to the state or feature. Some environmental research also works with multiple-factor conceptions when path analysis suggests several (social or environmental) variables may be contributing to the expression of a given behavior. The relation of each variable to the effect in question is still, however, linear, and their relation to each other additive. Thus, each of these three approaches investigates a distinct causal system. These causal systems intersect in complex organisms (from drosophila to humans) and so modify one another's effects. But the approaches dedicated to understanding a given causal system are not designed to investigate their interactions. The remaining developmental approach treats causality as nonlinear: all factors not only act together but also modify one another in a highly complex feedback process. This approach insists on the interaction of factors, but treats them as part of one system, the developmental system, rather than as elements of distinct systems that intersect. Hence, from a developmental systems perspective, other research programs that seek to establish the contribution of independent kinds of factors will come up empty-handed.

The second set of assumptions further guarantees that research within each approach will have little or no bearing on research from one of the other approaches. These assumptions concern the structure of the domain, that is, the space of possible causes assumed by the application of any given methodology. Figure 6.1 displays the possibilities prior to the imposition of structure. Each approach employs methodologies that require particular ways of understanding the causal space. Some phenomena regarded as causally active in one approach are simply not included in another. These differential selections result in incongruous causal spaces.

Quantitative behavior genetics divides the space between genes and shared family environment, that is, such gross family characteristics as socioeconomic status, since these are what are intended as contrasts in twin and adoption studies. The variation of genotype is studied against the background of a presumed stable macroenvironment, or a presumed shared genotype is studied in a set of different environments. Nonshared environment, such as differential parental treatment of siblings, gets treated as noise or as itself a genetic effect. While classical behavior genetics (Figure 6.2) can implicate genotypes, molecular genetics research (Figure 6.3) seeks to identify individual genes. The finding of significant heritability of some trait by classical behavior genetic methods is a signal to search for the relevant gene(s) using molecular methods. At that stage the point is to differentiate among candidate genes or gene complexes, not between genes and some other kind of factor. While physiological and intrauterine factors are incorporated within the scope of the genetic by the methods employed by classical behavior genetics, they are environmental factors from the molecular point of view.

Environmentalists (Figure 6.4) treat the space differently. They, too, divide it between genetic and environmental, but for them the genetic side includes the intrauterine and is treated as generic or uniform, while the environmental side is the location of effective variation, both in gross characteristics such as family SES and in more fine-grained ones, such as disciplinary practices, forms of endearment and other aspects of one-to-one interaction, maternal gender stereotypes and attitudes, and so on.

The neurobiologists (Figure 6.5) can be seen as opening up the middle area between genes and environment by focusing on the organic—anatomical

Genotype 1 [allele pairs]	Genotype 2 [whole genome]	Intrauterine environment	Physiology [hormone secretory patterns; neurotransmitter metabolism] Anatomy [brain structure]	Nonshared environment [birth order; differential parental attention; peers]	Shared (intrafamily) environment [parental attitudes toward discipline; communication styles; abusive/nonabusive]	Socioeconomic status [parental income; level of education; race/ethnicity]

Figure 6.1. Causal space

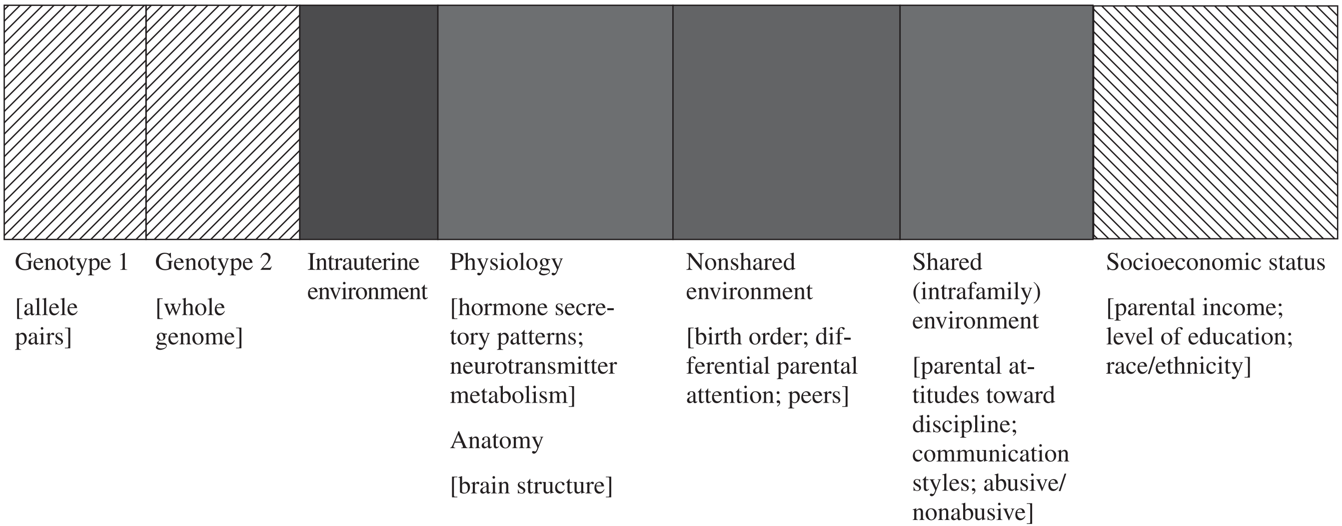

Figure 6.2. The potential causal space for behavior genetics. Diagonal lines represent active space; although in principle this could include features of shared (intrafamily) environment, in practice these are not taken into account or are subsumed under the socioeconomic status categories. Solid boxes indicate inactive space (if empty, either randomly distributed or effect of genotype).

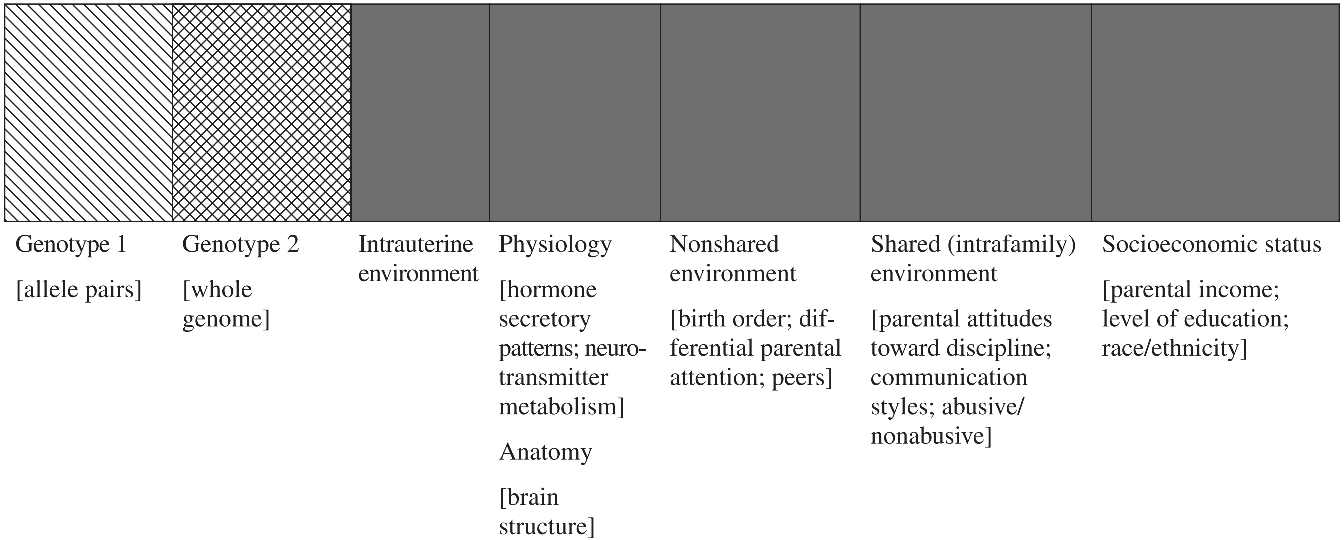

Figure 6.3. The causal space for molecular behavior genetics. Diagonal lines represent active space; the trellis pattern indicates possible interactions with factors in active space; the solid squares are inactive space.

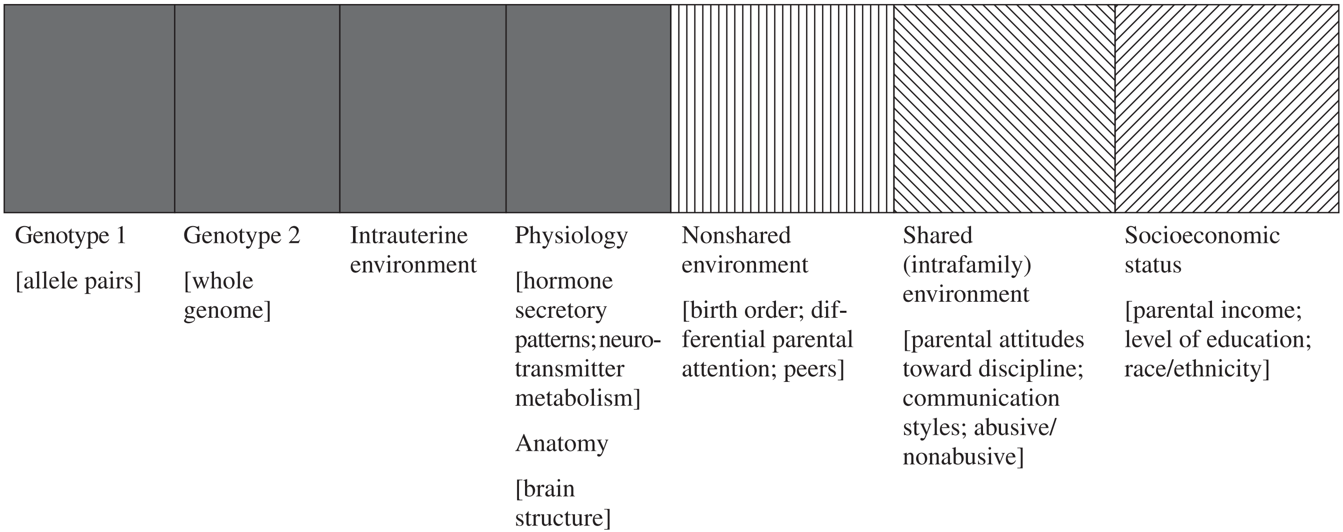

Figure 6.4. The potential causal space for social/environmental approaches. Vertical and diagonal lines are active space; solid squares are inactive space (treat as uniform background; variation randomly distributed).

and physiological—substrate. While many neurobiologists may think that this substrate is genetically determined, and behavior geneticists certainly act as though the organism inside its skin is just a direct expression of the genome, sufficient experimental work exists on the plasticity of neural structures that such a supposition is clearly just an assumption.[19] The organic, neural substrate must be treated as an independent factor relative to the organism's genotype, that is, in studying its effects researchers must treat it in isolation from the factors that produced it. But this will perforce ignore feedback effects that reset the system. Attempts to read significance into neuroanatomical correlates of sexual orientation treat the neural structures as potential causes rather than as effects of such orientation or as joint effects of a common cause. Thus, the effective causal space of the neurobiological approach is occupied by neural structures, systems and processes, with genes, environment, and other developmental factors interacting in a basically neutral background. The interest here is in the role and function of developed structures, not in the processes leading to their development. These can be ignored and hence play no role in the space investigated by the neurobiologists.

For the developmental theorists (Figure 6.6), all of these are potentially in the same causal space as are the interactions among them. Fine-grained environmental factors (like birth order) will interact differently with one set of endogenous factors as opposed to an alternative, that is, differently with one genotype as distinct from another. And these interactions will themselves vary depending on other factors, such as cultural values, differential social reward systems, and so on. For developmental theorists, the environment in and with which genes interact includes the intracellular and extracellular physiological environment in addition to the social and other factors external to the individual, while behavior geneticists and social environmentalists treat physiological factors as effectively equivalent to genes. This heterogeneity of the causal space prompts the behavior geneticists' complaint that the developmental approach is methodologically unrealistic. How is it possible to develop a model of intelligence differences, or sociality differences, if all those factors must be included? Indeed, developmental theorists may not be interested in modeling a single trait as distinct from articulating a general model of the interdependence of all the factors in the causal space. That is, they are less interested in understanding particular behaviors than in understanding the matrix in which any behavior develops. As noted above, to date its empirical program has consisted in demonstrating the insufficiency of single factors and a requirement for multiple factors. It has not developed resources for studying the inter- (or co-) action of these factors. And according to some of its critics, this is not even possible.

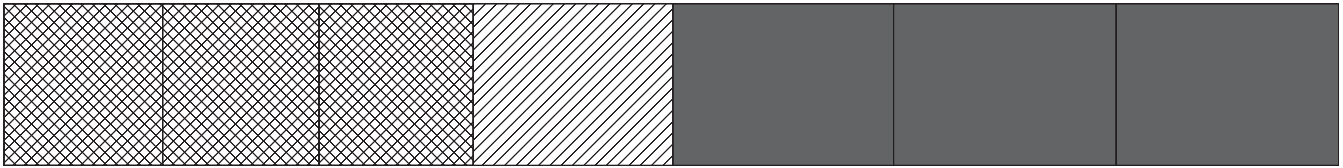

Figure 6.5. The potential causal space for physiological and anatomical research. Diagonal lines are active space; the trellis pattern represents possible determinants of active space; solid squares are inactive space, although some studies control for shared intra- and interfamily environments.

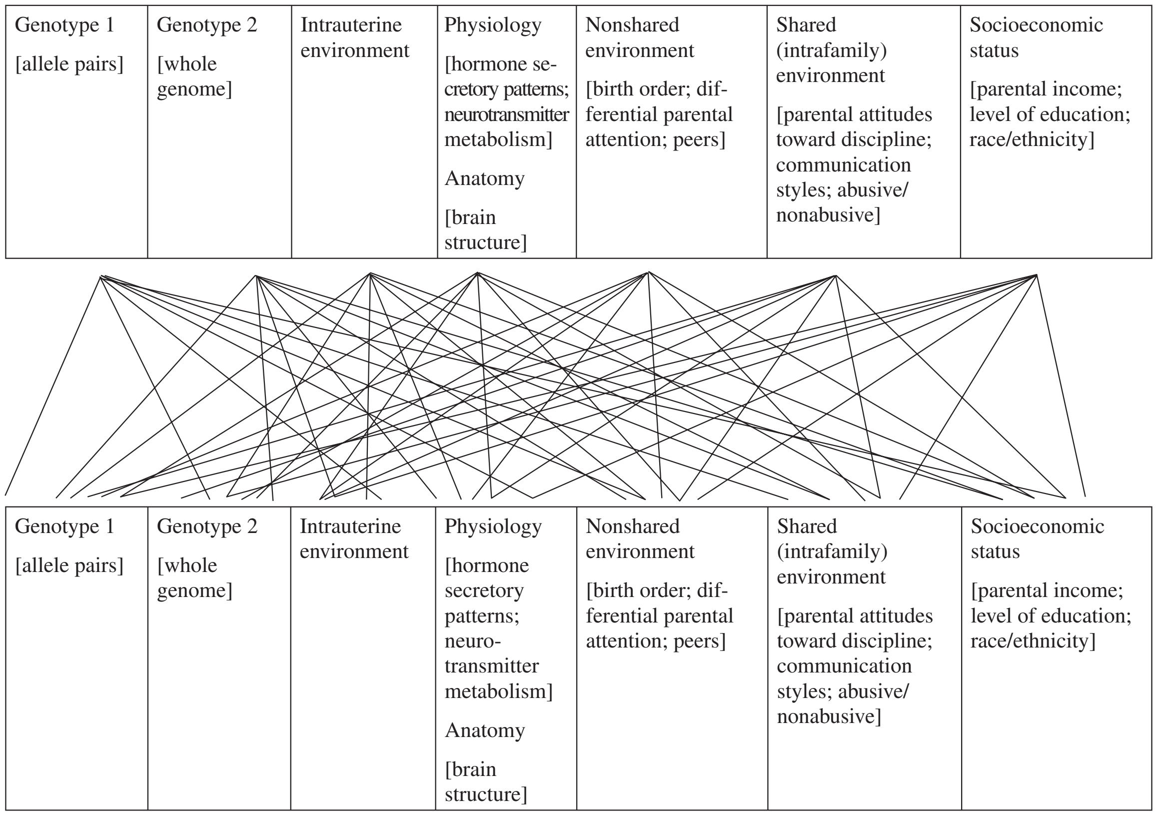

Figure 6.6. Developmental systems theory: a partial representation of the causal relations. Each type of factor can affect each other type of factor and affect how each other type of factor influences higher level organismic states.

Even though there is, from one point of view, a common phenomenon to be understood—behavior, or rather, specific behaviors or behavioral patterns such as aggressivity or a particular sexual orientation—each approach brings with it a prior and distinctive representation of the domain of investigation. Each treats different subsets of possible factors as possible causes among which to discriminate and each (with the exception of developmental systems theory) represents different areas as inactive or causally irrelevant. Each begins its research in a different causal field. Although for investigative purposes each must proceed as though its favored distinctive causal field were a complete characterization of the domain, each is really partial. The partiality is not a simple part-whole relation because the approaches cannot simply be integrated into a single picture. Since the methodologies developed within the different approaches are designed to differentiate between causal influences in a particular space of possibilities (biologically transmitted factors versus shared family characteristics; one genotype versus another genotype; one social influence versus another social influence; one brain area versus another brain area), and since the spaces differ from approach to approach, they do not have the capacity to discriminate between the causes proper to the different approaches (genotype G versus social experience S). While each individually can establish that its proper cause contributes something to the expression of a behavior (at least under certain conditions), superimposing the causal landscapes one upon another produces a blur instead of a basis for unification.

The situation is, therefore, analogous to an example Nancy Cartwright (1983) uses to talk about the ceteris paribus character of laws. We have a law from one domain of research that the higher the altitude, the lower the temperature at which water reaches boiling point; and we have a law from another domain that increasing the saline content of water increases the temperature at which it reaches boiling point. But no law explains what happens as we both move up the mountain and increase the saline content because each belongs to a different explanatory system, and each holds only as long as there are no changes in other conditions. In reality, of course, the water is always at some altitude and of some degree of salinity. In physics and chemistry we isolate the various causal systems involved and abstract them from the material context they jointly occupy in order to provide explanations, in order to know anything about those systems.

Conclusion

The features of competing behavioral research programs made salient by focusing on their critical interactions are better accommodated in a framework that is open to pluralism than one constrained by a commitment to monism. That is, at least in science studies or philosophy of science, the multiplicity of approaches is usefully addressed not by comparative evaluations directed at selecting the uniquely correct one, but by appreciating the partiality of each. If their partiality is accepted, each approach can be seen to produce some knowledge of behavior by answering the questions distinctive of it with methods that are also distinctive. But none of the approaches can yield a complete account. Conflict develops when approaches seek to displace one another, but this is more a function of external pressures than of features intrinsic to the research itself. I have suggested that in the case of the set of approaches studied here, diversity is generated partially by focus on specific kinds of causal factors and primarily by the representation of the causal space. Researchers must partition the causal space in order to begin formulating research strategies.[20] This partition, arbitrary as it is, nevertheless initiates a series of differentiating moves that particularize the knowledge that will be produced and the epistemologies within which it is so ratified. Each approach can produce partial knowledge. In concert, they constitute a nonunifiable plurality of partial knowledges. Whether a unified representation of the etiology of these behaviors is possible depends on whether the factors generating plurality change. While the status of knowledge need not be withheld on account of plurality, the dependence of the current understandings on epistemological factors (such as the representation of causal space or the classification of behavior) or ontological ones (such as the social environment) that might change in the future means that, in addition to being partial and plural, they are also provisional.

In spite of the fact that the approaches do not have the resources to show that one among them is correct in comparison to the others, the critical interactions that take place among their proponents do have consequences for the understanding of behavior. One consequence is that investigative resources proper to each approach are sharpened as a response to challenge and criticism. A second is that the limitations of each approach are made evident by the articulation of questions that they are not designed to answer. As prescribed by the social approach to scientific knowledge, knowledge is produced through critical interactions among adherents to different perspectives as well as through their interactions with the empirical domains under investigation. These interactions enable the refinement of methodologies, the clarification of concepts, the design of experiments and studies to control for causal factors demonstrated by others. All this makes for more knowledge, which, judged by means of the evaluative tools available within each perspective, is also better knowledge. Whether it is good knowledge or knowledge worth having in any more general context depends on assessment at a different level. This depends on (1) the relevance and connections that can be made to other research programs and (2) the relevance that can be established to social interests, that is, the uses that might be made of the knowledge. These matters, which are pragmatic and political as well as technical, must be determined by different standards than are involved in evaluations of the relative empirical adequacy or even correctness of the different approaches.

Notes

I am grateful to Ken Waters, Pete McGee, Steve Kellert and the participants in the Minnesota Center for Philosophy of Science "Workshop on Scientific Pluralism" in 2002 for comments on this paper. I am also grateful to Steve Fifield, who helped me with the research, and to the University of Minnesota Graduate School for the grant-in-aid that enabled me to employ him as a research assistant for the project. Additional research for the paper was conducted with the assistance of grant SBR9730188 from the National Science Foundation.

1. In previous papers (Longino 2001, 2002b) I have discussed the sociopolitical climate in which these behaviors are studied, some general features of behavioral research, and the difficulties of defining the behaviors under study. The social approach alluded to is that advanced in Longino 2002a.

2. Two exceptions to this generalization are the attempts of some researchers utilizing the neurophysiological approach to link up with work in behavior genetics and the mutual citation by some developmental systems and social environmental researchers. See Coccaro et. al. 1997.

3. The locus classicus for this argument is Lewontin 1974. Other versions are found in Gould 1981; and Lewontin, Rose, and Kamin 1984. The argument is repeated in Gottlieb 1995.

4. Scarr 1994. Weinrich (1995) argues that, in the case of sexual orientation, the political concerns of opponents to biological research on sexual orientation are misplaced and that these critics misunderstand the biological approach.

5. See, for example, DiLalla and Gottesman 1991.

6. If behavior geneticists really limited themselves to such claims, it is hard to see how disputes could arise. The controversial issues concern the inferences to genes and conflicting ideas of what the real questions are.

7. One research team does think it is possible to apply behavior genetic methods to draw conclusions about individuals. Burgess and Molenaar (1993) claim the procedure is equivalent to the use of factor analysis in interpreting results of personality tests.

8. An environmentalist or developmental systems theorist might cavil that "difference" is equivocated on here: the difference that behavior genetic methods can study are differences in the expression of a trait in a population, while in this context they seem to be traits of individuals.

9. See Longino 2001.

10. For discussion and criticism of this assumption, see Schaffner 1998.

11. Plomin uses the expression "complex" for traits that turn out to be families of symptomatically similar but distinct repertoires with distinct etiologies, such as the multiple forms of mental retardation (Plomin, Owen, and McGuffin 1994).

12. Sesardic (1993) highlights such assumptions in an attempt to overturn widely shared reservations about behavior genetics.

13. In the case of LeVay's (1991) work, this worked against his conclusion that the structural features he found in the brains of homosexual men were related to their sexual orientation, as the hypothalamic nuclei in rats that seem to influence sexual mate choice are not the same as the human ones in which LeVay found size differences.

14. This is the case with Heiligenstein et. al. (1992), whose research was sponsored by Eli Lilly.

15. *Pace* Scarr's presidency in 1991 of the Society for Research in Child Development.

16. Some socialization researchers treat behavior more relationally, but when drawn into debate with genetically oriented colleagues, they fall back on an internalist, individualist conception of behavior.

17. Impulsivity in the case of aggression, and childhood cross-gender behavior in the case of homosexuality.

18. They differ, of course, in their conceptions of naturalization.

19. The research showing the mutability of rate of testosterone secretion in response to the same stimulus after changes in social or environmental situation demonstrates the complexity of the interactions of the various systems involved. Similar results are being obtained for the serotonergic system. See Yeh, Fricke, and Edwards 1996.

20. This case, then, is most similar to that of research on the levels and forces of selection as analyzed by Waters (1991). For an account of varieties of pluralism, see Longino 2002a.

References

Baumrind, D. 1991. "Parenting Styles and Adolescent Development." In *Encyclopedia on Adolescence,* ed. J. Brooks-Gunn, R. Lerner, and A. Peterson. New York: Garland.

———. 1993. "The Average Expectable Environment Is Not Good Enough: A Response to Scarr." *Child Development* 64: 1299–317.

Bechtel, W., and R. Richardson. 1993. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research.* Princeton, N.J.: Princeton University Press.

Billings, P., J. Beckwith, and J. Alper. 1992. "The Genetic Analysis of Human Behavior: A New Era?" *Social Science and Medicine* 35, no. 3: 227–38.

Burgess, R. C., and P. Molenaar. 1993. "Human Behavioral Biology." *Human Development* 36: 36–54.

———. 1995. "'Some Conceptual Deficiencies in "Developmental" Behavior Genetics': Comment." *Human Development* 38, no. 3: 159–64.

Byne, W., and B. Parsons. 1993. "Human Sexual Orientation: The Biologic Theories Reappraised." *Archives of General Psychiatry* 50, no. 3: 228–39.

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

Coccaro, E. 1993. "Psychopharmacologic Studies in Patients with Personality Disorders: Review and Perspective." *Journal of Personality Disorders* Supplement (Spring): 181–92.

Coccaro, E. F., J. M. Silverman, H. M. Klar, T. B. Horvath, and L. J. Liever. 1994. "Familial Correlates of Reduced Central Serotonergic System Function in Patients with Personality Disorders." *Archive of General Psychiatry* 51: 318–24.

Coccaro, E. F., C. S. Bergeman, R. J. Kavoussi, and A. D. Seroczynski. 1997. "Heritability of Aggression and Irritability: A Twin Study of the Buss-Durkee Aggression Scales in Adult Male Subjects." *Biological Psychiatry* 41, no. 3: 273–84.

DiLalla, L., and I. Gottesman. 1991. "Biological and Genetic Contributors to Violence: Widom's Untold Tale." *Psychological Bulletin* 109, no. 1: 125–29.

Feldman, M. W., and R. C. Lewontin. 1975. "The Heritability Hang-Up." *Science* 190: 1163–68.

Gottlieb, G. 1991. "Experimental Canalization of Behavioral Development: Theory." *Developmental Psychology* 27, no. 1: 4–13.

———. 1995. "Some Conceptual Deficiences in 'Developmental' Behavior Genetics." *Human Development* 38, no. 3: 131–41.

Gould, S. J. 1981. *The Mismeasure of Man*. New York: W. W. Norton.

Hall, J. 1994. "The Mating of a Fly." *Science* 264, no. 5166: 1702–14.

Hamer, D., S. Hu, V. L. Magnuson, N. Hu, and A. M. L. Pattatucci. 1993. "A Linkage between DNA Markers on the X Chromosome and Male Sexual Orientation." *American Journal of Human Genetics* 53, no. 4: 844–52.

Haynes, J. D. 1995. "A Critique of the Possibility of Genetic Inheritance of Homosexual Orientation." *Journal of Homosexuality* 28, no. 1–2: 91–113.

Heiligenstein, J., E. F. Coccaro, J. H. Potvin, C. M. Beasley, B. E. Dornseif, and D. N. Masica. 1992. "Fluoxetine Not Associated with Increased Violence or Aggression in Controlled Clinical Trials." *Annals of Clinical Psychiatry* 4, no. 4: 285–95.

Jackson, J. F. 1993. "Human Behavioral Genetics, Scarr's Theory, and Her Views on Interventions: A Critical Review and Commentary on Their Implications for African American Children." *Child Development* 64: 1318–32.

LeVay, S.. 1991. "A Difference in Hypothalamic Structure between Heterosexual and Homosexual Men." *Science* 253, no. 5023: 1034–37.

Lewontin, R. 1974. "The Analysis of Variance and the Analysis of Causes." *American Journal of Genetics* 26: 400–411.

———. 1991. *Biology as Ideology: The Doctrine of DNA*. Concord, Ontario: Anansi.

Lewontin, R., S. Rose, and L. Kamin. 1984. *Not in Our Genes: Biology, Ideology, and Human Nature*. Harmondsworth, England: Penguin Books.

Longino, H. E. 2001. "What Do We Measure When We Measure Aggression?" *Studies in History and Philosophy of Science* 32, no. 4: 685–704.

———. 2002a. *The Fate of Knowledge*. Princeton, N.J.: Princeton University.

———. 2002b. "Behavior as Affliction: Common Frameworks of Behavior Genetics and Its Rivals." In *Mutating Concepts, Evolving Disciplines: Genetics, Medicine, and Society,* ed. L. S. Parker and R. A. Ankeny, 165–87. Dordrecht: Kluwer.

Mason, D., and P. Frick. 1994. "The Heritability of Antisocial Behavior: A Meta-Analysis of Twin and Adoption Studies." *Journal of Psychopathology and Behavioral Assessment* 16, no. 4: 301–23.

McGue, M. 1994. "Why Developmental Psychology Should Find Room for Behavioral Genetics." In *Threats to Optimal Development: Integrating Biological, Psychological, and Social Risk Factors,* ed. C. N. Alexander et al., 105–19. Hove, England: Lawrence Erlbaum Associates.

Plomin, R., M. Owen, and P. McGuffin. 1994. "The Genetic Basis of Complex Human Behavior." *Science* 264, no. 5166: 1733–39.

Scarr, S. 1987. "Three Cheers for Behavior Genetics: Winning the War and Losing Our Identity." *Behavior Genetics* 17, no. 3: 219–28.

———. 1992. "Developmental Theories for the 1990s: Development and Individual Differences." *Child Development* 63, no. 1: 1–19.

———. 1993. "Biological and Cultural Diversity: The Legacy of Darwin for Development." *Child Development* 64: 1333–53.

———. 1994. "Why Developmental Research Needs Evolutionary Theory: To Ask Interesting Questions." In *International Perspectives on Psychological Science,* vol. 1, *Leading Themes,* ed. P. Bertelson and P. Helen, 170–78. Hove, England: Lawrence Erlbaum Associates.

———. 1995. "Commentary." *Human Development* 38: 154–57.

———. 1997. "Behavior-Genetic and Socialization Theories of Intelligence." In *Intelligence, Heredity, and Environment,* ed. R. Sternberg and E. Grigorenko, 3–41. New York: Cambridge University Press.

Schaffner, K. 1998. "Genes, Behavior, and Developmental Emergentism: One Process, Indivisible?" *Philosophy of Science* 65, no. 2: 276–89.

Sesardic, N. 1993. "Heritability and Causality." *Philosophy of Science* 60, no. 3: 396–418.

Turkheimer, E., and I. Gottesman. 1991. "Individual Differences and the Canalization of Human Behavior." *Developmental Psychology* 27, no. 1: 18–22.

Wahlsten, D., and G. Gottlieb. 1997. "The Invalid Separation of Effects of Nature and Nuture: Lessons from Animal Experimentation." In *Intelligence, Heredity, and Environment,* ed. R. Sternberg, E. Grigorenko, et al., 163–92. Cambridge: Cambridge University Press.

Waters, C. K. 1991. "Tempered Realism about the Force of Selection." *Philosophy of Science* 58, no. 4: 533–73.

Weinrich, J. D. 1995. "Biological Research on Sexual Orientation: A Critique of the Critics." *Journal of Homosexuality* 28, no. 1–2: 197–213.

Yeh, S.-R., R. A. Fricke, and D. Edwards. 1996. "The Effect of Social Experience on Serotonergic Modulation of the Escape Circuit of Crayfish." *Science* 271: 366–69.

7
A New/Old (Pluralist) Resolution of the Mind-Body Problem

The identity theory, or thesis, of mind and body holds that the psychological, mental processes of an animal—sensations, perceptions, emotions, images, memories, inferences, thoughts, decisions, and volitions—are identical with physical processes in the animal's body. The thesis is an element of the more general theory of physicalism, the theory that all empirical phenomena are physical phenomena. Dualism is the competing view that any such psychophysical identity is false, and that mental processes are (at best) merely correlated with physical.[1] The mind-body problem is in reality a cluster of problems that includes the following: Which of the two competing views is true? Can we know which is true, and if so, how? Is the question of which is true a spurious, pseudo question? Is it scientifically or philosophically important to answer any of these questions? This essay is concerned in some measure with all these questions, but it focuses on the first.

I think most cognitive *scientists* would initially maintain that they subscribe to the identity theory but having seen my cluster of problems would say that they subscribe to it as a methodological assumption, or that the question is spurious, or that it is unimportant. I think most *philosophers* of cognitive science subscribe to the identity thesis, and not merely as a methodological assumption. But over the past two or three decades consciousness and qualia have become respectable philosophical topics, and dualism has become an almost respectable philosophical position. The first of these developments is attributable to an evolving philosophy of cognitive science no longer afraid to deal with metaphysical problems in its field. The second is mainly due to logico-linguistic objections to the identity theory advanced by Saul Kripke in the early 1970s and theoretical investigations spurred by these objections. In 1967 the identity theory was at its zenith. Herbert Feigl's compendious "The 'Mental' and the 'Physical'" (1958) had been required reading for nine years and had just been reissued with a postscript (1967). J. J. C. Smart's "Sensations and Brain Processes" (1959) was only one year younger and, short and simple as it was, had been read by almost everyone.[2] Although their views differed in some respects, both these philosophers took the identification of mental and physical phenomena to be a contingent thesis that is most likely true. And then in three lectures first published in 1972 as *Naming and Necessity,* Saul Kripke (1980) shook the establishment with a linguistic analysis of the identity theory that allegedly proved it to be necessarily false. Currently, we find David Chalmers (1996) and others proudly proclaiming that they are dualists and that their position is entirely compatible with contemporary science.

Should we be pluralist about this issue and recognize both physicalism and dualism as legitimate positions? Psychophysical pluralism of the mild variety is the view that physicalism and dualism are both valid, or useful, when construed as methodological assumptions, that is, when taken to be assumptions of distinct but valuable approaches to research in cognitive science. The defense of such a view is not difficult. Considerable success in explaining various sensory and motor capacities has been achieved by the physicalist approach, on which sensations are identified with receptor excitations and perceptions and volitions are identified with neural excitation in the sensory and motor cortices. But the approach has proved less successful in the investigation of higher cognitive processes, such as inferring, calculating, decision making, and other forms of thinking, which are vaguely associated with processes in the association cortex and other cortical areas. So it could be argued that physicalism is clearly a useful methodological assumption in the investigation of sensory and motor processes, but less clearly useful in investigations of higher cognitive processes. The conclusion in that case would be not that the statements that define physicalism and dualism are both true, but rather that the two approaches corresponding to those statements are both valid, or valuable, or useful in certain subfields of application. Alternatively, it could plausibly be maintained that each approach is valuable in its own way throughout the entire field of cognitive science. A pluralism of incompatible *approaches,* which are not properly called true or false, is, as in many other fields, easy to defend.

Psychophysical pluralism of the strong variety is the view that the statement defining physicalism and the statement defining dualism are both true. Pluralism in this sense appears either difficult or impossible to defend, depending on how physicalism and dualism are defined. On the usual definition, which is adopted here, physicalism includes the statement that mental processes are bodily processes and dualism states that mental processes are not physical processes, bodily or otherwise. These two statements are logically incompatible, and a pluralism of logically incompatible *theories* is surely impossible to defend.

My impression is that many—perhaps most—philosophers of psychology today are deeply conflicted about the relation between mind and body: inclined both toward physicalism and toward dualism, and not merely as approaches, but as theories expressed in conflicting, logically incompatible statements. It is a deep and long-standing conflict in the history of philosophy, and its long history and apparent unresolvability may seem to indicate that the issue is a pseudo problem—the conflicting positions meaningless, or "metaphysical," or confused—and that the philosopher's task is to reveal it as such.[3] Although this direction is inviting, in the present chapter I have taken a somewhat different tack to discover where it leads. When I embarked on the trip I thought it would lead to the pluralist, or perhaps one should say relativist, conclusion that dualism and physicalism are each true in their appropriate spheres of application. Now I am less certain, and I invite the reader to help determine my coordinates.

Here is the route so far. It begins with reasons for thinking that psychophysical identities are true: examining the procedures by which they are confirmed, and refuting objections to these procedures. Next I consider the strongest reasons for judging psychophysical identities false—one based on the law of the indiscernibility of identicals, the other based on the law of the necessity of identity—and I refute them. The refutations suggest a distinction between two senses or types of identity: logical identity and what I call empirical identity,[4] with the indiscernibility of identicals and the necessity of identity governing only the former and physicalism involving only the latter, which is not subject to the two laws. Whether this qualifies as pluralism, or relativism, or conceptual clarification, or something else is left unsettled.

Confirmation of Psychophysical Identities

That water is H_2O and that temperature is molecular kinetic energy are physical identities. That the pain of a pinprick in my finger is neural excitation in my somatosensory cortex and that my current visual percept of this page is neural excitation in my visual cortex are psychophysical identities. Like physical identities, psychophysical identities are usually taken to be empirical statements that can be justified in a process of confirmation by evidence obtained through observation. Confirmation of a statement is the acquisition of evidence that the statement is true, and disconfirmation is the acquisition of evidence that the statement is false. It is assumed here that in practice confirmation and disconfirmation are rarely, if ever, conclusive. Evidence can be used to support a statement in two different ways. We will call the first experimental confirmation, and the second theoretical confirmation, although this familiar terminology can mislead.[5] We will examine methods of both types and objections to the assumption that they can confirm psychophysical identities.

Experimental Confirmation

I can confirm that the greenish gray glob hanging from the limb of a distant tree is a swarm of bees by approaching it until individual bees appear. I can confirm that the white flecks I see on the countertop are particles of sugar by touching them and tasting them. If necessary I can employ a magnifying glass to look more closely at the flecks. I cannot confirm that my pain is neural excitation in my somatosensory cortex or that my visual percept is excitation in my visual cortex by comparable methods. Introspecting my pain or my percept more closely or attentively does not cause either to appear as neural excitation,[6] and there is no method for observing neural excitation that can be correlated with introspection in the way that vision and touch can be correlated, so as to indicate that the two observations have the same object. Leibniz illustrated this difference by comparing the body with a mill, entering which could reveal levers and gears to the visitor but not mental processes. For a contemporary illustration, imagine that, as in the film *Fantastic Voyage,* a miniaturized explorer is inserted into my brain and sees flashes of light produced by neural excitation in area V4 of my visual cortex when I see a red cherry or a red afterimage. Neither I nor she (it was Raquel Welch) would thus have observed neural excitation to be or even have the location of my visual percept. And neither of us could have observed this in the way I visually observe the white flecks to be the sugar particles I feel, since we are distinct observers.

From such considerations many philosophers have fallaciously concluded that mental processes are not identical with brain processes or any other processes characterized solely in the terms of physiological or physical science.[7] The strongest legitimate conclusion is that the identity of mental processes and brain processes cannot be confirmed in the way a certain narrow class of physical identities can be confirmed. These are physical identities in which the phenomena can be observed under each of two descriptions by some mode of ordinary perception, and a single observer can thus perceive the phenomenon perceived under one description to have the location of the phenomenon perceived under the other description. I see the object as a greenish gray glob and see it maintain its location as I come to see it as a swarm of bees. I see that the white flecks I see have the locations of the sugar particles I feel by seeing that my finger touches the white flecks. I see that the magnifying glass that reveals the particles I see is focused on the white flecks I see. There is no technical term for this specific type of confirmation, but since it appears to be the most direct confirmation available for identities, we will for convenience call it "direct confirmation."[8] Psychophysical identities cannot be directly confirmed in our sense because introspection, which is the counterpart of ordinary outer perception, is the method employed to observe the phenomenon under its psychological description. I can introspect my visual percept to be a percept, but I cannot introspect my cortical excitation to be cortical excitation. Nor can I introspect my visual percept to have the location of my visual cortical excitation.[9]

That psychophysical identities are not directly confirmable does not entail that they are false or unconfirmable. For if it did, it would also entail that such well-accepted physical identities as "Water is H_2O" and "Lightning is an electric discharge in air" are false or unconfirmable, since they too cannot be directly confirmed. Water can be observed as water by ordinary perception (visual, tactual, possibly olfactory), but molecules of H_2O cannot be observed as molecules of H_2O by ordinary perception. (Indeed, it is unclear whether molecules of H_2O can be "observed" at all in any sense of the term that distinguishes observation from inference.) And water cannot be observed by ordinary perception to have the location of molecules of H_2O. The identity of water and H_2O is inferred by the chemical molecular theory of matter from the outcome of such processes as electrolysis, in which water is electrically converted into gases of hydrogen and oxygen and these are determined to be present in the proportions required by molecules of H_2O. Similarly, lightning can be observed to be lightning by visual perception, but an electric discharge in air cannot be observed to be an electric discharge by ordinary perception. Its presence is inferred by the theory of electricity from the readings of appropriately placed potentiometers, instruments that the theory predicts will respond in specific ways to electric discharges in a gas (sequential ionizations of its molecules). The use of special processes and instruments to make observations (findings, detections, etc.) of theoretical phenomena such as molecules and electrical currents requires a joint theory of processes/instruments and theoretical phenomena that permits the observer to infer from perceivable features of the process or instrument the presence of the phenomenon. For this reason, such observations are theory dependent to a greater degree, and in a different way, than unassisted perceptual observations.

"My visual percept is neural excitation in my visual cortex" is usefully compared with "Lightning is an electric discharge in air," an example bequeathed to us by Smart (1959). Neither identity can be directly confirmed in our sense, but this feature has no bearing on their truth or their indirect confirmability. The physical identity can be indirectly confirmed by a single observer from her visual observation of lightning and her inference to the electric discharge from readings of appropriately located potentiometers. Of course, a theory of electricity is required to make the inference, and in this respect the observation of the electric discharge is theory dependent. But this feature is no reason to conclude that the physical identity is not thus confirmed. The theory of electricity has independent confirmation, and the observations provide additional evidence for the conjunction of the theory of electricity and the identity. Analogously, the psychophysical identity can be confirmed by introspective observation of the visual percept and an inference from fMRI (functional magnetic resonance imaging) displays to the presence of visually stimulated neural excitation in my visual cortex. A neurophysiological theory of visual perception is required to make the inference from fMRI images to the presence of visually stimulated neural excitation in cortical area V4; in this respect the observation of such excitation is theory dependent. But again, this feature is no reason to doubt that the psychophysical identity between neural excitation and the percept of a cherry is confirmed by fMRI-mediated observation of the one and introspection of the other.[10]

It may seem that a single observer cannot confirm psychophysical identities. For if another person is the subject, I cannot introspect her pains and percepts, and if I am the subject, I cannot simultaneously introspect the psychological phenomenon and observe the correlated neural phenomenon. Although correct, this point does not constitute a reason to think that psychophysical identities are false or unconfirmable. Most accepted physical identities have been confirmed using distinct observers for the phenomena under their different descriptions. Furthermore, a single observer can be employed. I can be placed in an fMRI scanner that displays neural excitation as a pattern of colors on a computer screen, so positioned that I can see the screen of the device as it almost instantaneously displays patterns produced by neural excitation in the part of my visual cortex that subserves conscious perception. I cannot thus simultaneously introspect my visual percept of a red, round cherry and visually observe the computer image of the correlated neural excitation. Since the percept is simultaneous with the cortical excitation, it must, like the excitation, be produced prior to the computer image of the excitation; and my observation of the image even a second later requires eliminating my percept of the cherry. However, such an observation would approximate a simultaneous one, becoming in effect an instant replay of a simultaneous computer image, as compelling as an instant television replay showing that it was indeed a foul that the referee called a few seconds earlier. Neural excitations are, of course, not colored and probably not shaped in the usual macroscopic senses of the terms, and the colors and shapes on the fMRI screen representing neural excitations are artifacts of the particular electronic technique of representation employed. But suppose these colors and shapes are so selected that when I see a red, round cherry the computer displays a red, round image on the area of the screen activated by the correlated neural excitation. Such confirmation (or disconfirmation) of the identity between my percept and neural excitation in area V4 of my visual cortex would be compelling, at least as compelling as single-observer confirmation of the identity between lightning and an electric discharge in air.[11]

In spite of the similarity between the two cases, a crucial difference remains and threatens to invalidate the suggested method of confirming psychophysical identities. I can see that lightning occurs between the cloud and the steeple, each of which is adjacent to a potentiometer that monitors the electric discharge. But I cannot introspect that my percept of a cherry is in area V4 of my visual cortex, the area being scanned by the fMRI device. To the extent that my percept of something round and red has an introspectively determinable location, it is not a location in my nervous system. The cherry I see is seen to be located in the fruit bowl on the table, but the percept itself is not seen to have a location. The red, round afterimage I see with eyes closed is seen to have a location in some area of the visual field, usually the center. But the complete percept is the entire visual field—afterimage plus its dark surround—and the visual field is not introspected to have a location. In every normal case the conscious percept occurs simultaneously with excitation of a specific type in some area of my visual cortex and some milliseconds later than the related excitation in my retina. (Abnormal cases include those in which retina and cortex are disconnected and the afterimage is produced in the retina alone and consequently is unconscious, and those where percepts are produced centrally without retinal excitation.) But in no case can I see my percept to be located either in my cortex or on my retina. How then can I conclude that my percept is neural excitation in my visual cortex rather than neural excitation in the interneurons of my retina? How, more importantly, can I conclude that my percept is identical with neural excitation of any kind, or with any physical process or phenomenon that has a physical location?

The first difficulty is resolvable. Suppose that when I have a red, round afterimage in the center of my visual field there is excitation of a specific sort in the area of my visual cortex to which optic nerves from the central part of my retina project, and that when there is no such excitation I have no such afterimage, even when the retina has been stimulated in the manner required to produce red, round afterimages. These results would constitute significant evidence that my (conscious) percept of a red, round afterimage is cortical and not retinal excitation. Of course, the evidence is not conclusive, since empirical evidence never is. The second difficulty—whether I can conclude that the percept is neural excitation or a physical process of any kind—is deeper than that of the general inconclusiveness of empirical evidence. Dualists will object that the evidence is significant evidence only that the afterimage is *merely correlated* with cortical excitation. Physicalists, of course, can reply that the results are equally significant evidence for identity. This apparently unresolvable dispute is a symptom of a specific type of theoretical underdetermination. The experimental results provide evidence for either the identity of the afterimage with cortical excitation or for their mere correlation, but not for the one theory rather than the other. As we will see, the same problem arises in theoretical confirmation of psychophysical identities, and its solution rests on the analysis of identity and correlation provided in the final section of this chapter.

Theoretical Confirmation

This section examines what we have called the theoretical confirmation of psychophysical identities: confirmation by conjoining the identity hypothesis with others and inferring from the conjunction an experimentally confirmable consequence describing some empirical phenomenon. Confirmation of the consequence confirms the joint hypothesis that includes the identity statement. ("Confirm" is used here to mean "provide confirming evidence," with the assumption that confirming evidence is usually if not always inconclusive.) Thus confirmed, the joint hypothesis is said to predict and hence explain the phenomenon described in its consequence. This method is often called the hypothetico-inferential method of confirmation. Two examples will be provided.

Example A. The pain caused by pricking one's left index finger with a pin is identified with neural excitation in the right somatosensory cortex, excitation normally produced by nerve impulses from sensors in the skin of the finger that respond to injury such as that produced by the pinprick. We will call such cortical excitation P-excitation. This identification predicts that and explains why anesthetizing the nociceptive sensory nerves leading from the injury to the cortex will prevent pain when the finger is pricked, and also predicts that and explains why direct electrical stimulation of the cortical area in question produces the pain of a pricked finger in the absence of any pinprick. Note that the theory of pain employed here is at best a simplification of current theories, most of which analyze pain as having a sensory component and an affective (emotive) component and do not locate the affective component in the somatosensory cortex. Furthermore, the usual theory takes pain to be bodily injury (mainly tissue destruction), which may or may not be (consciously or unconsciously) perceived, and identifies the conscious *percept* of pain, not the pain, with neural excitation in the cortex.

Example B. Conscious visual images are identified with patterns of neural excitation in the visual cortex, let us suppose in area V4. This identification makes it possible to explain the phenomenon of double vision, in which focusing on a near object with a slightly more distant object visible in the background produces a single image of the near object and two side-by-side identical images of the distant object. The explanation is that the two retinal images of the near object are "fused" into a single conscious brain image in area V4 of the visual cortex because they have corresponding locations in the retina, whereas the two retinal images of the far object have noncorresponding locations and consequently are projected onto separate areas in V4 as distinct conscious images.[12]

Dualists will object that the predictions/explanations in these examples do not require positing psychophysical identities, only psychophysical correlations. Suppose in example A that somatosensory excitation of the type produced by the pinprick and the pain of such a pinprick invariably occur simultaneously if at all, but that the two are distinct. That assumption can equally well be used to predict that anesthetizing the appropriate sensory nerves leading from the injury will prevent pain when the finger is pricked and that appropriate electrical stimulation of the cortical area in question will produce the pain of a pricked finger in the absence of any stimulus. The objection is perhaps less obvious when applied to example B, but no less applicable. Suppose the explanatory theory states that conscious visual images are distinct from any of the patterns of neural excitation that invariably accompany them in area V4 of either cortex. On this assumption, retinal patterns in corresponding retinal locations will fuse into a single neural pattern in V4 that is accompanied by a single visual image, and retinal patterns in noncorresponding retinal locations will produce doubled neural patterns in V4 that are accompanied by doubled visual images. Thus single and double images respectively of the fixated and nonfixated objects are predicted and explained by positing psychophysical correlations rather than identities. Physicalists can of course reply that experimental confirmation of the theoretical predictions provides equally good evidence for the theory that the pain is identical with somatosensory cortical excitation and for the theory that conscious visual percepts are identical with excitations in the visual cortex. This dispute, like its counterpart regarding experimental confirmation, is an extreme case of the familiar problem of competing theories underdetermined by the available evidence, a case where it seems impossible even to imagine evidence that would decide in favor of the one theory over the other.

It is not often noticed that the same dispute can arise concerning what are universally agreed to be true physical identities, statements such as "Lightning is an electric discharge in air" and "Water is H_2O." A long tradition in philosophy regards colors, hot and cold, sounds, tastes, and so on as perceptual, mental entities—impressions, ideas, qualities of sensation, and so on—and consequently regards such macroscopic phenomena as water and lightning, which are defined or at least identified in terms of such perceptual qualities, to be either mental phenomena or physical phenomena of a different type than electric discharges and H_2O molecules. On this view, lightning should be held, by those of dualist inclination, to be merely correlated with electric discharges, and water merely correlated with collections of H_2O molecules. The epistemological situation here is analogous to that involving psychophysical identities: observations confirm either the identity or the correlation but not the one rather than the other. Physicalists can use the analogy to argue that the dualist view of the psychophysical relations is just as unscientific as the now outmoded dualist view of physical relations. And they can buttress the argument with an appeal to the scientific principle of parsimony, or theoretical simplicity, which is that the number of theoretical entities in a theory should be no greater than what is required to derive observational consequences from the theory.[13] Application of this principle selects identities rather than their correlational counterparts as confirmed. Dualists can reply that the principle of parsimony is merely a methodological recommendation for simplifying theories, without any ontological force. And they will surely reply that there are objections to psychophysical identifications that do not apply to the physical. To the most important of these we now turn.

Logical Objections to Psychophysical Identities

Previous sections examined methods for confirming physical and psychophysical identities and objections to their efficacy in confirming identities of the latter type. All such epistemological objections were satisfactorily refuted except one: that such methods no more strongly confirm identities than they do mere correlations between psychological and physical entities. This objection will be addressed in the final section of the chapter. We turn now to two logical objections to psychophysical identities: the first from Leibniz's law of the indiscernibility of identicals, the second from the necessity of identity, which we will honorifically call Kripke's law. If either of these is successful, then the methods examined confirm psychophysical correlations at best.

Objection from Leibniz's Law

In its contrapositive form, Leibniz's law of the indiscernibility of identicals states that if X and Y do not have the same properties then they are not identical. The objection from this law to mind-body identity maintains that mental processes possess some properties not possessed by neural processes and hence cannot be identified with such processes. For example, although the roundness of your red, round afterimage seen with eyes closed can perhaps be identified with some property of neural excitation, say the roundness of the area of correlated neural excitation, its redness cannot; for no brain process in its normal condition is red. Again, for example, although the neural excitation correlated with pain has properties that correlate with its duration and intensity, the pain has other qualities that neural excitations seem not to have. Pains are burning, aching, dull, and so on, but neural excitations are not. Even if such phenomenal qualities of pains are correlated with neural properties, they are distinct from the correlates.

In the case of the red afterimage, the usual reply is that the sensation, which consists in having a red afterimage, is not itself red but is rather a sensation *of* red. What object then is red? The intentionalist answer is that it is an "intentional object," the object of a mental act of the same type as an unfulfilled expectation or a false thought. This answer is objectionable to physicalists because intentional objects, if they exist at all, are surely nonphysical. Most physicalists subscribe to one of the answers suggested by U. T. Place and J. J. C. Smart. According to Place (1956, 49), a sensation is a sensation of red if it has the quality sensations possess when produced by red physical objects such as ripe cherries. This quality is not *red:* it is the physical object that is red. Thus described "topic neutrally" (to use Smart's term), a sensation of red can be identified with neural excitation of the type produced by seeing a red physical object: its type defined in terms of intensity, frequency, and/or other properties of neural excitation. The difficulty with this suggestion is our overwhelming impression that something is red when we have a red afterimage or a hallucination of a ripe cherry. It is not a ripe cherry or other red physical object, none being present, and it cannot be a neural process since no such process is red. To what is this phenomenal quality *red* to be assigned, if not to the cherry and not to the sensation? It is, to use Feigl's term, "homeless."

Smart's suggestion (1959, 148–50) is easily confused with that of Place. Like Place, he maintains that a sensation is a sensation of red if it has the neural quality sensations possess when produced by physical objects such as ripe cherries. But he does not say "when produced by *red* physical objects." According to him, colors are "powers to evoke certain sorts of discriminatory responses" (149) in perceivers, such as the response of discriminating ripe cherries from lettuce leaves. He admits that colors are powers that produce sensations, but "sensations . . . identifiable with brain processes" (ibid.), which are in a sense colorless. "Sensations are colorless for the same reason that *something* is colorless" (150), for example, the "something" referred to in my statement that there is something going on in me like what is going on when I see a ripe cherry. Apparently Smart must be interpreted as an extreme eliminative materialist, eliminating what we commonsensically regard as color not only from the sensation but also from the external world, and taking color terms to refer either to unknown properties of neural excitation or to the microphysical properties of objects that produce such excitation. On this view, phenomenal properties such as color are not simply homeless; they are fictitious.

Physicalists who subscribe to Leibniz's law tend to categorize pain as the perception of bodily injury and then apply a Place or Smart analysis of perception to yield the following result. Strictly speaking, it is bodily injury and not the sensation it produces that is painful; and the quality the subject feels—the so-called phenomenal quality—is the neural quality bodily sensations have when produced by painful bodily injury, whether the sensation is actually produced by such injury or, as in cases of phantom pain, is not. On this analysis the phenomenal qualities of pain become, like the color of an afterimage, homeless or fictitious.

A relatively unfamiliar reply to the objection from Leibniz's law is that phenomenal qualities such as redness and painfulness are physiological properties, but of a type distinct from those that are the subject matter of physics and current neuroscience. A version of this view, which can be extracted from the writings of Grover Maxwell (1976, 318–25; 1978, 395–98) and possibly Bertrand Russell (1948, 229–31; 1959, 25–27), maintains that phenomenal qualities are *intrinsic* properties of the subject's neural processes, while the qualities of neural processes posited by current neurophysiology—intensity, frequency, and duration of electrochemical neural discharges—are *structural* properties. On this view, when I see something red, whether an afterimage or a ripe cherry, I am directly aware of an intrinsic quality—redness—of some entity in my neural system; and when you see something red, you are directly aware of an intrinsic quality of some entity in your neural system. I cannot perceive the redness in your neural system. If I open your skull and look inside at your brain, I will see gray matter, and the grayness is an intrinsic quality in *my* neural system. By analogy I can infer that if you look at my brain you will see gray matter. But if I try to observe the color you see by using some instrument such as a PET or fMRI instrument to scan your brain, I will fail because such instruments reveal only structural properties of physical entities, not intrinsic properties such as redness and grayness. Although color patterns are usually the outputs of such devices, they are merely a convenient way of representing neurophysiological properties such as heat or blood flow, which in turn indicate neural activity. And according to the view at issue, these colors are properties of the brains of the person using the instrument.

At times Maxwell suggests that the properties treated by science are inevitably structural, and that in consequence the science of neurophysiology will never reveal or even satisfactorily conceptualize the intrinsic neurophysical qualities of redness and painfulness. At other times his suggestion is that an entirely novel neurophysiology, unlike anything we have at present, will be required. In either case, his fundamental claim is the same: current neurophysiology does not enable a scientist to directly observe phenomenal qualities in the neural systems of subjects or even to hypothesize their existence with its limited theoretical resources, but this fact is no objection to the view that phenomenal qualities are intrinsic physical properties of neural systems, since the subject and only the subject can, through introspection, directly observe such qualities.

Although Maxwell's view is intended to be physicalist, with both intrinsic and extrinsic properties classified as types of physical properties, it is difficult to distinguish from the view called property dualism—a dualism between mental and physical properties or manifestations of a single neutral type of substance. It is not parallelistic dualism, in which every mental phenomenon has a distinct neural correlate that can be discovered by the standard psychophysiology envisaged by Fechner and practiced today by psychophysiologists. Instead, it is a dualism in which both mental and physical events are at the same theoretical level, in a single system of causal relationships whose complete discovery must await the development of a futuristic neuroscience, perhaps forever. One problem in the view is that attempts to explicate the vague distinction between intrinsic and structural properties in a way that makes it correspond to that between phenomenal and other physical properties have not been successful. That an afterimage has a blue center with a red surround is a relational, and hence structural, property of the afterimage; but it is also a phenomenal property.

The reply to the objection from Leibniz's law recommended here is none of those considered above. It is, rather, that the indiscernibility of identicals is not a law that can be used to reject proposed identities such as the psychophysical and macrophysical-microphysical identities we have been considering. If it had been so employed, then many scientifically accepted identities of these types would never have been accepted. For example, according to current theory, a drop of water is identical with an enormous collection of molecules of H_2O separated by relatively vast amounts of empty space. But the pretheoretic understanding was that water is a spatially continuous substance, whereas a collection of entities with space between them is not, and so by Leibniz's law could not on that understanding be identical with a drop of water. We now have a strong inclination to insist that a drop of water does have space between its parts—between its molecules—and so is discontinuous, merely appearing to unaided perception to be continuous. But this contention assumes that the drop of water is identical with a collection of H_2O molecules, thus begging the question that the application of Leibniz's law is supposed to resolve. If the law had been applied to water prior to any proposal of an atomic theory of matter, the outcome would have been that water is not identical with a collection of H_2O molecules or any other particles, on the ground that water is spatially continuous and such a collection is not. Now that the atomic theory of matter and the theory that water is H_2O have been accepted, and we have attributed the properties of a collection of H_2O molecules to water, we find—unsurprisingly—that Leibniz's law is satisfied.

Proponents of the law may reply that Leibniz's law is not intended to apply to identities between an entity and its components. But then the law cannot be used to argue that an aching pain is not neural excitation on the ground that none of the excitations in the collection with which the pain is identified has the aching property.[14]

Proponents of Leibniz's law often reinterpret identifications, such as that between a piece of gold and a collection of atoms, by employing Locke's distinction between primary and secondary qualities as follows: In the strict, "primary quality" sense, yellow is not a physical property either of atoms of gold or of a collection of these, but rather a property of sensations produced in perceivers by light waves reflected from physical surfaces characterized by primary qualities; and the term "yellow" is applied to the physical surfaces in a derivative, "secondary quality" sense.

This interpretation ultimately sacrifices the very identity it aims to preserve. Obviously it can be applied to the macro properties of taste, odor, and felt heat, and it is easily extended to hardness and weight (as opposed to impenetrability and mass). Less obviously, the interpretation can be extended to most properties traditionally regarded as primary, such as spatial continuity, shape, and so on. As we noted, collections of atoms are not spatially continuous. (And what could be meant by saying that an individual atom is spatially continuous is unclear.) A lump of gold has shape; but the collection of its atoms does not, at least not in the sense of shape as the boundary of a spatially continuous object. When the claim of secondary, nonphysical character for properties is in this manner nonselectively applied and extended, every macro property of gold, and finally macro gold itself, is found to be nonphysical and hence not in any sense identical with physical atoms and their properties.[15]

The appropriate conclusion is that Leibniz's law cannot be employed to accept or reject empirical identities. Although, as our examples show, the law cannot be employed to determine whether microphysical-macrophysical identifications are in principle acceptable, it can be and is used to determine whether a particular identification of the sort is acceptable. For example, the successful identification of a sample of water with a collection of H_2O molecules must explain why water is transparent and visibly continuous by specifying microproperties of collections of H_2O molecules that correspond to the two macroproperties. Leibniz's law thus serves a regulative principle for proposed macro-microphysical identifications. It serves the same function for proposed specific psychophysical identifications. But it cannot be employed to determine whether psychophysical identifications are in general and in principle acceptable.

The conclusion that Leibniz's law cannot be employed to accept or reject empirical identities removes a major problem for the physicalist. Just as it removes the objection from Leibniz's law to identifying samples of water with collections of H_2O molecules, it removes the objection from that law to identifying qualities of pains and percepts with properties of neural excitation. The result thus also removes a loop in the "world knot"[16] that includes the mind-body problem. However, a serious problem remains as to what should be identified with what.

It seems obvious that my red, round afterimage should be identified with excitation in that part of the visual cortex that subserves conscious perception, there being no external entity with which to identify it. The red color of the ripe cherry I see could be identified with the "red" band of wavelengths reflected from the surface of the object (the band centered on wavelengths of approximately 700 nanometers) and its roundness identified with the shape of the surface—were it not for an objection that derives from the classical argument from illusion. It is possible to have a percept of a ripe cherry that is intrinsically indistinguishable from a red, round afterimage, that is, indistinguishable on the basis of momentary perception.[17] Consequently, it seems that the redness and roundness of my percept of a ripe cherry should be identified with excitation in my visual cortex like that with which my red, round afterimage is identified, and that the two cases should be distinguished by the causes of the percepts. The cortical excitation that is my percept of a cherry is caused by reflected light in the "red" wavelength band. The cortical excitation that is my red, round afterimage is caused by reflected light in a band complementary to the "red," which when intense or prolonged produces an aftereffect in the retinal receptors like the effect of "red" light that is then transmitted to the visual cortex and there produces the neural excitation that constitutes a red percept.

These identifications will appeal to scientists of perception and scientifically oriented philosophers of perception, especially those who have thoroughly considered the alternatives. But even they may share the uneasiness of many philosophers with the identification of perceived redness and roundness of external objects with qualities of sensations, which seems to violate common sense. Surely—as I say in correct English—it is the ripe cherry in my hand that is red and round, not my sensation of the cherry, whether the sensation is identified with neural excitation or not. The oddity of saying otherwise may be muted by the Lockean suggestion that the redness of the cherry is the capacity of its surface structure to reflect light waves that produce red sensations in normal human perceivers. But the oddity is not thus removed, as the need to refer to "normal" and "human" perceivers shows. The same surface structure might produce green sensations in an abnormal human, or in a bee. Furthermore, the roundness of the cherry also must be identified with its capacity to produce round sensations in normal human perceivers, and so on for all its other perceivable properties, thus divesting the cherry of most of the properties that characterize the macroscopic cherry of common human perception and common sense.

We find, then, that although it may remove a loop from the world knot, repealing Leibniz's law does not completely untie the knot. It is to be hoped that the repeal constitutes progress and not—as often happens with stubborn knots—the creation of new loops by removal of an old one.

Objection from Kripke's Law

What is here called Kripke's law is a version of the law of the necessity of identity, the law that genuine identity statements are necessarily true if true and necessarily false if false. In its traditional, pre-Kripkean form, the objection from this law to the mind-body identity thesis takes a simple form. Statements that identify mental and bodily processes are not necessarily true; hence, either they are false (necessarily) or they are not genuine identity statements. For example, the statement that my earache is P-excitation in my somatosensory cortex is not necessarily true. Such excitation could have occurred without my earache or without any pain in me at all. And my pain could have occurred together with P-excitation in my hippocampus and none in my somatosensory cortex or with Q-excitation in one of these organs, or without excitation in any part of my neural system. Hence, if the statement expresses an identity, then it is false. If it expresses a mere correlation between earache and P-excitation, it may be true, since statements of mere correlation are contingently true, true as a matter of contingent fact.

This objection is usually accompanied by a conceivability premise to buttress the contention that mind-body identities are not necessary. Using the previous example, this would consist in pointing out that we can conceive that P-excitation occurs without my earache occurring and that my earache occurs without P-excitation or any neural excitation occurring. Therefore (the argument goes), the statement that my earache is identical with P-excitation in my hippocampus is not necessarily true, and hence is either false necessarily or is not a genuine identity statement. Thus buttressed, the objection is a neurophysiologically embellished instance of Descartes's argument in the second of his *Meditations* that mind and body are distinct. He could conceive that the mind exists without the body and so concluded that mind and body are distinct.

Twentieth-century identity theorists such as Feigl and Smart were of course familiar with the objection, but they rejected it on the ground that in addition to necessary identities there are contingent identities that rest on empirical evidence and that true mind-body identities are in the latter class. Feigl likened statements such as "My earache is P-excitation" to "The morning star is the evening star," and Smart likened "My orange afterimage is a brain process" to "Lightning is an electric discharge in air," both of which appear to rest on empirical evidence and to be contingent. Then, in a series of lectures in 1970, first published in 1972,[18] Kripke reformulated the objection from the necessity of identity in a way that seemed to invalidate the Feigl/Smart defense of the identity thesis. Here in three steps is my construction of his reformulation.

Step 1. Identity theorists have failed to realize that necessary statements can and do sometimes rest on empirical evidence, and that statements such as "Water is H_2O" and "Pain is neural excitation" are among these. If I have learned that the number $10^{100}+1$ is a prime number by programming my computer to calculate all number products less than or equal to $10^{100}+1$, and it determines that the only such product is $(10^{100}+1)\cdot 1$, then I have come to know that the number is prime through the empirical evidence of my computer's calculation. And yet the statement "$10^{100}+1$ is a prime number" has been determined to be necessarily true. Statements such as "Water is H_2O" and "Pain is neural excitation" are also statements that rest on empirical evidence, evidence that may include the output of an instrument or calculating device.[19]

Step 2. Identity theorists have also failed to understand that contingent identity statements must be distinguished from necessary identity statements and that mind-body identity statements such as "Pain is neural excitation" are among the latter. In a necessary identity statement, both terms flanking the identity sign are rigid designators. A rigid designator is a term whose referent has been fixed in a causal and/or referential act of dubbing so that it refers to that object at every possible world in which the object exists. There are two main types of rigid designators: names of individual persons and things and names of kinds.

"Cicero" and "Tully" are examples of the former type. These are names by which the man was known in his capacities as, respectively, powerful orator and Roman statesman. Since the names are rigid, the statement "Cicero is Tully" is true at every possible world, necessarily true. Another familiar example: "Hesperus" and "Phosphorus" are names that were given respectively to the evening star and morning star (so-called) before anyone realized that they designate a single heavenly body, the planet Venus. (As does Kripke, we ignore the complication that Venus is not a star.) Each name, being rigid, designates that body at every possible world. Consequently, "Hesperus is (identical with) Phosphorus" is true at every possible world, necessarily true. "The morning star" and "the evening star" are examples of nonrigid designators. At one possible world "the morning star" and "the evening star" designate distinct objects, the first following the actual-world orbit of Venus in the morning, the second following it in the evening. At yet another possible world, both objects follow the Venus orbit, but one pops into existence as the other pops out at the appropriate times in the morning and the evening. At both worlds, "The morning star is the evening star" is false, and so is "Hesperus is the morning star." These identity statements are therefore contingent and true at the actual world.

Examples of the second type of rigid designator are "water" and "H_2O," the former a name of a natural kind, the latter a name of a theoretical kind, as they are often called. "Water," through its evolution from synonyms in prior languages and its cultural transmission in English, is the name given to the transparent, tasteless substance that in liquid form falls in rain and flows and waves in streams, lakes, and oceans of Earth. "H_2O" is the name scientists gave to the kind of molecule that consists of one oxygen atom and two hydrogen atoms, as they made the theoretical discovery that the substance consisting of such molecules is identical with water. Because "water" and "H_2O" are rigid designators, "Water is (identical with) H_2O" is either necessarily true or necessarily false. Consequently, when scientists discovered that it is true, they discovered that it is necessarily true. The discovery was made through the acquisition of empirical evidence, for example, evidence that water is transformed into hydrogen gas and oxygen gas through electrolysis. But as was seen in step 1, necessarily true statements can, and some do, rest on empirical evidence.

Similar examples of necessarily true macro-microphysical identity statements are "Heat is molecular motion" and "Gold is the element with atomic number 79." An example that provides a better analogy for psychophysical, mind-body identity statements is "Lightning is electricity" or "Lightning is an electric discharge." Here too the flanking terms "lightning" and "electricity" and "electric discharge" are rigid designators, and so the statements are either necessarily true or necessarily false. And here too they have been empirically discovered by scientists to be necessarily true.

Step 3. Had they understood the logic and epistemology of identity outlined in steps 1 and 2, the identity theorists would have realized that statements of the mind-body identity thesis such as "Pain is C-fiber neural excitation" also contain flanking terms that rigidly designate names of kinds and consequently are necessarily true or necessarily false. They would then surely have seen the force of Descartes's objection to physicalism, which is that such statements are possibly false and consequently necessarily false. For it seems possible (is conceivable)[20] that I am in pain when there is no C-fiber neural excitation in my body but only A-fiber neural excitation. It even seems possible that I am in pain when there is no neural excitation of any kind in my body or anywhere else in the universe.[21] "Pain is neural excitation" is therefore not necessarily true and consequently is necessarily false, false in every possible world, including our actual world. Extensions of this argument establish that pain is not identical with any type of bodily or physical process, and parallel arguments show that percepts, images, thoughts, and all other mental entities are not identical with neural excitation or with any bodily or physical process.

Kripke's theory of the logic and epistemology of identity is flawed in my opinion, but its apparently seamless character makes it difficult to determine where the flaws lie, whether in the epistemology or the logic or the linguistics. To focus the task I will point out two inconsistencies (or at least asymmetries) in Kripke's application of his theory to the identities under consideration, and then try to identify their source. Both inconsistencies are exhibited in Kripke's finding that "Pain is neural excitation" is false (necessarily) but that "Water is H_2O" is true (necessarily).

The first inconsistency lies in Kripke's recognition that "Pain is neural excitation" is possibly (conceivably) false and his failure to recognize that "Water is H_2O" is also possibly (conceivably) false.[22] The first possibility is sufficiently apparent, and it entails that on Kripke's theory "Pain is neural excitation" is possibly false, hence necessarily false and thus not true. As for the second possibility, we may suppose that at some period in ancient Greek or Mesopotamian science the prevailing hypothesis was that water is composed of tiny, transparent, gelatinous globules and at some earlier period was that water is composed only of water—a cup of it composed of smaller volumes, and those volumes of still smaller volumes, and so on ad infinitum. Whatever the history, either hypothesis is today still possibly (conceivably) true. Such possibilities show that on Kripke's theory "Water is H_2O" is possibly false, hence necessarily false and thus not true. But Kripke maintains that science has established the statement to be true.

The second inconsistency lies in Kripke's willingness to accept empirical evidence for the truth of "Water is H_2O" and his unwillingness to accept similar evidence for the truth of "Pain is neural excitation." Among the evidence for the former identity is the confirmation of its prediction that electrolysis of water produces hydrogen gas and oxygen gas. Among the evidence for the latter is the confirmation of its prediction that anesthesia of specific neural systems eliminates pain. The evidence for the pain-neural excitation identity is perhaps not as strong as that for the water-H_2O identity, but it is surely strong enough to consider the identity probable. This second inconsistency is connected with the first. Kripke is disinclined to accept empirical evidence that pain is neural excitation because of his intuition that the statement is possibly false and hence necessarily false. But if empirical evidence can override the intuition that "Water is H_2O" is possibly false and hence necessarily false, then it can override the intuition that "Pain is neural excitation" is possibly false and hence necessarily false.

These inconsistencies indicate that something is wrong with Kripke's theory, in particular with his test for nonidentity or its application. The test is obviously unreliable when applied in the obvious way to established empirical identities of the first type, those whose flanking terms are both names of physical continuants, such as "Hesperus is Phosphorus." As we have seen, it is easy to imagine that this statement is false, for example, by supposing that both bodies follow the orbit of Venus and one pops into existence as the other pops out. The test thus indicates that Hesperus is not identical with Phosphorus, which we know to be false and hence the wrong result. It is also unreliable when applied to identities of the second type, whose flanking terms are both names of kinds of things or substances. Again, as we have seen, it is easy to suppose that water is transparent gelatinous globules, and ancient scientists may have so supposed. The test thus indicates that water is not H_2O, which again we know to be the wrong result.

In an attempt to discover the flaw, let us examine the underlying logic of the test. It employs the following general version of the central axiom of Kripke's theory of identity:

> (*) If (1) "A is identical with B" is true (at our actual world) and (2) "A" and "B" are rigid designators (each designating the same entity at every possible world in which it exists), then (3) "A is identical with B" is true at every possible world.

Kripke's application of this axiom proceeds as follows: Statement (3) is correctly found to be false by conceiving (thinking possible) a world in which "A is identical with B" is false. It is then inferred—correctly by basic propositional logic—that (1) and (2) are not both true. Finally, it is assumed that (2) is true, from which and previous results it follows that (1)—the identity hypothesis—is false. One fallacy in this reasoning lies in its assumption that (2)—"A" and "B" are both rigid designators—is true.

In applying the test to the pain-neural excitation case, Kripke correctly notes that (3) is false, that "Pain is identical with neural excitation" is not true at every possible world, and using (*) he correctly infers that (1) and (2) are not both true. But then he assumes that statement (2)—"pain" and "neural excitation" are rigid designators—is true. The fallacy in so doing is not obvious to someone like Kripke, who believes that statement (1)—"Pain is neural excitation" is true—has not been scientifically established. So he concludes that (1) is the false conjunct. If his conclusion were sound, it would be sound in every instance of the test's application; but obviously it is not sound in instances where (1) has been independently established to be true.

In the Hesperus-Phosphorus case it is clear that statement (2)—"Hesperus" and "Phosphorus" are rigid designators—cannot be assumed true without begging the question of statement (1)—"Hesperus is Phosphorus" is true, a question long ago settled in the affirmative by astronomers. For since "Hesperus is Phosphorus" is false at some possible worlds, i.e., (3) is false, and (1) and (2) cannot both be true; and since (1) is known to be true, the false conjunct must be (2) instead of (1). Accordingly, "Hesperus" and "Phosphorus" cannot both be rigid designators. Since they obviously have the same linguistic status, neither is a rigid designator as it functions in Kripke's test. The term "Hesperus" functions in the test as the description "the bright star that appears in the western sky on clear evenings" and "Phosphorus" functions as the description "the bright star that appears in the eastern sky on clear mornings." If they did not so function, if they functioned merely as "this" and "that" with no attached descriptive content even in thought, the test would not be possible. For the conceptual materials required to imagine (describe) a possible world in which the named objects are clearly not identical and thus show that (3) is false would be unavailable. And if the descriptions were somehow "rigidified," then it would not be possible to employ them—as we have—to imagine both a possible world in which their referents are identical (the actual world, for example) and a possible world in which their referents are not identical (for example, the world in which their distinct referents occupy the Venus orbit at appropriately different times). If two such distinct worlds are thus designatable, the designators employed cannot be rigid in Kripke's sense.

To apply Kripke's test in the water-H_2O case, "water" is substituted for "A" and "H_2O" for "B" in axiom (*). Our critique takes the same form as in the Hesperus-Phosphorus case. Assumption (2)—that "water" and "H_2O"

are both rigid designators—begs the question of whether water is identical with H_2O, a question already settled in the affirmative by physical science. Since "Water is identical with H_2O" is false in some of the possible worlds imagined earlier, either "Water is identical with H_2O" is false or "water" and "H_2O" are not both rigid designators. Science has established that "Water is identical with H_2O" is true; consequently, "water" and "H_2O" are not both rigid designators. Perhaps "H_2O," being the name of a theoretically constructed entity, designates that entity in every possible world by stipulation. But "water" is not a term of this sort. When I imagine water composed of transparent gelatinous globules or composed of smaller and smaller bits of water, I conceptualize water as "the transparent, odorless liquid found in streams, lakes, and oceans of earth," thus in effect using the term "water" as an abbreviation of the quoted definite description that expresses my conception. This description is not rigid: it describes H_2O in the actual world and gelatinous globules in a possible nonactual world.[23]

In examples such as the two above, where a physical identity statement has been scientifically established to be true, the application of Kripke's test for nonidentity can lead only to the conclusion that the terms flanking the identity sign are not both rigid designators. If, as seems clear, physical identity statements that have *not* been established to be true are of the same type as one of our two examples, Kripke's test cannot be applied to such a statement without begging both the question of whether the statement is true and the question of whether its flanking terms are rigid designators. For the statement may, like those in our two cases, be true although conceivably false, in which case its flanking terms cannot both be rigid designators. There is disagreement as to whether psychophysical identity statements such as "Pain is neural excitation," whose truth is controversial, are of the same type as one of our examples (the second). If, as I maintain, they are of the same type, the application of Kripke's test to psychophysical identity statements likewise begs the questions of their truth and the rigidity of their flanking terms. The test is therefore useless when applied to empirical identity statements of every type, since it cannot show that they are empirically false.

This result is unsurprising, since Kripke designed the test for application to *logical* identity statements, incorrectly assuming that all identity statements are logical. Even for this application the test is generally useless if understood to involve ascertaining that the statement in question is false in some possible (conceivable) world. To determine that "59 × 37 = 2173" is false, we do not consider whether it is false in some possible world. We do the multiplication. The distinction between empirical and logical identity statements is described in the next section.

The Resolution: Empirical versus Logical Identity

Logically possible worlds are worlds whose descriptions are consistent with the laws and generalizations of logic and pure mathematics. Empirically possible worlds are worlds whose descriptions are in addition consistent with the laws and generalizations of empirical science and those of enlightened common sense. Empirically necessary statements are true (or false as the case may be) only in every empirically possible world. Logically necessary statements are true (or false as the case may be) in every logically possible world. Empirically rigid designators denote the same entity in every empirically possible world. Empirical identity statements with empirically rigid flanking designators, such as "Water is H_2O," are empirically necessary. Logically rigid designators denote the same entity in every logically possible world. Logical identity statements with logically rigid designators, such as "2 + 3 = 5," are logically necessary. The set of empirically possible worlds is a proper (partial) subset of the set of logically possible worlds. Consequently, an empirical identity statement—such as "Water is H_2O"—that is true in all empirically possible worlds and thus empirically necessarily true is false in some logically possible worlds and thus not logically necessarily true.

In empirical identity statements such as "Water is H_2O," at least one of the flanking terms is not logically rigid. Consequently Kripke's (logical) possibility test cannot be applied to such statements. My hypothesis is that logically rigid designators (names that denote the same entity in every logically possible world) and logically necessary identities (identities true [or false] in every logically possible world) are found not in our ordinary or scientific empirical languages, but only in (a) artificial and theoretical languages in which designators are *stipulated* to be rigid and (b) the languages of pure mathematics, set theory, and logic, where designators are rigid either by virtue of designating themselves, so to speak (the nominalist view), or by virtue of designating abstract entities (the Platonist view).

We can now turn to the promised explanation and resolution of the conflict philosophers of psychology undergo in their inclination to subscribe both to dualism and to physicalism. On the one hand, we are inclined toward physicalism by the mountain of direct and indirect evidence for correlations between psychological processes and neural excitation, and by the apparent lack of any better reason to regard these correlations as mere correlations instead of identities than we have so to regard correlations such as that between a sample of water and a collection of H_2O molecules. On the other hand, we are inclined toward dualism by the objection from Leibniz's law that percepts and pains cannot be identical with neural excitations because they have different properties, and by the objection from Kripke's law that such psychophysical identities are not necessarily true and hence are false necessarily.

The resolution of the conflict was prefigured in our fundamental replies to the two objections. The fundamental reply to the objection from Leibniz's law is that psychophysical (as well as physical) identity statements involve a sense or type of identity that is not governed by Leibniz's law. For, as was argued above, if the psychophysical and physical identity statements accepted in science were so governed, most of the latter would have to be evaluated as false. Since a drop of water has the property of continuity and a collection of molecules does not, the statement that the drop is a collection of molecules of H_2O would be false if it were governed by Leibniz's law. It is a true statement; therefore, it is not governed by Leibniz's law. If physical identities are not governed by Leibniz's law, then neither are psychophysical identities.

Precisely stated, the objection from Kripke's law is that identity statements whose identity sign is doubly flanked by rigid designators are necessary: necessarily true or necessarily false; and since psychophysical identities flanked by rigid designators are not necessarily true, they are false (necessarily). The fundamental reply is that psychophysical as well as physical identity statements involve a type of identity not governed by Kripke's law. Some philosophers have called it "contingent identity" to distinguish it from the necessary identity governed by Kripke's law, but this terminology, as we shall see, is potentially misleading. Just as there is a type of identity that is not governed by Leibniz's law of the indiscernibility of identicals, there is a type that is not governed by Kripke's law of the necessity of identity. These two types combine in a single type that is naturally called *empirical identity* to contrast it with the *logical identity* governed by Leibniz's law and Kripke's law.

The ultimate resolution of conflicting inclinations toward dualism and physicalism requires distinguishing between logical and empirical identity, realizing that they have different contexts of application, and restricting the concept of logical identity to logical (pure logical, pure mathematical, set theoretical, etc.) contexts and the concept of empirical identity to empirical (physical, biological, psychological, etc.) contexts. Pure mathematical and set theoretical statements, among others, employ logical identity and satisfy both Kripke's and Leibniz's laws. For example, the statement "2 + 3 = 5" is logically necessarily true, true at all logically possible worlds; and the identified numbers have all their properties in common. The sum of 2 and 3 is not divisible without remainder by 2 and neither is 5; the sum is the square root of 25 and so is 5; and so on. Physical and psychophysical identity statements, of which we have supplied many examples, employ empirical identity and do not satisfy either Kripke's law or Leibniz's law. But when we contemplate them, we wrongly assume that they too involve logical identity and apply the two laws. My red afterimage is neural excitation in my visual cortex? How can that be, since the one is red and the other is not, and since the neural excitation could occur without the afterimage occurring?

Those who, with Kripke, hold there is only one type of identity, governed by Leibniz's law and/or Kripke's law, will complain that my "empirical identity" is in reality mere correlation with an inappropriate, tendentious name. They will argue that to say that A is identical with B is to say that 'A' and 'B' denote a single thing, a single thing in every possible world if the meanings (denotations) of 'A' and 'B' do not change from world to world. Consequently, if A and B do not have the same properties, then they are not identical but are at best merely correlated; and if 'A' and 'B' denote distinct things in other possible worlds, then their meanings (denotations) are not rigid. This objection begs the question of whether there are two types of identity. It merely defines logical identity and derives Leibniz's law and Kripke's law from the definition.[24]

Of course, we can call the identity I dubbed "empirical identity," which is governed neither by Kripke's law nor Leibniz's law, by some other suitable name; but "correlation" and "constant correlation" are not suitable. One reason is that these names apply to cases in which the correlated entities have distinct locations and/or distinct times, and cases in which one of the entities is the cause of the other. Neither of these features is a feature of the relation between a sample of water and the collection of H_2O molecules with which it is identified, nor are they features of the relation between a visual percept and the cortical neural excitation with which neuropsychologists identify it. Water exists in the same place and at the same time as the collection of its molecules, and water is neither a cause nor an effect of the collection of molecules. Similarly, my visual percept exists in the same place (although not thus introspected) and at the same time as the cortical neural excitation with which it is identified, and is therefore neither a cause nor an effect of the cortical neural excitation.

An additional reason for distinguishing empirical identity from constant correlation is that the relation is necessary, albeit empirically necessary and not logically necessary. It is for this reason that calling it "contingent identity" is misleading. To say that water is identical with H_2O entails that water is correlated with H_2O in all relevant empirically possible worlds, in all worlds that satisfy the true laws of any relevant empirical science—physics, chemistry, and so on. To say that mammalian pain is neural excitation in the animal's somatosensory cortex entails that pain is correlated with neural excitation in the somatosensory cortex of every mammal in all empirically possible worlds, all worlds that satisfy the true laws of any relevant empirical science—physics, chemistry, biology, neurophysiology, psychology, and so on.

Physical necessity is a type of empirical necessity. Consider a very small body at the surface of a very large spherical body with a radius of 6.37×10^6 meters and a mass of 5.98×10^{24} kg (Earth, for example), both so distant from other bodies that only the gravitational attraction between the two need be taken into account. It is physically necessary that the first body accelerates toward the center of the second at very nearly the rate of 9.8 meters per second. That is to say, it is true in all physically possible worlds, as these are defined by the law of gravitation, which governs at this macrophysical level. But it is not logically necessary, not true in all logically possible worlds, whether "logically possible" is taken to mean "logically consistent" or to mean "conceivable." For there is a logically consistent, conceivable world in which a body on the surface of the one described above accelerates at some different rate or in the opposite direction or not at all, gravitational force in that case being different or absent. For example, if "A lump of gold is a collection of atoms with atomic weight 79" is necessary, it is physically necessary, not logically necessary. For there is a logically consistent, conceivable world in which lumps of gold—the yellow, malleable metal from which we make rings and money—have atoms of some atomic weight other than 79, and one in which they are not composed of such atoms or of atoms of any kind.

There is no better name for the relation at issue than "empirical identity," a name that classifies it as an identity but distinguishes it from the "logical identity" governed by Leibniz's and Kripke's laws. It is useful to compare empirical identity with equality in length of rods, both being (like logical identity) an equivalence relation.[25] A is equal in length to B if and only if the length of A is identical to the length of B. Here A and B may be logically identical or logically distinct. On my analysis, A is empirically identical with B if and only if the spatiotemporal location of A completely coincides with that of B. Here too A and B may be logically identical or logically distinct. Where A is my afterimage and B is the correlated excitation in my visual cortex, A and B satisfy neither Leibniz's nor Kripke's laws (as we have seen) and hence are logically distinct; nonetheless they are empirically identical. In the case where A and B are logically identical, empirical and logical identity coincide.

It may be objected that identity is by definition the relation that an entity can have only to itself, and that since my so-called empirical identity is not such a relation, it is not identity. But this begs the fundamental question at issue, the question of whether there is another type of identity distinct from logical identity. Furthermore, there is a sense of "itself" in which empirical identity *is* a relation of an object to itself (see note 14).

My analysis of empirical identity is analogous to the twentieth-century modal version of Hume's analysis of causation as a type of constant spatiotemporal conjunction. The modal version of his analysis adds the condition that the conjunction is empirically necessary, present in all empirically possible worlds.[26] The resemblance is not accidental. My analysis is offered in the spirit of logical empiricism, the empiricism of Hume, Mill, Carnap, Feigl, Smart, and Quine, which tries to remove logical necessity from the empirical world and locate it in the logico-linguistic systems with which we conceptualize that world. Where (as here) the term "correlation" is employed neutrally, as implying neither identity nor nonidentity, empirical identity can be succinctly characterized as spatiotemporally equated empirically necessary correlation.

There is an obvious objection to this analysis—an application of the general objection from Leibniz's law—to which at least a brief reply is required. The objection is that although bodily sensations (pains, for example) and some percepts (afterimages, for example) are located in a subjective sensory space, they do not have physical spatial locations, at least not the locations of the neural processes in the central nervous system with which they are identified by neurophysiologists; and such mental entities as desires, fears, memories, thoughts, etc., do not have even subjective location. In reply, first consider thoughts and similar mental processes. They do not have an introspectible location, subjective or physical, and there is no biological reason for them to have one. But the fact that they have no introspectible location is not evidence that they have no bodily location. Although it initially seems otherwise, a similar analysis applies to percepts and bodily sensations. They too have no introspectible location. Some are sensations of located items, real or hallucinated, and the introspection—a sort of duplicate but conscious sensation—is of the *perceived* location of the item, not the location of the sensation. For example, when I introspect the pain of a pinprick in my finger, I introspect the location of what apparently causes the pain, not the location of the pain itself, and often not the location of what really causes the pain. Again, the fact that sensations have no introspectible location is no evidence that they have no location.[27]

Coda

A tenacious dualist will refuse to concede, offering the following reason. "I grant—at least for the sake of argument—that you have neutralized all the empirical objections and all the logical objections to dualism. But you have not shown why we should accept physicalism rather than dualism. You hold that only empirical evidence is relevant, and you have granted that all present and probably all future evidence will not decide between the theory that mental phenomena are identical with physical phenomena and the theory that mental phenomena are merely correlated with physical phenomena."

My reply is that we should accept physicalism because dualism leaves the relationship between mental and physical phenomena unclear, undecidable, and unexplainable. Why is my afterimage correlated with excitation in my visual cortex? Does the neural excitation cause the pain (epiphenomenalism)? Does the pain cause the neural excitation (interactionism)? Do the pain and the correlated neural excitation have a common cause, and if so what is it (god perhaps)? These questions have been debated for hundreds of years and they seem unanswerable. They vanish on the identity theory and leave nothing in their wake. Why is my afterimage identical with excitation in my visual cortex? Why is lightning an electric discharge in air? Why is water H_2O? The only possible answer is, as the evidence shows, "It just is," which shows that the question has no answer and hence is illegitimate.

As indicated in the introduction, I have some uncertainty about the precise character of the resolution I have proposed. Here is a version that currently seems to qualify as pluralist. The distinction between logical identity and empirical identity resolves the perennial conflict between dualism and physicalism by revealing what is correct and what is incorrect in each position. Dualism is correct that mind and body are not identical in the logical sense, but it fails to realize that in the empirical sense mind and body are identical. Physicalism is correct that mind and body are identical in the empirical sense, but it fails to realize that in the logical sense mind and body are not identical. The distinction between logical identity and empirical identity also explains the ambivalence of philosophers of psychology about the relation between mind and body. When inclined toward physicalism we implicitly define it in terms of empirical identity, as the view that psychological and physical phenomena are empirically identical, which is true. When inclined toward dualism, we implicitly define physicalism in terms of logical identity, as the view that psychological phenomena and physical phenomena are logically identical, which is false.

This resolution is pluralist in at least the respect that it distinguishes two equally legitimate senses of identity and their appropriate fields of application: logical identity for pure logic and pure mathematics, empirical identity for the empirical sciences. Whether it is pluralist in any stronger respect is unclear to me. My main uncertainty is whether dualism is correct in its contention that in some sense mind and body are distinct. As we have seen, the empirical evidence suggests that they are not distinct in the empirical sense of identity. And if the concept of logical identity is—as I suspect—not only not intended to apply to empirical entities, but is thus *inapplicable,* then mind and body cannot correctly be said to be distinct or to be identical in the logical sense.[28]

Notes

This chapter is dedicated to the memory of Paul E. Meehl (1921–2003). He was a founding member of the Minnesota Center for Philosophy of Science and he participated in activities of the Center until a few months before his death. His work on my topic, together with that of Feigl, Maxwell, and Gunderson is in the Minnesota tradition of attempts to solve the mind-body problem with a noneliminative, scientific mind-body identity theory. He had the conference version of the paper read to him—his vision having become impaired—and in a private communication expressed his general agreement and made what he called, in characteristically understated fashion, some "minor points." These provoked several significant improvements in the section on empirical confirmation.

1. In contexts such as the one noted, the term "identity" is employed as an abbreviation for "identity statement" with no implication as to truth or falsity of the statement. The terms "correlated" and "correlation" will be employed neutrally, implying neither identity nor nonidentity and applying both to items that are *identical* and to items that are not identical but *merely correlated*.

The theory of mind-body identity—identity theory, briefly—is the theory that all things and properties are physical and some among them are mental or phenomenal as well. The term "physicalism" is ambiguous in the literature, sometimes used to include the identity theory, as in this essay, sometimes used to exclude it. In its latter use the term refers to "eliminative materialism" (often simply called "materialism"), the theory that everything is material and that mental and phenomenal terms, such as "painful," "red," "hot," etc., should be replaced by scientific terms of physics, chemistry, biology, and neuroscience. The identity theory does not replace or eliminate mental and phenomenal terms, but identifies their referents with those of scientific physical terms. The term "monism" is used in our context to denote the mind-body identity theory, although both eliminative materialism and idealism (the theory that everything is mental) are monistic in admitting only *one* type of thing.

2. In his first footnote Smart cites U. T. Place's essay of two years earlier, "Is Consciousness a Brain Process?" (1956), as the precursor to his own. He adds that his own essay is "meant also to supplement" Feigl's (1958).

3. Apparently Carnap regarded the mind-body problem as a pseudo problem whose genuine counterpart is resolved by a linguistic decision influenced by empirical evidence, and Feigl regarded the problem as genuine but one that could not be resolved without philosophical interpretation in addition to empirical evidence. Whether their positions ultimately differ, and if so how, are difficult questions.

4. Empirical identities hold on empirical (concrete) entities such as pains and neural excitations or samples of water and collections of H_2O molecules. Logical identities hold on logical (abstract) entities such as 5 and 2+3, or the complement of the intersection of two sets and the union of their complements, or a proposition and its double negation. My distinction between empirical and logical identity does not correspond to that between contingent and necessary identity (which Smart employed) if, as I maintain, identities

such as “Water is H_2O” are empirically necessary. It corresponds to the distinction between synthetic and analytic identity (which Feigl mainly employed) only if logicism—the view that logical truths are reducible to analytic truths—is true. The distinction apparently corresponds exactly to that between a posteriori and a priori identities.

5. The terminology obscures the fact that both methods are experimental in the sense that both require observation of a correlation between the phenomena that are inferred to be identical. It also obscures the fact that both methods are theoretical in the sense that some background theory is almost always tacitly employed in making the inference. In experimental confirmation this theory is usually acquired through normal learning and so is part of common sense, but it is perhaps sometimes innate. For example, the perhaps innate theory that identicals exhibit perceptual continuity is tacitly employed in inferring that the greenish gray glob is a swarm of bees. In theoretical confirmation the background theory is usually part of specialized science. Theoretical confirmation is often called “hypothetico-inferential” or “abductive” confirmation. No elegant pair of labels for the two methods that is suggestive without being misleading comes to mind.

6. Place makes this point in a memorable passage: “A closer introspective scrutiny will never reveal the passage of nerve impulses over a thousand synapses in the way that a closer scrutiny of a cloud will reveal a mass of tiny particles in suspension. The operations required to verify statements about consciousness and statements about brain processes are fundamentally different” (1956, 47). The same point is the focus of an essay by Gunderson (1970).

7. Gunderson (1970, 1974) examines differences in the ways we confirm statements about our own mental processes and statements about the mental processes of others, differences that can seem to be an argument for dualism but in fact reveal only a profound “asymmetry” between our access to mental processes in ourselves and our access to such processes in others, an epistemological rather than an ontological dualism. However, the main source of the asymmetry does not seem to be the difference between self and other but rather the difference between introspection and other methods of observing mental processes. This difference is illustrated in confirming one’s own perceptions by introspection and confirming one’s own (even unconscious) perceptions with an instrument such as an fMRI (functional magnetic resonance imaging) device.

8. The familiar terminology is again potentially misleading. Probably all confirmation of identities is indirect in the sense that the observations employed involve (possibly unconscious) interpretation, and the correlation of these observations involves (possibly unconscious) inference.

9. This difference between physical and psychophysical identities is related to, but not supportive of, a classical objection to the identification of mental phenomena and physical phenomena, namely, that physical phenomena have physical location and mental phenomena do not. My characterization does not imply that mental phenomena do not have physical location. It implies only that the location of a mental phenomenon cannot be determined to be identical or nonidentical with that of a physical phenomenon by observational or quasi-observational methods—such as introspection—that are employed to detect the presence of the phenomenon. See the final section of this chapter for a reply to the classical objection.

10. Physical identities such as “Lightning is an electric discharge in air” and psychophysical identities such as “Pain is neural excitation in the somatosensory cortex” are sometimes contrasted with identities such as “That greenish gray glob is a swarm of bees” by referring to the former as theoretical identities and the latter as observational identities. This nomenclature is misleading, since theories are employed in confirming identities of both types. When I see the object as a greenish gray glob and then as a swarm of

bees, I unconsciously employ the theory that when a temporal sequence of perceptions of increasingly "grainy" objects in the same position in my field of vision is interspersed between the two perceptions they are of the same object, a theory that is either innate or was acquired in infancy. Consequently, it seems that I can simply, immediately see that the object under the one description is the object under the other. In reality an intuitive theory of continuity is mediating an unconscious inference from their converging visual appearances that the one object is the other. For a relevant treatment of observation see Savage 1992.

11. Thus employed, the fMRI scanner would be a realization of an instrument vaguely imagined and called an "autocerebroscope" by Feigl (1958, 456). Meehl (1966) imagined the first detailed realization, in which simultaneous introspection of the percept and observation of its neural correlate is apparently achieved by programming the instrument to display an R (for "red") or a G (for "green") when the neural excitation produced by the perceived object is that normally produced by light waves reflected from surfaces that the subject sees to be, respectively, red or green.

12. The process of fusion is complicated. See Haber and Hershenson (1980, 232–40, 247–52) for a description of the process.

13. Feigl (1967) says that "the step from parallelism to the identity view is essentially a matter of philosophical interpretation" and that "the principle of parsimony as it is employed in the sciences contributes only one reason in favor of monism" (94). A few pages later, he says that "between the parallelism and the identity doctrines . . . there are no [empirically testable differences]" and that the issue is in that respect "similar . . . to such 'metaphysical' issues as realism versus phenomenalism. . . . These issues unlike disputes regarding scientific theories cannot be decided by empirical test [but are rather] a subject matter for logical analysis" (96–97). See Carnap 1963 (882–86) for a complementary view of the issue.

Smart says: "There is no conceivable experiment that could decide between materialism [the brain-process theory] and epiphenomenalism [dualism]. . . . If it be agreed that there are no cogent philosophical arguments which force us into accepting dualism, and if the brain-process theory and dualism are equally consistent with the facts, then the principles of parsimony and simplicity seem to me to decide overwhelmingly in favor of the brain process theory" (1959, 165–66).

14. Arguments that Leibniz's law cannot be rejected without contradiction are frequently advanced, all apparently begging the question at issue. Here is a representative example: "For some property P, it is alleged [by those who reject Leibniz's law] that *a* has P, while *b* lacks P. Yet *a* and *b* are the same entity. A single entity is thus alleged both to have P and not have P, which is contradictory." (For another example, see Kim 2006, 101.) This argument relies on the following suppressed subargument: "Since *a* is identical with *b* and *a* has P, *b* has P," which is an application of Leibniz's law, the very law in question. The argument's persuasiveness is due in large measure to its use of the phrase "a single entity," which is ambiguous. The phrase must be understood in terms of identity—*a* and *b* are a single entity if and only *a* is identical with *b*. But according to the view here, we have a choice between logical identity governed by Leibniz's law and empirical identity not so governed, and only the latter is appropriate for such cases as the identity of a sample of water with a collection of H_2O molecules.

15. For a brief indication of how the claim can thus be extended, see Savage 2001, especially pages 130–32. For a full explanation see Savage 1970, chapter 10.

16. Schopenhauer's term—*Weltknoten*—is adopted by Feigl (1967, 6), who characterizes the problem broadly as "a cluster of intricate puzzles—some scientific, some epis-

temological, some semantical, and some pragmatic [to which are related] sensitive and controversial issues regarding teleology, purpose, intentionality, and free will." The loops in the knot most relevant to the present chapter are puzzles regarding psycho-physical (mind-body) identities and macrophysical-microphysical identities, which I discuss next.

17. A case of this type was experienced by the author when the flash of a camera at a child's birthday party produced in his visual field a round, purple afterimage that he initially mistook for a purple balloon floating among others of the same and different colors in the room.

18. In 1980 a revised version with an important preface and some useful addenda was published. References are to this version.

19. This example—without a specific number—is from Kripke (1980, 35), where it is used to show that an a priori statement, that is, one that can be known without using experience, can at the same time be a posteriori, one that can be known through experience alone. The example can also be used to argue, as Kripke does on page 36 using Goldbach's conjecture, that a necessary statement can at the same time be a posteriori, and it is simpler so to use it. However, with either example the argument is ineffective. A person who comes to know that a number is (or is not) prime through a computer's calculations is directly using experience of the computer's result and reliability and indirectly using its calculation, the latter of which is not experiential. The epistemic process involved is analogous to using the calculations of another person to establish a mathematical truth. Knowledge thus obtained is based not solely on experience but also on the nonexperiential knowledge of the informant.

Kripke holds that identity statements such as "H_2O is water" and "Pain is neural excitation" are *logically* necessarily true (true in all logically possible worlds) and at the same time a posteriori—knowable when true through experience. The passage in question is in part an attempt to buttress his view by providing an example of a statement of this type that is not an identity statement. I doubt that any such example exists. On the view presented here, "H_2O is water" and "Pain is neural excitation" are indeed a posteriori and necessarily true, but they are *empirically* necessarily true.

20. Where Kripke says "seems possible" or "is possible," others use the more familiar "is conceivable." Since something's "seeming possible" must be based on some epistemic activity of reasoning, intuiting, conceiving, imagining, etc., it is tempting to employ one of the epistemic terms. In lieu of a theory of the relevant epistemic activity, the unanalyzed familiar terminology of "conceiving" will often be employed here.

21. The first of these possibilities is supposed to show that mind-body type identity—the identification of types of mental entity with types of physical entity—is false. The second possibility is supposed to show that even mind-body token identity—the identification of an individual of whatever mental type with an individual of whatever physical type—is false. Kripke's argument from his law applies to identifications of both kinds, as he points out in note 73, page 144.

22. Kripke (142ff.) attempts to explain away the apparent inconsistency as follows: "When we think we imagine heat without molecular motion we are in reality imagining the sensation of heat (which was used to fix the reference of 'heat') without heat. That we imagine heat without molecular motion is a conceptual illusion. But when we imagine pain without neural excitation, it cannot be that we are in reality imagining the sensation of pain without pain. For the sensation of pain is not distinct from the pain. It is therefore no illusion that we imagine pain without neural excitation."

This explanation is unsuccessful. By imagining heat as a subtle, invisible fluid I imagine heat as not identical with molecular motion and at the same time distinguish heat from the sensation of heat. By imagining water as nonparticulate, composed only of smaller and smaller volumes of water ad infinitum, I imagine water as not identical with H_2O while distinguishing water from the sensation of water. That the explanation is unsuccessful is unsurprising. For it is offered to explain how, *given that heat is necessarily identical with molecular motion and hence not possibly not identical with molecular motion,* it can seem possible that heat is not identical with molecular motion. Then it is shown that a similar explanation cannot be provided for how it can seem possible that pain is not neural excitation. The class of permissable explanations is thus artifically and question-beggingly limited to those that can show only how an identity can seem to be possibly false when it is necessarily true, excluding those that show how an identity can seem to be possibly false when true empirically.

23. Kripke's view is that "water" is not an abbreviation of a definite description, but rather a name for a kind of substance whose superficial, inessential properties are transparency, liquidity, and so on, and whose essence is now known to be H_2O. The essence or essential nature of a kind of substance is that property or set of properties without which the substance could not be what it is. The essence of "water" was unknown when the word first came into use. Nonetheless, the use of the word in the presence of its inessential properties—transparency, liquidity, and so on—"fixed" H_2O as the reference of "water" even before the essence was known. We now know that there is no possible world in which water is gelatinous globules or anything other than H_2O.

This essentialist view entails Kripke's theory that identities such as "Water is H_2O" are necessarily true if true, and so it inherits all objections to that theory. An independent objection is that the view incorrectly assumes that the reference of a denoting expression is not completely determined by the linguistic intentions and practices of the speaker or by the linguistic community of the user(s). Ancient scientists and the other speakers in their communities did not know or believe that water is H_2O—never having heard of H_2O—and therefore could not have been using their synonym for "water" to refer to H_2O. They intended it to refer to and used it to refer to the transparent, odorless liquid found in streams, lakes, and oceans of earth. And the reference of the term was "fixed" by what they intended and used the term to refer to, not by its being uttered in the presence of H_2O. Even if the substances to which they applied the term had been gelatinous globules, its reference would have been the same.

Even if this objection is set aside, the essentialist view does not support, and indeed undercuts, Kripke's argument that pain is not neural excitation. For he admits that our knowledge of the essences of empirical kinds such as water and pain is empirical, and pain is an empirical kind. Empirical science has discovered that water is H_2O. Whether it has also discovered that pain is neural excitation is controversial, at least among philosophers. But it is clear that science has not discovered pain to possess some essence other than neural excitation. Therefore, Kripke cannot argue from any such discovery that pain is not neural excitation, and, as I have shown, his possibility test does not provide any support for this conclusion. It is difficult to see how the test could provide such support, the test being a priori and the conclusion empirical and thus apparently capable only of a posteriori support.

24. Deeper circularities are also involved. The suggested definition of identity is metalinguistic, which in formal terms reads as follows: *a* is identical with *b* if and only if (1) *'a'* and *'b'* are rigid and unambiguous designators, and (2) there is an object *x* such that *'a'* denotes *x* and *'b'* denotes *x*. The notions of rigidity and lack of ambiguity in clause (1) cannot be defined except by means of the notion of identity. An unambiguous designator is by definition one that denotes *one and only one* object (every denoted entity being *identical* with the one) in any world in which it denotes something. A rigid designator is

by definition one that denotes the *same identical* object in every world in which it denotes something. In clause (2) the scope of the existential quantifier in the definiens must be understood as requiring that in every interpretation (model) of the sentence the object assigned to the first occurrence of *'x'* is *identical* with the object assigned to the second occurrence of *'x.'* These circularities prevent the metalinguistic definition from uniquely specifying logical identity. For if, as claimed here, there is another type of identity—empirical identity—and the italicized phrases explaining the terms of the definition are taken to refer to that type, then the definition becomes a definition of empirical identity.

As regards definitions in the object language, Kripke's is of course not a candidate since it is metalinguistic, essentially the metalinguistic one just considered. Apparently the only candidate is the one that employs Leibniz's law: (x)(y)(x = y if and only if (F)(Fx if and only if Fy), where 'F' is a variable ranging over properties. But it is less a definition than a characterization, and apparently cannot be derived from any definition or deeper principle that is not circular. As a result, although it qualifies as a characterization of logical identity, the claim that it characterizes the only legitimate concept of identity cannot be justified.

25. An equivalence relation is symmetrical, reflexive, and transitive. In symbols, relation R is an equivalence relation if and only if $(x)(y)(x\mathrm{R}y$ and $y\mathrm{R}x)$, $(x)(x\mathrm{R}x)$, and $(x)(y)(z)$(if $x\mathrm{R}y$ and $y\mathrm{R}z$ then $x\mathrm{R}z$).

26. Feigl (1967, 97) must have some such analogy in mind if, as it seems, he compares the difference between parallelism and the identity theory with the difference between Hume's analysis of causation and the modal version of Hume's analysis.

27. For the cognitivist analysis of sensation and perception that underlies this reply, see Savage 1989.

28. This chapter contains two distinct replies to Kripke's argument against mind-body identities. The first reply accepts his central axiom (*) (page 151) and his metalinguistic definition of empirical identity and maintains that (2) is the false conjunct of the antecedent because at least one of 'A' and 'B' functions as a definite description and so is not logically rigid. The second, later reply rejects his metalinguistic definition and proposes that "A is empirically identical with B" means that A and B have the same spatiotemporal location. Under this definition the central axiom becomes false. Even assuming that "water" and "H_2O" are logically rigid designators, "Water has the spatiotemporal location of H_2O" is true at the actual world and false at a possible world in which water has the spatiotemporal location of gelatinous globules, or at a world in which H_2O does not exist.

The second reply seems preferable. It avoids limiting the rigidity of empirical designators to empirically possible worlds, a limitation that seems ad hoc. And it does not accept the metalinguistic definition of identity, a definition that is implicitly circular. It also provides a unified reply to the objections from Leibniz's law and Kripke's law. Although a drop of water has the spatiotemporal location of a drop of H_2O and is therefore identical with H_2O, water is continuous and H_2O is not. And there are possible (conceivable) worlds in which a drop of water is a drop of gelatinous globules. A corresponding analysis applies to the identification of pain with neural excitation.

References

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

Chalmers, D. J. 1996. *The Conscious Mind*. New York: Oxford University Press.

Feigl, H. 1958. “The ‘Mental’ and the ‘Physical.’” In *Concepts, Theories, and the Mind-Body Problem,* ed. H. Feigl, M. Scriven, and G. Maxwell, 370–497. Minneapolis: University of Minnesota Press.

———. 1963. “Physicalism, Unity of Science, and the Foundations of Psychology.” In *The Philosophy of Rudolf Carnap,* ed. P. A. Schilpp, 227–67. La Salle, Ill.: Open Court.

———. 1967. *The “Mental” and the “Physical”: The Essay and a Postscript.* Minneapolis: University of Minnesota Press.

Gunderson, K. 1970. “Asymmetries and Mind-Body Perplexities.” In *Analyses of Theories and Methods of Physics and Psychology,* ed. M. Radner and S. Winokur, 273–309. Minneapolis: University of Minnesota Press.

———. 1974. “The Texture of Mentality.” In *Wisdom: Twelve Essays,* ed. R. Bambrough, 173–93. Oxford: Oxford University Press.

Haber, R. N., and M. Hershenson. 1980. *The Psychology of Visual Perception.* 2nd ed. New York: Holt, Rinehart and Winston.

Kim, J. 2006. *Philosophy of Mind.* 2nd ed. Boulder, Colo.: Westview Press.

Kripke, S. 1980. *Naming and Necessity.* 2nd ed. Cambridge, Mass.: Harvard University Press. See especially Lecture III, 106–55.

Maxwell, G. 1976. “Scientific Results and the Mind-Brain Issue: Some Afterthoughts.” In *Consciousness and the Brain,* ed. G. Globus, G. Maxwell, and I. Savodnik, 329–58. New York: Plenum Press.

———. 1978. “Rigid Designators and Mind-Brain Identity.” In *Perception and Cognition,* ed. C. W. Savage, 365–403. Minneapolis: University of Minnesota Press.

Meehl, P. 1966. “The Compleat Autocerebroscopist.” In *Mind, Matter, and Method,* ed. P. Feyerabend and G. Maxwell, 103–80. Minneapolis: University of Minnesota Press.

Place, U. T. 1956. “Is Consciousness a Brain Process?” *British Journal of Psychology* 47: 44–50.

Russell, B. 1948. *Human Knowledge: Its Scope and Limits.* London: George Allen and Unwin.

———. 1959. *My Philosophical Development.* New York: Simon and Schuster.

Savage, C. W. 1970. *The Measurement of Sensation.* Berkeley: University of California Press.

———. 1975. “The Continuity of Perceptual and Cognitive Experiences.” In *Hallucinations, Behavior, Experience, and Theory,* ed. R. K. Siegel and L. J. West, 275–86. New York: John Wiley.

———. 1989. “Epistemological Advantages of a Cognitivist Analysis of Sensation and Perception.” In *Science, Mind, and Psychology,* ed. M. L. Maxwell and C. W. Savage, 61–84. Lanham, Md.: University Press of America.

———. 1992. “Foundationalism Naturalized.” In *Cognitive Models of Science,* ed. R. N. Giere, 207–36. Minneapolis: University of Minnesota Press.

———. 2001. “In Defense of Color Psychophysicalism.” *Consciousness and Cognition* 10: 125–32.

Smart, J. J. C. 1959. “Sensations and Brain Processes.” *Philosophical Review* 68: 141–56.

Carla Fehr

8
Explanations of the Evolution of Sex: A Plurality of Local Mechanisms

Sex represents a pointed challenge to evolutionary theory. On the one hand, sexual reproduction is ubiquitous and abundant; on the other hand, it is incredibly expensive. Imagine two organisms, identical except for the fact that organism A reproduces asexually and organism B reproduces sexually. Organism A will pass its entire complement of genes on to its daughters while organism B passes on only 50 percent. From an evolutionary perspective sex has a twofold cost. This means that for a sexual population to resist invasion by an asexual clone, or for sexual and asexual reproduction to exist concurrently, or for sexual reproduction to evolve from asexual reproduction, there must be some benefit to reproducing sexually that overcomes this cost (Maynard Smith 1978; Williams 1975). Since it is not only obvious that sex exists but also that it is common, biologists since Darwin have been searching for this benefit and have come up with over twenty kinds of explanations for the evolution of sex (Kondrashov 1993).[1] Supporters of most of these explanations hold a traditional, monistic view. They work under the assumption that these explanations are in competition with one another and that one will eventually prevail over the others. But a growing number of scientists are willing to consider a plurality of explanations for this phenomenon.

In previous work I have defined *explanatory pluralism* as the state of affairs in which more than one explanation is required to account for a phenomenon. According to explanatory pluralism, some of our best scientific accounts of the world are incomplete, and accounts that explain one aspect can obscure other aspects of the same phenomenon. Explanatory pluralism can be contrasted with *explanatory monism,* which assumes that there is a single best possible account for every phenomenon. Explanatory monism is consistent with the scientific monism described by Kellert, Longino, and Waters in the introduction to this volume. They describe scientific monism as the view that "the ultimate aim of a science is to establish a single, complete, and comprehensive account of the natural world." Scientific monism assumes that, in principle, the world can be explained by such an account and that there are, or can be, methods that can yield such an explanation. Further, according to scientific monism, methods and theories ought to be evaluated on the basis of their ability to provide such an account.

Kellert, Longino, and Waters define modest pluralists as those who acknowledge a plurality of accounts, but treat that plurality as resolvable. *Modest pluralism* is a state of affairs in which multiple explanations of a phenomenon are tolerated because it is expected that they will eventually resolve into monism. Some accept modest pluralism because the world is patchy, and a phenomenon may require different explanations in different patches. This case slides into monism because it is assumed that there is a single best explanation relative to a particular patch, and a conjunction of these explanations may account for the phenomenon in its entirety. I consider cases in which multiple explanations can be subsumed under a single explanatory framework as cases of modest pluralism because the conjunction of these explanations can be viewed as a complicated case of monism. I argue that the evolution of sex is best explained not by monism or by modest pluralism, but by explanatory pluralism. Explanation of the evolution of sexual reproduction requires multiple accounts, which cannot be integrated with one another without loss of content or explanatory information.

Three Explanations for the Evolution of Sex

Sexual reproduction can be broadly defined as including any reproductive systems and life cycles that involve an exchange of DNA. More specifically, sex is often defined as those life cycles that include the formation of eggs and sperm involving a process called meiosis, and the subsequent joining of those gametes during fertilization. The *Red Queen explanation* (RQ) takes its name from the queen in the story *Through the Looking Glass,* who points out that in her country one had to run as fast as one could just to stay in the same place. The RQ argues that sex is beneficial because it allows sexually reproducing organisms to stay adapted to a continually changing biotic environment (Hamilton 1980; Lively 1987, 1989). In particular, this explanation depends on coevolution between host organisms and the parasites that attack those hosts. Parasites evolve rapidly and can adapt to overcome host resistance during the lifetime of a host organism. When hosts reproduce sexually, they create genetically variable offspring. Such offspring are identical to neither parent. Sexually reproducing parents can produce rare offspring with parasite resistance mechanisms to which parasite populations are not already adapted. This explanation postulates a time lag oscillation between hosts and parasites in which negative frequency-dependent selection protects organisms that carry genes for sexual reproduction. The RQ explanation is powerful because coevolution has been shown to occur between hosts and parasites, and parasite infection is an omnipresent selective force. This explanation provides some understanding of the ecological distribution of sex; sexual reproduction is more commonly found in environments in which parasite pressure is stronger. The weaknesses of the RQ are twofold. First, modeling has shown that the effect of the parasite on host fitness must be very high in order to induce an arms race between parasite and host populations. Second, the RQ selects for rarity among diversity. This diversity could be provided by subpopulations of clones as well as sexually produced offspring. Since these clones would not have to pay the cost of sex, oscillating clonal populations could drive the sexual population extinct (Dybdahl and Lively 1995a, 1995b).

Muller's Ratchet (MR) describes an asexual population's inability to recover genotypes that have been corrupted by deleterious mutation (Muller 1932, 1964). The portion of a population carrying the least number of mutations is called the least-loaded line. Chance events will lead to the extinction of the least-loaded line, either because the members of that line will acquire mutations or fail to reproduce. When this happens in an asexual population, the least-loaded line is lost forever because each individual passes on an exact copy of its DNA to its offspring. With the loss of the least-loaded line, the equilibrium number of mutations in that population goes up. The ratchet has advanced a notch and the process begins again. In this manner mutations accumulate and eventually lead to the extinction of the asexual population. In a sexual population, the least-loaded line can be recovered when individuals with different mutations mate. Because each individual only passes half of its genetic material to its offspring, it is possible to produce an offspring that has fewer mutations than either parent, and the least-loaded line can be re-created. Muller's Ratchet is a long-term group selection mechanism and has been used to explain why there are very few ancient asexual species. This model works best in small populations of organisms with large genomes and a high rate of mildly deleterious mutations. The drawbacks of this explanation are that (1) this mechanism functions too slowly to provide a short-term advantage to sexual reproduction; (2) it doesn't take obligate sexual reproduction to halt the ratchet, sex every few generations will do; and (3) the mechanism is sensitive to the form of selection against mutations. If there are strong interactions between mutations, the ratchet may slow down or even halt. Even given these caveats, MR is an important explanation of sexual reproduction. Biologist Graham Bell writes, "As a general rule, . . . no germ line can persist for geologically substantial periods of time in the absence of cross-fertilization" (1988, 158).

According to the *DNA Repair* explanation, the benefit of sex is due to the ability of sexual organisms to repair genome damage (Dougherty 1955).

Bernstein (1983), Bernstein, Hopf, and Michod (1988), and Bernstein, Byerly, Hopf, and Michod (1985a, 1985b) explain the evolution of sexual reproduction in terms of the function of the molecular mechanisms of meiosis. They considered the structure and biochemistry of the reactions that take place during meiosis, and the regulation of those reactions, and found that meiosis is very good at, and appears to be well designed for, repairing double-strand DNA damage.[2] Supporters of this view (Bernstein, Byerly, et al. 1985a; Michod 1995; Michod and Long 1995; Long and Michod 1995) argue that the explanations of both the origins and the maintenance of sex are "based on a selective advantage arising from recombinational repair of genetic damage." This repair process allows organisms to produce a greater number of genetically intact gametes and hence a greater number of viable offspring.

DNA damage is not a mutation but rather an alteration of the structure of a DNA molecule. This sort of damage occurs frequently and, if not repaired, results in the loss of genetic information, which can be harmful or lethal to an individual's gametes and offspring. During meiosis the homologous chromosomes of diploid organisms are brought into intimate association with one another.[3] While they are close together, one chromosome can copy information from the other and use it to patch damaged areas and recover lost information.

In this explanation, the force of the argument is found at molecular and cellular levels of biological organization in the elucidation of the biochemical steps of the cellular process of meiosis. It is a very different type of explanation from the other two described because it gains its strength not from population genetic modeling, but from a structural account of the biochemical steps that actually occur during meiosis. Other researchers are beginning to look at the evolution of sexual reproduction from a molecular level. Redfield (1999) forcefully argues that population geneticists ignore cytological and molecular function at their own peril.[4]

This chapter does not depend on the continued acceptance of these three explanations. The point is that they are all well-supported theories and that by attending to a plurality of theories researchers can consider the ways that theories conflict *and* interact. This is proving to be a fruitful approach to an old and difficult problem for evolutionary biology.

Pluralistic Approaches to the Evolution of Sex

While many biologists hope for a single answer to the problem of sex and conduct research under the assumption that these different theories are in competition with one another, there is a growing group of scientists who either happily or unhappily are starting to see a pluralistic approach as their best option. One researcher in the field writes:

> I do not like this possibility [of pluralism] because such a beautiful phenomenon as sex deserves a nice, simple explanation and messy interactions of very different processes would spoil the story. Of course, this does not mean that such interactions are not, nevertheless, essential. (Kondrashov 1999, 1031)

I will review three pluralistic approaches to the evolution of sex. These approaches involve different combinations of explanations, and their supporters have different reasons for endorsing pluralism.

Bell

Graham Bell (1988) argues that we need three explanations for the evolution of sex: DNA Repair, Muller's Ratchet, and something like the RQ.[5] Bell makes the obvious point that it is important to keep the germ line of a population in good working order, and since this system will naturally degrade, it must have a repair mechanism. There are two sorts of repair systems. One of these is endogenous repair, a conservative process in which a self-correcting system scrutinizes and repairs itself. This is the sort of DNA repair that Bernstein, Byerly, Hopf, and Michod argue is crucial in the evolution of sex. A second sort of repair system is exogenous repair, in which the repair mechanism and its instructions are outside of the system being repaired. This sort of repair mechanism only works in replicating systems because it entails a functional test of the adequacy of the offspring produced. The functional test is natural selection. Bell argues that endogenous repair cannot be perfect because the repair mechanism is itself vulnerable to corruption; hence, there is also a need for an exogenous repair mechanism.

Also, there are two sorts of genetic error: DNA damage, which both DNA repair mechanisms can detect and remedy, and mutation, which endogenous repair mechanisms cannot detect and which are inherited by an individual's offspring. The first of these explains the role of meiosis but does not address outcrossing. The second of these mechanisms requires outcrossing to function. Mutations are accumulated in an asexual population via Muller's Ratchet. According to this explanation, the accumulation of mutations in an asexual population will lead to their extinction in the long run due to this group selection process, while sexual populations, with the ability to re-create the least-loaded line, will maintain their viability. The taxonomic distribution of asexuality supports this argument. Bell writes, "The taxonomically isolated position of most groups of obligately parthenogenic [asexual] animals suggests that they are the short-lived offshoots of sexual stocks" (1988, 172). Even though the MR hypothesis explains the taxonomic distribution of sex, Bell believes that we still need to account for its ecological distribution. As a result Bell argues for something like the Red Queen hypothesis.

Bell is interesting because he argues that a complete account of sex will require explanations for three different aspects of sexual reproduction: endogenous DNA repair, which explains sex in terms of how the biochemical steps of meiosis ensure the health of an organism's gamete's DNA and apply to organisms that undergo meiosis as well as exogenous protection of the repair mechanism itself; Muller's Ratchet, which explains the presence of sex and the absence of taxonomically ancient asexual reproduction in terms of a group selection argument; and explanations such as the Red Queen, which explain the ecological correlates of sex and apply in many but not all ecological circumstances.

West, Lively, and Read

Howard and Lively (1994) and West, Lively, and Read (1999) note that although the existing explanations for the evolution of sex all describe some benefit that a sexual organism enjoys relative to an asexual organism, this benefit is rarely large enough to counteract the twofold cost of sex. In order to increase the benefit described by an explanation to the magic number "2," scientists are forced to impose strict and sometimes unrealistic restrictions on their models. In light of this, West et al. propose that if models don't formally contradict one another, then rather than searching for predictions that distinguish among them, we ought to look for ways that we can combine them and hence combine the benefit they attribute to sex.

They combine the RQ and the MR (Howard and Lively 1994).[6] Recall that the Red Queen requires very severe fitness consequences of parasite infestation and that the presence of multiple clones can erode the benefits of sex. The main drawback of the MR is that it needs small populations to work fast enough to provide a short-term benefit for sex.

When the RQ and the MR interact, they cover each other's weaknesses (Howard and Lively 1994; West, Lively, and Read 1999). Computer simulations show that this interaction of mechanisms can provide a sufficient benefit to explain sexual reproduction. The RQ drives the asexual clones through periods of decreased population numbers, during which time the clonal population is small enough for the MR to function reasonably quickly. The MR removes the asexual clones that would drive the sexual population extinct if the RQ alone were operating. The RQ also explains the ecological correlates of sex, which the mutation-based explanation cannot. In addition, the presence of mutations may make the effects of parasite infestation more severe; individuals carrying many mutations may be considerably sicker as a result of parasites or disease. West et al. propose pluralism because these synergistic interactions between the RQ and MR make it a better account of sex than either mechanism alone.

Fehr

I have argued that scientists need to employ multiple explanations to account for the evolution of sex because it is a complex phenomenon composed of imperfectly overlapping and interacting parts, or domains of phenomena (Fehr 2001a, 2001b). One particular and familiar grouping of the parts of sex includes the production of eggs and sperm, which importantly includes meiosis, and the mixing of these meiotically produced gametes during outcrossing, or cross-fertilization. Meiosis and outcrossing go hand in hand. But even though there is a functional unity to meiosis and outcrossing, they have contingent evolutionary histories that aren't necessarily connected and currently aren't actually connected in many extant organisms.

Currently the DNA Repair explanation can be used to account for meiosis and the RQ can be used to account for outcrossing. Sex in some domains will be explained by a constellation of explanations that includes the RQ and not DNA Repair, in other domains by a constellation that includes DNA Repair and not the RQ, and in yet others by a constellation that includes both or, for that matter, neither. The fact that sex can be divided into different domains that imperfectly overlap and require different explanations is constitutive of explanatory pluralism. In a later section of this chapter I take this argument a step further and consider the case of a single domain in which each of the members undergoes meiosis and outcrossing and that requires more than one explanation. I argue that this doesn't slide into a monistic view. First, I turn to a more general account of various challenges to pluralism.

Challenges to Pluralism: The Problem of Conjunction

Common challenges from the monist view include criticisms that pluralism simply represents ambiguity in the characterization of the phenomenon being explained or ambiguity in the explanation-seeking question being asked. These are possible reasons why more than one explanation may be proffered for a phenomenon. It is appropriate to address these issues, and in some cases it may turn out that pluralism is a temporary result of ambiguity. But it is not appropriate to assume that ambiguity is the sole cause of pluralism. In previous work, I rule out these exact two possibilities with regard to pluralism in the case of the evolution of sex (Fehr 2001a, 2001b).

A monist may also argue that pluralism is the result of the immaturity of the investigation of the phenomenon in question. From this point of view, accepting a pluralistic account with respect to a domain of phenomena is a cop-out; ceasing an investigation before a plurality of explanations has been resolved by the formulation of a single correct story is forfeiting a game that, in time, may well be won. But again, this is a case that time and not philosophical assumptions should rule on. It is important to note that problems of immaturity and ambiguity are not limited to situations in which pluralism exists. An ambiguous question in an immature area of investigation could very well lead to a single account of a phenomenon.

A crucial worry that a monist may have with pluralism is what I call the problem of conjunction. A monist may argue that if we were just more careful in our description of pluralistic accounts we would find that they represent single, complicated causal stories rather than a number of distinct positions. In other words, all pluralism is of the modest type and hence leads to monism. According to this perspective, the existence or acceptance of multiple explanations need not entail pluralism, but rather may represent a single complex explanation. Say that phenomenon X is accounted for by explanations A, B, and C. If A, B, and C do not conflict with one another, they may represent a single unified, albeit complicated, story. Sandra Mitchell's account of pluralism presses this point (Mitchell 1992, 2002). She argues that there may be several abstract models that one needs to employ to account for different idealizations of a complicated evolutionary phenomenon, but at the concrete level, each phenomenon is the result of a single causal history. Hence she concludes that there can be pluralism at the abstract but not the concrete level. According to Mitchell, the explanation of a phenomenon at the concrete level must be the conjunction of the events in that phenomenon's causal history that led to its existence.

With this position in mind, one needs to demonstrate that the explanations involved in a pluralistic account are importantly distinct and that it would be a mistake to construe them as parts of a single unified story. To clarify what makes one explanation importantly distinct from another, one must turn to the epistemic context in which the various explanations were developed and continue to function. I argue that the accounts of the evolutionary maintenance of sex are explanatory when considered in a particular epistemological context and that unifying the explanations of sex under a single framework obscures these differences in context. It is not only that contextual information is lost in the combination of these explanations, which in itself is a reason to consider these explanations as being distinct from one another. But it is this local contextual information that allows researchers to make a description of a mechanism an explanation.

If one were to push a position such as Mitchell's, there is one possible (although I will argue ineffective) unifying framework that comes to mind, which is that all of the explanations considered are mechanistic. Although Mitchell's approach is mechanistic (it entails the conjunction of a series of explanations of an event's causal history), mechanism as a unifying framework is not limited to this sort of historical argument. Mechanism could provide a conceptual unity to a plurality of explanations, meaning that an account is eligible as an explanation insofar as it is mechanistic. An argument along these lines would postulate that if multiple explanations count as explanatory in light of the same criteria, then there would be good grounds for considering those explanations as a group of conjuncts that slides into monism. Not only does mechanism not provide this unifying framework, a close analysis of the mechanisms involved in these explanations provides grounds for holding that the evolution of sex is a case of explanatory pluralism.

Mechanism

What is a mechanism? Salmon (1989, 1994) has provided much of our current philosophical ideas and intuitions concerning mechanism in relation to his causal mechanical view of explanation. Salmon defines causal mechanisms as "the underlying microstructures of what [researchers] endeavor to explain" (1989, 184). Explanation involves describing the causal structure that supports a phenomenon or describing the causal history of a phenomenon. Salmon points out that individuating these processes may be problematic and argues that the determination of whether a complex process is a single process or is an aggregate of processes is a pragmatic matter. In Salmon's view, an explanation is an idealization of a causal mechanism; the determination of just what aspects are idealized will depend on the context of particular scientific investigations. Salmon's conception of a mechanism relies on an account of causation as a process in which a conserved quantity is transmitted through an interaction of world lines. His approach has been criticized because the concept of a conserved quantity applies uneasily to the life sciences. But the characterization of causation as a process and his point that explanations can idealize different aspects of a causal history are useful.

Machamer, Darden, and Craver (2000) point out that although the concept of mechanism is often used in science and philosophy of science, it has not been explored as thoroughly as it deserves. They define "mechanism" as follows: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (2000, 3).

According to their view, activities produce changes in entities, moving from set-up conditions to termination conditions. This definition does not specify the level of organization of the entities or activities. The phrase "regular changes" refers to the fact that mechanisms run from beginning to end in a typical fashion. Complete descriptions of mechanisms exhibit productive continuity without gaps from the setup to termination. Productive continuities are what make the connections between stages intelligible (Machamer, Darden, and Craver 2002, 3).

Mechanisms can be multilevel, and different disciplines "bottom out" at different levels of organization. Machamer, Darden, and Craver argue that the point of describing a mechanism is to make a phenomenon intelligible. In their view, a description of a mechanism accomplishes this by including entities and activities that are "accepted as relatively fundamental or taken to be unproblematic for the purposes of a given scientist, research group or field" (2002, 13). They point out that the explanatory privilege of particular bottom-out activities, and the concept of intelligibility itself, may well be "historically constituted and disciplinarily relative" (22). This concept of mechanism, as entities regularly going through activities, is not hobbled by worries about causation and is also open to a pluralist analysis because mechanisms can occur at different levels of organization or even be multilevel; researchers determine the set-up and termination conditions and mechanisms can bottom out at different levels.

Both Salmon's more narrow approach and Machamer et al.'s wider approach allow for significant leeway in the individuation of mechanisms. For Salmon, explanations are idealizations of parts of causal processes; the choice of which part of the causal process to idealize is local and context dependent. Machamer et al. state that mechanisms render phenomena intelligible by showing how they are produced using the fundamental entities (and the activities that these entities undergo) of a particular discipline; what is fundamental will differ among disciplines. Also, Machamer et al.'s point that the choice of set-up and termination conditions is a pragmatic matter is copacetic with Salmon's point that different aspects of the world can be idealized in the description of a mechanism.

Within a causal mechanical framework of explanation, for an account to be explanatory, it must be mechanistic. Within these views of mechanism it is apparent that for a mechanism to function as an explanation there are several factors that need to be made explicit. This list of factors includes but may not be limited to the following: background conditions that allow researchers to recognize or discover a mechanism, the features of the world idealized in a description of a mechanism, the set-up and termination conditions of the mechanism, and the bottom-out entities and activities that make the phenomenon being explained intelligible. I turn to an examination of the epistemological contexts of the explanations of sex to investigate the mechanisms involved and analyze the role that these factors play in making these mechanisms explanatorily satisfactory.

Explanatory Contexts of Explanations of Sex

Helen Longino (2002) has developed links between pluralism and local epistemologies by arguing that differences in intellectual context can generate different explanations for a phenomenon. She lists various assumptions made in epistemological communities that create the intellectual architecture on which explanations hang. These include substantive assumptions about what the world is constituted of and the processes in the world that need to be explained, as well as methodological assumptions about how knowledge is created. Further, there are assumptions about the form of knowledge that involve the epistemic virtues that a community embraces. The characterization of such a local epistemology includes these methods and standards of a community and the sort of knowledge a community is seeking, the justification for the methods that are used, as well as pragmatic and professional reasons for these epistemological choices. Each of the three explanations of sex was developed in a different, local epistemological context. A closer look at these contexts will show that it is not just the fact that these explanations are mechanistic that make them explanatory, but rather that they are particular types of mechanisms.

Red Queen

The RQ explanation was developed in the fields of population genetics and evolutionary ecology. It is used to explain the short-term evolutionary benefit of sexual reproduction. There are three complicated selection mechanisms at play in the Red Queen explanation. These accounts of selection mechanisms are idealized models of selection at the level of the biological individual. Evolution by natural selection at the individual level involves the change in gene frequencies in a population from one generation to the next. This change is explained by the fitness of individuals of a particular type in that population. As a result, the set-up conditions for the mechanisms involved in the Red Queen explanation are the fitness of sexually reproducing individuals of a particular type, and the termination conditions are the fitness of the offspring that those individuals produce. These mechanisms span a single generation. The three mechanisms are hitchhiking, coevolution, and negative frequency-dependent selection. These mechanisms were developed over a long period of time in population genetics and have been honed until they became part of the standard arsenal of the population geneticist.

The RQ involves a *hitchhiking model of selection* (Maynard Smith and Haigh 1974) because an organism receives no direct benefit from being sexual. The genes associated with sex allow for the production of novel and rare genotypes that can resist parasites. It is this parasite resistance that is directly selected. Sex genes are hitchhikers; they tend to spread in a population because they become linked with genes that are associated with parasite resistance. Hitchhiking mechanisms are extremely important for explaining evolutionary change where there doesn't appear to be direct selection and for explaining lack of evolutionary change in cases where one would expect to find strong selection pressure.

The RQ, with its evolutionary interactions between host and parasite populations, is also part of a larger series of explanations based on *coevolutionary models*. Coevolution has been defined as an interaction in which "two or more lineages . . . evolve specifically and reciprocally in response to each other" (Futuyma 1986, 483). This type of mechanism has been used not only to explain evolutionary interactions between hosts and parasites, but also interactions between predators and prey (Stenseth and Maynard Smith 1984), herbivores and target plant species (Ehrlich and Raven 1964), and interactions among species that compete for a resource, the most famous example of this being the adaptive radiation of finches on the Galápagos Islands (Lack 1947).

Finally, the RQ is a case of *negative frequency-dependent selection* (Futuyma 1986). The fitness of a particular host genotype is not absolute. It depends on the frequency of that type in its population. The benefit depends not on having a particular genotype, but on having a rare genotype. Frequency-dependent selection has been employed to explain such phenomena as mate choice (Ehrman 1967) and prey choice (Clark 1962). Coevolution, negative frequency-dependent selection, and hitchhiking models of selection are all types of mechanisms that are so generally recognized within population genetics and evolutionary theory that they are listed and defined in standard textbooks. Even though the RQ explanation of sex was a novel explanation produced in the 1980s, the different mechanisms that it employs were all commonplace in the discipline before it was developed.

There are three types of evidence that have been used to support the RQ view. The majority of the support comes from theoretical evidence generated by computer-modeling techniques. Practical difficulties make the RQ very difficult to test empirically. Recently, researchers have developed a body of comparative and correlative evidence (for example, Lively 1987), as well as evidence based on experimental manipulations (such as Lively 1989). This explanation does a good job of explaining the ecological correlates of sexual reproduction. We tend to find sex in old stable habitats in which it is presumed that there would be a strong parasite presence.

Muller's Ratchet

Muller's Ratchet is also a population genetic mechanism. It is based on idealized models of selection at the level of the group. This is a long-term explanation because it relies on the generation time of populations or groups rather than the generation time of individuals. As a result, the set-up condition involves the birth of a new lineage, and the termination condition is the extinction of that lineage and the production of daughter lineages. Asexual populations that can't re-create the least-loaded line will accumulate mutations that will eventually lead to their extinction.[7] The loss of the least-loaded line is due to processes that have been modeled in many other circumstances (Futuyma 1986). The chance loss of the least-loaded line is due to a random walk process, in other words, genetic drift. The evidence for this explanation is based on computer models that are used to predict the mutation rates, population size, and genome size needed for this mechanism to obtain. This model does a good job of explaining the taxonomic distribution of sexual reproduction: ancient extant asexual organisms are extremely rare.

DNA Repair

The DNA Repair explanation for sex was developed in the fields of molecular, cellular, and evolutionary biology. Supporters of this explanation argue that selection occurs at the level of the biological individual because producing healthy gametes confers an obvious fitness advantage. But the force of this account does not lie with the evolutionary story. This explanation is powerful because it offers an idealized model of how the cellular steps of meiosis repair damaged DNA. In this explanation the set-up condition is the beginning of meiosis, and the termination condition is the end of meiosis. DNA repair occurs at the intermediate stages of this mechanism. The timescale of this mechanism is very short. Whereas the RQ explanation, with its reliance on a hitchhiking mechanism, describes an indirect benefit of sex, the DNA Repair explanation offers a direct benefit to sexual reproduction. The evidence for this explanatory mechanism consists of a detailed elucidation of the mechanism of meiosis, as well as experimental studies in which either DNA damage is induced or the repair systems in meiosis are inhibited.

Factors Dependent on Epistemological Context

In review, it is not surprising that the epistemological or explanatory context of these different explanations of sex vary given that they were developed in different biological subdisciplines. These mechanisms act at different levels of selection: the RQ and DNA Repair are cases of individual selection while MR is a group-selection explanation. These different explanations are used to account for different aspects of sexual reproduction. RQ focuses on explaining outcrossing, it demonstrates how a sexual population could resist invasion by an asexual clone, and it best accounts for the ecological correlates of sexual reproduction. DNA Repair accounts for the presence of meiosis. MR accounts for the taxonomic distribution of sex.

These explanations all rely on different kinds of mechanisms. The RQ and MR use types of mechanisms that are well accepted in the biological subdisciplines in which they originated. These explanations are novel and important, even though they rely on mechanism types that were previously developed and are frequently used to address a variety of problems within their disciplines of origin. The DNA Repair view receives its explanatory power from an engineering-type mechanism that works at a low level of biological organization, while the explanatory power of the other two lies with evolutionary stories that take place at the individual and population levels. The mechanisms all have different set-up and termination conditions. The timescale of these mechanisms ranges widely: DNA Repair spans a molecular process, the Red Queen spans a generation in a population, and Muller's Ratchet spans the life of an entire population.

These explanations refer to different kinds of evolutionary benefits of sex. The RQ focuses on the short-term, indirect benefit of sex. MR relies on a long-term, indirect benefit, and the DNA Repair view relies on a short-term, direct benefit. Finally, these different explanations are supported by very different kinds of evidence. MR is primarily supported by theoretical evidence provided by population genetic models. RQ was originally supported by population genetic models, but more recently has been supported by ecological studies that are both correlational and manipulative. DNA Repair is supported by studies that document and describe the repair mechanism as well as experimental manipulations of that mechanism. All of these factors constitute the epistemological context of these three explanations.

Returning to Longino's characterization of local epistemologies, these explanations assume different views about the composition of the world. They account for different aspects of sex. The Red Queen accounts for outcrossing, and the DNA Repair for meiosis. The group of organisms that undergoes outcrossing is not identical to the group of organisms that undergoes meiosis. If one is focusing on outcrossing, sex is defined in terms of outcrossing; if one is focusing on meiosis, sex is defined in terms of meiosis. If one is sorting sexual organisms from asexual organisms, the joints at which the world can be carved differ depending on which framework one is using to define sex.

In the various epistemic contexts the explanations focus on different processes. They differ in terms of the level at which selection is occurring, and they also differ because selection is the target process for two of the mechanisms but a molecular mechanism is the target process of the DNA Repair. The methods used to test these hypotheses also differ. MR, because of the timescale, is limited to computer-simulation and correlational studies. The RQ, although primarily tested in simulation and correlational studies, is beginning to be tested empirically. The evidence for the DNA Repair view is primarily descriptive, although there are some manipulative tests.

The supporters of these explanations are all seeking causal knowledge, but this knowledge is motivated by different factors. The MR and RQ views are motivated by taxonomical and ecological distributions of sex respectively, while the DNA Repair view focuses on descriptive knowledge of the repair process. The epistemic contexts of these explanations are clearly different. This in itself is interesting because the explanations are all aimed at the same phenomenon, sex. Since these contexts are different, combining the RQ, MR, and DNA Repair views into a single explanation would require abstracting the various explanations from their unique contexts.

Returning to the Pluralistic Accounts

The three pluralistic accounts of sex make salient different variables along which the contexts of these explanations differ. Bell's view reveals the need to explain different aspects of sex. West, Lively, and Read's account highlights different selection processes. My account brings to light different kinds of mechanisms acting at different levels of biological organization.

Bell

Bell offers an early account of pluralism that is revolutionary because he does not assume that the explanations involved must compete with one another. He makes it clear that a complete explanation of sex will need to account for several different aspects of sexual reproduction, including the fact that there are practically no ancient asexual species, as well as the fact that sex is more common in particular kinds of habitats. MR and the RQ account for these aspects of sex. Another aspect of sex that Bell contends is of central importance is the maintenance of the integrity of the germ line. He argues that something like the DNA Repair explanation is important for ensuring that the DNA passed from one generation to the next remains constant and that a selective mechanism is crucial for the maintenance of the DNA repair apparatus. Bell's account focuses on these explanatory needs and not on the nuts and bolts of the interactions among the actual explanations. Bell seems to be proposing a modest form of pluralism in which there is a one-to-one correspondence between an explanation and the aspect of sex that needs to be explained.

West, Lively, and Read

West et al.'s pluralistic account makes salient the interaction of mechanisms that occur at different levels of selection. From a biological perspective, this account is highly pluralistic. From a philosophical perspective, this account seems to be a case of modest pluralism that has already slid into monism. These two mechanisms, which act at different levels of selection, have been conjoined in a single model. This particular moment in the history of biology is fascinating because West et al., in combining the RQ and MR mechanisms, propose a significant change in the epistemic context of these investigations of the evolution of sex.

Before West et al. developed their pluralistic account, MR was considered to be inherently problematic because of the timescale on which it acts. This form of group selection stretches over the time it takes for populations to come into being and go extinct. Individual selection stretches over the generation time of a biological individual. Given the twofold cost of sex, individual selection would likely drive sexual populations to asexuality before the sexual population could reap the group selection benefits provided by the MR.

Further, although mathematical models have been used to investigate the effects of opposing group and individual selective pressures, at a more concrete level the vastly different timescales of these two levels of selection make it difficult to focus on both processes simultaneously. In fact, one of the weaknesses of various models of group selection is that the temporal scope of the process makes it extremely difficult to investigate empirically. That is why the evidence for MR is based on computer models and comparative taxonomic studies while the evidence for the RQ view is more open to the possibility of empirical studies. These empirical studies have become increasingly predominant.

The change in context that West et al.'s pluralism effected involves conceiving of a circumstance in which the temporal scope of different levels of selection converge. The action of the RQ forces invading asexual clones to cyclically go through population bottlenecks in which the size of the asexual populations are very small. MR can act quickly in small populations, forcing them to have shorter generation times. In combination, these two models act on a more similar timescale than they would act alone. In terms of Machamer, Darden, and Craver's work on mechanism, before West et al.'s synthesis the context in which these mechanisms were explanatory involved individuating them in such a way that the set-up and termination conditions existed on such different timescales that they couldn't practically be combined. In the new context the set-up and termination conditions of these mechanisms exist on a similar timescale and hence can be combined.

This change in context is highlighted by the fact that West et al. explicitly state that their new pluralistic approach requires a change in the methodological approaches and epistemic values of the scientific community. This pluralistic view, because of its complexity, is very difficult to test or falsify and definitely does not exhibit the traditional virtue of simplicity. West, Lively, and Read call for the replacement of falsifiability and simplicity with realism. Their model, with its intricacy, has a better chance of capturing the complexities of the system being investigated than other simpler, monistic, and more easily falsifiable models. They also point out that this change in values will result in a change in methods. They argue that researchers should stop focusing on developing predictions that will distinguish among these models and instead should focus on parameter estimation, which would allow them to investigate the strength of the various explanations involved in a pluralistic analysis.

Fehr

My account of pluralism makes salient explanations of sex focusing on different levels of biological organization. I point out that sex is a mechanism that is composed of partially overlapping biological kinds (Fehr 2001a). One of these kinds includes instances of meiosis, a molecular and cellular mechanism, and the other includes instances of outcrossing, an individual-level mechanism. These kinds are distributed differently among different species; in other words, there is a huge variety in extant mating systems—some of which self-fertilize with meiotically produced gametes, some of which cross-fertilize with gametes that are not produced by meiosis, as well as the mating system that is most familiar to us that combines outcrossing and meiosis. These kinds can also be distributed differently in the same species, or even in the same biological individual; there are many species that have mixed mating systems. For example, in the common morning glory, an individual plant will have some seeds that are the result of outcrossing and some seeds that are the result of self-fertilization.

If the two mechanisms that I refer to are required to explain sex in a single biological individual, why is this not a case of modest pluralism that suffers from the conjunction problem? Why is this not a case of Mitchell's monism at the concrete level? There are two reasons why the RQ and DNA Repair views cannot be combined under a single framework even when they both act in a single individual. First, these mechanisms are explanatory because they are based on accepted kinds of mechanisms in their epistemological contexts. Second, although these mechanisms act in an individual, they are explanatory in terms of abstracting from the individual, and this abstraction is different in the explanatory contexts of the RQ and the DNA Repair views.

There are three reasons why the RQ and the DNA Repair accounts are explanatory not simply because they are mechanistic but because they are examples of kinds of mechanisms accepted as explanatory in their respective epistemological contexts. First, each account is not accepted as explanatory in the context of the other account. It is not that those who hold the DNA Repair view consider the RQ to be a bad explanation; it is not accepted as a possible explanation, and vice versa. Population genetic critics of the DNA Repair view believe that at best this view could be an explanation for the origins of meiosis or diploidy, but that it is not an explanation for the maintenance of sexual reproduction (Maynard Smith 1987). Notice that this criticism relies on the assumption that sex is defined not in terms of meiosis but rather in terms of outcrossing. Supporters of the DNA Repair account point out that the RQ account relies on genetic variation. They argue that their analysis of meiosis shows that its function is to repair DNA, not to create genetic variation. Supporters of this view do not comment on the various modes of outcrossing. If mechanism was the criterion for determining what accounts get to be candidates for explanations, then both of these explanations ought to be considered explanatory, or at least as possible explanations, in the other discipline. This is not the case.

Second, these explanations are examples of the kinds of mechanisms that are well accepted in their respective epistemic contexts. Hitchhiking, coevolution, and frequency-dependent selection all have a long history in population genetics, are well respected, and have a record of explanatory success. These accepted kinds of mechanisms were brought to bear on the problems of sex in the particular constellation of mechanisms called the

RQ approach. The DNA Repair view is not based on previously existing theories in the way that the RQ view is. But it does represent an accepted way of generating knowledge in cellular and molecular biology. Much of the work of these disciplines is based on generating descriptive knowledge of the sequential biochemical reactions that take place in the cell and in looking at the conditions necessary for these sequential processes to work. The DNA Repair hypothesis is based on generating a similar knowledge of meiosis and then using that knowledge to hypothesize about the function of this process. The DNA Repair explanation brings approaches and methods that are well accepted in its epistemic context to bear on meiosis. These methods generated an engineering-type mechanism, which is a common kind of mechanism in that context.

Third, in terms of Machamer, Darden, and Craver's characterization of mechanism, these explanations bottom out at different levels of organization. In other words, the fundamental processes that make the phenomenon intelligible are different: the DNA Repair explanation relies on biochemical processes, and the RQ relies on selection processes. These explanations not only count as explanations, but also are recognized as explanations because they are examples of particular kinds of mechanisms acceptable in their different explanatory contexts. Decontextualizing these mechanisms by uniting them under a single explanatory framework forces one to ignore the factors that made them satisfactory explanations in the first place. My account of the evolution of sex represents a case of explanatory pluralism because in order for the mechanisms involved to function as explanations they need to be considered as a part of different explanatory contexts.

Another reason why the RQ and DNA Repair view cannot be combined under a single framework, even when they both act in a single individual, is because they employ different kinds of abstractions. These accounts are explanatory in terms abstracted from the individual, and this abstraction is different in the explanatory contexts of the RQ and the DNA Repair views. The RQ and DNA Repair explanations are both mechanisms, and both offer reasons why sex is selected for in the face of the twofold cost of sex. Obviously, because they focus on different definitions of sex, one is abstracting from meiosis and the other is abstracting from outcrossing. Even more important, these explanations make different sorts of abstractions in the construction of the idealized models that they employ. John Dupré (1993, 116, 133), in his arguments against reductionism, points out that reductionism fails because accounts at different levels of organization rely on different kinds of abstract individuals. A similar point about abstraction can be made here. The DNA Repair view, in its reliance on the structure and function of meiosis, abstracts away the individual idiosyncrasies of instances of meiosis to describe an idealized cellular process.

It involves the creation of an abstract cellular model of meiosis. The Red Queen explanation abstracts away idiosyncratic environmental, genetic, and other information to create idealized ecological and population models. In these models there is an abstract host and an abstract parasite. In this abstraction, the instances of meiosis and the abstract cellular model of meiosis disappear.

Further, the DNA Repair model removes individual variation in its description of a molecular process, but the RQ model relies on variation within a population. The RQ view involves idealizations of populations that have a particular structure. This structure is represented by frequencies of hosts with different forms of parasite resistance and frequencies of parasites with different ways of overcoming these resistance mechanisms. The abstraction in the DNA Repair explanation involves abstracting away the variation found in instances of meiosis, and the RQ model involves highlighting variation within the populations. The population and ecological factors needed to model the RQ explanation disappear in the abstraction necessary to model the DNA Repair mechanism. The instances of meiosis and the abstract model of meiosis disappear in the creation of RQ models. DNA Repair and the RQ represent explanatory pluralism and should not be subsumed under a single unifying framework because the abstractions involved in the formation of one explanatory model obscure the aspect of sex accounted for by the other explanatory model and vice versa. A critic, such as Mitchell, who supports monism at the concrete level may point out that this part of my analysis turns on a discussion of idealized models. But even in the most concrete cases, for example, when the RQ account is applied to a particular extant population, these abstractions are still present in the explanation.

The modest pluralist model proposed by West, Lively, and Read could take over the role of the RQ model in my analysis of the evolution of sex. West et al.'s model offers an explanation for outcrossing, but does not account for the cellular and molecular aspects of sexual reproduction. This is not a weakness of their explanation; they are merely focusing on one aspect of sexual reproduction. Both Bell and I argue that there are other aspects of sex that need to be accounted for, and hence we add repair mechanisms to our analyses.

Conclusion

In my pluralistic account of sex, there are two reasons why, even within a single species, or in some cases within an individual, the different explanations of sex cannot be unified under a single framework and hence represent a case of explanatory pluralism. First, the mechanisms involved function as explanations within different epistemological frameworks. Outside of those frameworks the explanations are not satisfactory. Also, the mechanisms that are acceptable types of accounts differ between these epistemic frameworks. Further, these mechanisms make sex intelligible by referring to different kinds of processes that exist at different levels of organization. Second, in these frameworks, different aspects of sex are abstracted in the construction of explanatory models. In these abstractions, the aspect of sex that is accounted for by one model obscures the aspect of sex accounted for by the other model and vice versa.

If one endorses explanatory monism, the current plurality of explanations for the evolution of sex represents a serious problem. But explanatory monism is an inappropriate standard for evaluating *this* science because the evolution of sex represents a case in which the nature of the phenomenon does not support a single unified account. This should not be surprising since the evolution of sex is complex and involves processes acting at different levels of organization. These different processes are investigated from different disciplinary perspectives that involve methods well suited to generating local accounts of parts of the phenomenon. Although I believe that the choice between monism and pluralism ought to be conducted on a case-by-case basis with reference to the particular part of the world and science in question, I think that we ought not to be surprised to find explanatory pluralism with respect to other cases of evolutionary phenomena involving processes that act at multiple levels of organization.

Notes

Many thanks to Travis Butler, Ondrea Fehr, Margaret Holmgren, Helen Longino, Peter Vranas, Ken Waters, Mark Wunderlick, and the contributors to this volume. This material is based on work supported by National Science Foundation Grant 0450821.

1. Bell (1982) also provides a useful classification of the various evolutionary explanations for sex.
2. Double-strand damage is damage to both strands as opposed to just one of the strands of a DNA molecule.
3. Diploid organisms are those organisms that have two copies of every chromosome and these two copies are called homologues.
4. Several criticisms have been raised against this explanation, primarily from within population genetics (Charlesworth 1989; Kondrashov 1993; Maynard Smith 1988) but many of these objections have been answered by Bernstein, Hopf, and Michod (1988). In my opinion the persistence of these objections, given Bernstein, Byerly, et al.'s responses, represents the Balkanization of different biological subdisciplines and itself is worth further study.
5. Bell has supported the Tangled Bank hypothesis, which postulates that sex is beneficial in complex, spatially heterogeneous environments. The Red Queen can also explain

the ecological correlates of sex and is more widely accepted. For the sake of simplicity I do not describe the Tangled Bank explanation.

6. West, Lively, and Read include yet another explanation, a mutational deterministic hypothesis, in their account of sex that I omit from this analysis for the sake of simplicity.

7. There are two taxonomic groups that appear to be old and are asexual, the bdelloid rotifers and the chaetonotus gastrotrichs.

References

Bell, G. 1982. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*. Los Angeles: University of California Press.

———. 1988. *Sex and Death in Protozoa*. Cambridge: Cambridge University Press.

Bernstein, H. 1983. "Recombinational Repair May Be an Important Function of Sexual Reproduction." *Bioscience* 33: 326.

Bernstein, H., H. Byerly, F. Hopf, and R. Michod. 1985a. "Genetic Damage, Mutation, and the Evolution of Sex." *Science* 229: 1277–81.

———. 1985b. "DNA Repair and Complementation: The Major Factors in the Origin and Maintenance of Sex." *The Origin and Evolution of Sex,* ed. H. O. Halvorson and A. Monroy, 29–45. New York: A. R. Liss.

Bernstein, H., F. Hopf, and R. Michod. 1988. "Is Meiotic Recombination an Adaptation for Repairing DNA, Producing Genetic Variation, or Both?" In *The Evolution of Sex,* ed. R. Michod and B. Levin, 139–60. Sunderland, Mass.: Sinauer Associates.

Charlesworth, B. 1989. "The Evolution of Sex and Recombination." *Trends in Ecology and Evolution* 4: 264–67.

Clark, B. 1962. "Balanced Polymorphism and the Diversity of Sympatric Species." *Systematics Association Publication* 4: 47–70.

Dougherty, E. 1955. "Comparative Evolution and the Origin of Sexuality." *Systematic Zoology* 4: 145–69.

Dupré, J. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, Mass.: Harvard University Press.

Dybdahl, M., and C. Lively. 1995a. "Host Parasite Interactions: Infection of Common Clones in Natural Populations of a Freshwater Snail *(Potamopyrgus antipodarum)*." *Proceedings of the Royal Society of London* B260: 99–103.

———. 1995b. "Diverse, Endemic, and Polyphyletic Clones in Mixed Populations of a Freshwater Snail *(Potamopyrgus antipodarum). Journal of Evolutionary Biology* 8: 385–98.

Ehrlich, P., and P. Raven. 1964. "Butterflies and Plants: A Study in Coevolution." *Evolution* 18: 586–608.

Ehrman, L. 1967. "Further Studies in Genotype Frequency and Mating Success in *Drosophila*." *American Naturalist* 101: 415–24.

Fehr, C. J. 2001a. "The Evolution of Sex: Domains and Explanatory Pluralism." *Biology and Philosophy* 16: 145–70.

———. 2001b. "Pluralism and Sex: More Than a Pragmatic Issue." *Philosophy of Science* 68: S237–50.

Futuyma, D. 1986. *Evolutionary Biology.* 2nd ed. Sunderland, Mass.: Sinauer Associates.

Hamilton, W. 1980. "Sex versus Non-Sex versus Parasite." *Oikos* 35: 282–90.

Howard, R., and C. Lively. 1994. "Parasitism, Mutation, and the Maintenance of Sex." *Nature* 367: 554–57.

Kondrashov, A. 1993. "Classification of Hypotheses on the Advantage of Amphimixis." *Journal of Heredity* 84, no. 5: 372–87.

———. 1999. "Being Too Nice May Not Be Too Wise." *Journal of Evolutionary Biology* 12: 1031.
Lack, D. 1947. *Darwin's Finches*. Cambridge: Cambridge University Press.
Lively, C. 1987. "Evidence from a New Zealand Snail for the Maintenance of Sex by Parasitism." *Nature* 328: 519–21.
———. 1989. "Adaptation by a Parasitic Trematode to Local Populations of Its Snail Host. *Evolution* 43, no. 8: 1663–71.
Long, A., and R. Michod. 1995. "Origin of Sex for Error Repair: 1. Sex Diploidy and Haploidy." *Theoretical Population Biology* 47: 18–55.
Longino, H. 2002. *The Fate of Knowledge*. Princeton, N.J.: Princeton University Press.
Machamer, P., L. Darden, and C. F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67, no. 1: 1–25.
Maynard Smith, J. 1978. *The Evolution of Sex*. Cambridge: Cambridge University Press.
———. 1988. "The Evolution of Recombination." In *The Evolution of Sex,* ed. R. Michod and B. Levin, 106–25. Sunderland, Mass.: Sinauer Associates.
Maynard Smith, J., and J. Haigh. 1974. "The Hitchhiking Effect of a Favorable Gene." *Genetical Research* 23: 23–35.
Michod, R. 1995. *Eros and Evolution: A Natural Philosophy of Sex*. New York: Addison Wesley.
Michod, R., and A. Long. 1995. "Origin of Sex for Error Repair: II. Rarity and Extreme Environments." *Theoretical Population Biology* 47: 56–81.
Mitchell, S. 1992. "On Pluralism and Competition in Evolutionary Explanations." *American Zoologist* 32: 135–44.
———. 2002. "Integrative Pluralism." *Biology and Philosophy* 17: 55–70.
Muller, H. 1932. "Some Genetic Aspects of Sex." *American Naturalist* 66: 118–38.
———. 1964. "The Relation of Recombination to Mutational Advance." *Mutation Research* 1: 2–9.
Redfield, R. J. 1999. "A Truly Pluralistic View of Sex and Recombination." *Journal of Evolutionary Biology* 12: 1043–46.
Salmon, W. 1989. "Four Decades of Scientific Explanation." In *Scientific Explanation,* ed. P. Kitcher and W. Salmon. Minnesota Studies in the Philosophy of Science, vol. XIII, 3–196. Minneapolis: University of Minnesota Press.
———. 1994. "Causality without Counterfactuals." *Philosophy of Science* 61: 297–312.
Stenseth, N., and J. Maynard Smith. 1984. "Coevolution in Ecosystems: Red Queen Evolution or Stasis." *Evolution* 38: 870–80.
Waters, C. K. 1998. "Causal Regularities in the Biological World of Contingent Distribution." *Biology and Philosophy* 13, no. 1: 5–36.
West, S. A., C. M. Lively, and A. F. Read. 1999. "A Pluralist Approach to Sex and Recombination." *Journal of Evolutionary Biology* 12: 1003–12.
Williams, G. C. 1975. *Sex and Evolution*. Princeton, N. J.: Princeton University Press.

9
A Pluralist Interpretation of Gene-Centered Biology

Genes are at the center of research throughout many of the biological sciences, especially the sciences aimed at explaining what is happening within organisms. I call this "gene-centrism." The common *interpretation* of gene-centrism, advanced not just by popular writers but also by many scientific leaders, is that genes take center stage because biologists have learned that genes direct all the important developmental and metabolic processes in living systems. It is often reported that genes provide the information, the master plan or the blueprint, for the development of individual organisms. The reason biologists are (and should be) preoccupied with genes, according to this interpretation, is because biologists know that once they learn what all the individual genes of an organism do, they will understand everything essential about what is going on within the organism.

Gene-centrism has attracted a lot of criticism in the philosophical community. Criticisms are often based on the premise that development is brought about by, and living systems are sustained by, dynamic interactions among many kinds of elements. Privileging genes is allegedly inappropriate because focusing attention on genes (and on DNA) obscures the dynamics of the developmental systems of which they are only parts.[1] Developmental systems include DNA, and also RNA, enzymes, lipids, cell membranes, and ecological environments. This kind of critique has attracted a lot of attention among philosophers of biology, and some are actively advancing an alternative proposed by Susan Oyama called DST (developmental systems theory). Although it is unclear what proportion of practicing biologists are seriously entertaining DST (or what proportion are even aware of it), a good proportion of philosophers of biology are supportive, or at least sympathetic.

The view among philosophers on the DST wagon seems to be that gene-centrism rests on fundamentally flawed concepts and a seriously mistaken global theory about the way organisms function.[2] It is difficult to read philosophical critiques of gene-centrism without thinking that biologists' obsession with genes will impede the advance of scientific knowledge and lead to harmful social policies if it is not defeated and superceded by DST (or a close analogue).

It is not clear that gene-centrism (or DNA-centrism) impedes the advance of science. In fact, the strategy of focusing research on genes has led to many advances in biologists' investigations of a wide variety of phenomena, ranging from the early stages of embryological development to the variation of traits in populations. It seems to me that philosophers should be interested in understanding the nature of this research, not dismissing it on the basis of abstract theoretical arguments. I think this is particularly important because misunderstandings of this scientific research are leading society toward mistaken views and harmful policies. On the issue of harmful policies based on gene-centrism, I agree with the critics. But I am not convinced that the problem stems from the practice of placing genes and DNA at the center of biological research. I think it is appropriate, for instance, that much of the research into the workings of yeast, nematodes, and fruit flies centers on genes and DNA. The problem stems not so much from preoccupation with genes and DNA, but from the common interpretation of this preoccupation. I will argue that this interpretation is mistaken and that the roots of this misinterpretation go back to our basic assumptions about the nature of scientific knowledge.

Two common assumptions about scientific knowledge impede the understanding of gene-centrism. The first, what I call scientific monism, includes the idea that a successful science ought to provide a single, comprehensive account or theory of whatever part of the world it is intended to cover. The second, perhaps more common among philosophers than others, is that successful sciences are ultimately based on such comprehensive theories and that the activities of scientists are organized around the attempt to validate and fill out the grand theory on which their science is allegedly based. Together, these assumptions lead advocates of gene-centrism to infer that because gene-centric science is successful, it must be based on a comprehensive explanatory theory that identifies the essentials behind the processes being investigated. These epistemological assumptions lead critics, on the other hand, who recognize that gene-centric science obscures a lot of essentials, to infer that because gene-centric science does not provide a comprehensive account of everything relevant to complicated phenomena, such as the development of fruit flies, that it must not be a good science of complicated phenomena. The aims of this paper are (1) to argue that the dangers associated with gene-centrism stem from a bad interpretation of gene-centrism and (2) to introduce and defend a pluralist interpretation of gene-centered science.

I begin my argument in the next section by examining the DST critique to make salient the underlying epistemological assumption of monism. I go on to briefly describe the common misinterpretation of gene-centrism, which is advanced by leading scientists as well as critics of gene-centrism. I introduce a new interpretation of gene-centrism based on the epistemological idea that a body of scientific knowledge should be understood as one of a potential plurality of viable approaches for investigating and explaining a domain, which raises an important question. If gene-centrism is viewed as one of a potential plurality of approaches, and if the gene-centered approach fails to provide a comprehensive account of complicated biological processes, then why does it dominate biologists' attention? An important part of the answer is that a scientific approach can be organized around a central investigative strategy rather than a grand theory. Gene-centrism dominates because it is organized around an investigative strategy with broad investigative reach. I argue that genes are central to the investigation of a broad range of biological processes not because they figure into a comprehensive theory covering the processes being investigated, but because they provide a useful entry point for investigating and modeling those processes. I conclude by suggesting dangers posed by gene-centrism are largely due to the mistaken ideal, accepted by gene-centrists as well as their foes, that successful science must offer a comprehensive and global theory of the phenomena being investigated.

The DST Critique and Monism

Susan Oyama has provided much of the basis and inspiration for philosophers who criticize gene-centrism and advance DST as an alternative, and it will be helpful to examine possible connections between her views and monism. It turns out that she is more of a pluralist than many of her philosophical followers, but I will argue that her criticism ultimately relies on a demand for a comprehensive and all-inclusive theory or framework, a demand linked to monism. Since my target is the common misinterpretation of gene-centric biology and not DST per se, I will not try to offer a general description of the elusive DST alternative (Oyama, Griffiths, and Gray 2001 offers an introductory account). Instead, I will examine one of the themes at the heart of the DST critique, the idea that dichotomizing between inside forces and outside forces—for example, between nature and nurture, between genes and environment, or between internal constraints and natural selection—somehow misses *the* point: "I submit that these quarrels over causal responsibility miss the point" (Oyama 2000, 99).

Oyama suggests that the only way to resolve sterile debates about causal responsibility is to withdraw such distinctions. She writes:

> No one claims that genes alone are sufficient for development, or denies that environments, organic and inorganic, microscopic and macroscopic, internal and external, change over organismic and generational time. What is missing from most accounts is the synthetic processes of ontogenetic construction. Inheritance is not atomistic but systemic and interactive. It is not limited to genes, or even to germ cells, but also includes developmentally relevant aspects of the surround—and "surround" may be narrowly or broadly defined, depending on the scope of the analysis. Inheritance can be identified with "nature" only if it embraces all contributors to that nature, and nature does not reside in genes or anywhere else until it emerges in the phenotype-in-transition. Nature is thus not properly contrasted with nurture in the first place; it is the product of a continual process of nurture. (Oyama 2000, 71–72)

First, let me note that I agree with much of what is stated here. I will not take issue with what Oyama or her sympathizers in the philosophical community say about what is happening within biological systems. My disagreement concerns tacit epistemic premises about the nature of scientific knowledge.

Biological research, much of it gene-centered, has indeed shown that inheritance is systematic and involves the interaction of genes, accessory molecules, cellular structures, and the surround. Furthermore, Oyama has a valid point about those who would want to identify nature with genes and nurture with environment (also see Keller 2001). What I want to emphasize is that this argument does not show, and apparently does not purport to show,[3] that *genes* cannot be "properly contrasted" with *environment* in certain contexts.[4] The context here includes the part of the world scientists are modeling as well as their goals in modeling that part of the world. An experiment described in an introductory genetics text can be used to illustrate a context in which drawing a distinction between genes and environment (or between genetic and environmental variables) makes sense.

Suzuki, Griffiths, and Lewontin (1981) describe an experiment intended to illustrate the concept of norm of reaction. Clones from different plants of the same species *(Achilillea millefolium)* were planted, one clone from each plant, at an elevation of 30 meters above sea level, a second clone from each plant in the foothills at 1,400 meters, and a third clone from each plant in the mountains at 3,050 meters. The results for clones from seven different plants are shown in Figure 9.1. The authors use these results to demonstrate the point that one cannot use the relative heights of the different clones at one elevation to predict the relative heights at another elevation. One genotype results in greater height at the lowest elevation, another genotype at the medium elevation, and a third at the highest elevation.

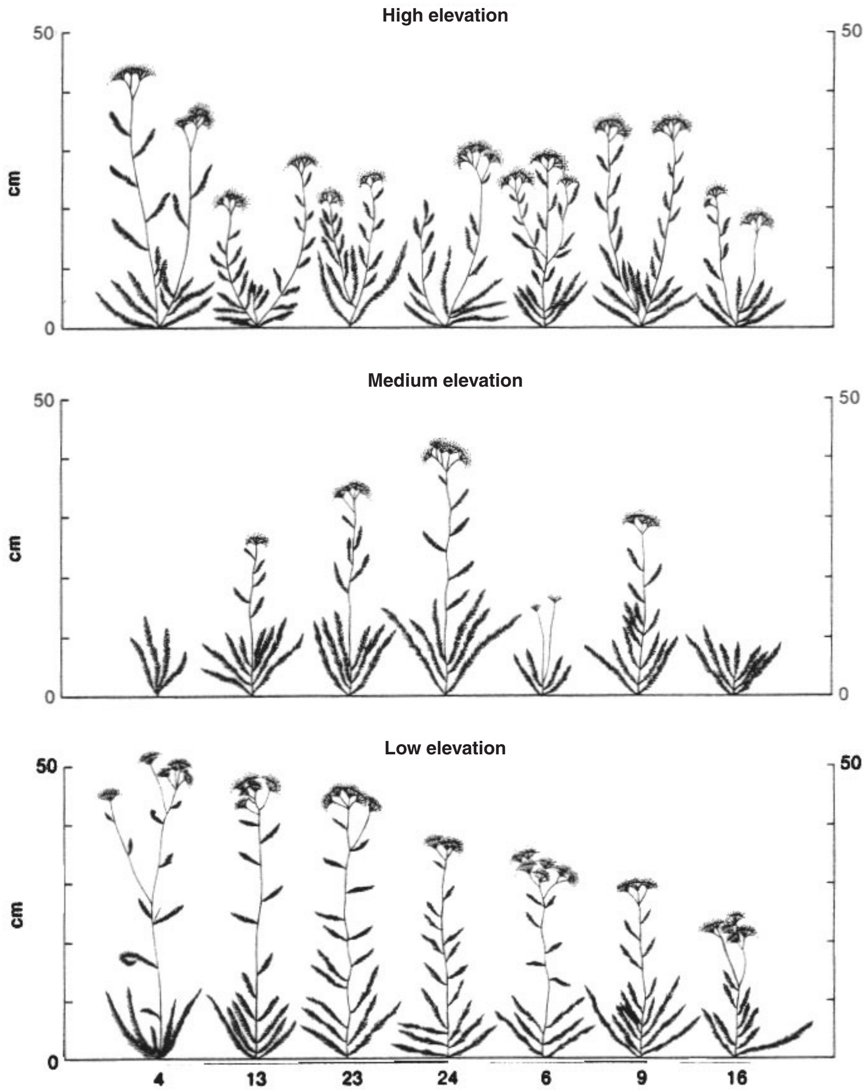

Figure 9.1. Parental plant (source of cuttings). Norms of reaction to elevation for seven different Achillea *plants (seven different genotypes). A cutting from each plant was grown at low, medium, and high elevations (Carnegie Institute of Washington). From* An Introduction to Genetic Analysis, *ed. David T. Suzuki, Anthony J. F. Griffiths, and Richard C. Lewontin, 2d ed. (New York: W. H. Freeman and Company, 1981), Figure 1-8, page 18; reprinted with permission from W. H. Freeman and Company.*

Distinguishing between genes and environment makes sense in this situation with respect to the goal of identifying what factors are causally relevant to the *difference* in plant height *in particular sets of plants*. Consider first the set of different clones grown at the low elevation. What differences caused the actual differences in height among this set? In this case, the differences are in the genes. Differences in the genotypes, not differences in the environment, apparently cause some plants to be taller than others.[5] Consider next the set of three clones from plant 4 that were planted at the three different elevations. What differences caused the differences in height among this set? The differences in this set weren't caused by differences in genes; environmental differences caused some plants to grow taller than others. What about the case involving all twenty-one plants? What differences caused the differences in height among this set? The answer is both differences in the genes and differences in the environments of these twenty-one plants. There is nothing "improper" about distinguishing between genes and environments in this context. The science of classical genetics didn't account for the appearance of characteristics in later generations, it accounted for the appearance of *differences* in characteristics. Genes and environment, or, to be more precise, genetic and environmental variables, can be properly contrasted when the aim is to investigate the causes of differences within a particular set of individuals in a particular range of environmental conditions. Oyama acknowledges this point.

Two caveats are in order. First, to say that the differences in height among the three plants in the first set were caused by differences in genes is not to say that the height of any particular plant is determined by the genes and not by the environment. Height is a characteristic constructed in a process of interaction among many causal elements including elements in the DNA, elements in the cytoplasm, and elements in the organism's environment. Given what biologists have learned about development, it simply wouldn't make sense to say that the reason an individual plant is a certain height is because of its genes and not because of its environment. An individual plant is the height it is because of genes and environment (as well as other factors internal to the organism but not part of its genetic makeup). However, *differences* in height between individuals are sometimes caused by differences in their genes, sometimes by differences elsewhere, and sometimes because of differences in both. Much has been written about attempts to measure the relative contributions of gene differences, environmental differences, and other differences to differences in particular traits. I will not go further into these issues here, but simply maintain that the distinction on which this apportionment rests is coherent, provided it is carried out with respect to a specified set of individuals in their actual environment(s).

The issue of whether or how such results of apportionment studies can be used to make inferences about future individuals raises a host of troubling issues. This brings me to the second caveat. Having evidence that genotypic differences caused the differences in height among the seven individuals at the low elevation does not imply that changes in environmental factors couldn't also affect height. In fact, if the plants were grown in a different environment, the heights of plants with different genotypes might be the same (as they are for plants of genotypes 13 and 6 in the high-elevation environment) or the relative heights might be reversed (as they are for plants of genotypes 4 and 9 at the low and medium elevations). The evidence that genes made the difference applies *only to a fixed range of genotypes* and *in a particular kind of environment*. This is obvious in the example, which of course was chosen to make this point clear.[6] This is much less obvious in the context of human genetics, where evidence that trait differences are caused by genetic differences is too often assumed to mean that our only means to control the trait is by changing the genes, not modifying the environment. This assumption is seriously mistaken. It may well be that an environmental difference not tested for, even a minor environmental difference, could make a tremendous difference in traits.

Establishing the fact that the difference that produced a difference in one group was genetic (or environmental) does not show that possible differences in environments (or genes) could not easily eliminate the difference or reverse it.

Although my causal analysis of the apportionment of causation between genes and environment may seem to contradict blatantly the views of philosophers who advance DST-type critiques, it is perfectly consistent with Richard Lewontin's criticisms of ANOVA. In fact, Lewontin is a coauthor of the textbook I used to substantiate my analysis (Suzuki, Griffiths, and Lewontin 1981). And provided the caveats are kept in mind, it is also in line with Oyama's account of this kind of reasoning (2000, 107–8). It may appear that Oyama's position is that dichotomizing between insides and outsides and apportioning causal responsibility between the two is conceptually flawed. But her real position seems to be that dichotomizing between insides and outsides and apportioning causal responsibility in *some* ways and contexts are flawed in the first place (e.g., nature and nurture), but are not flawed in the case of dichotomizing between genes and environment in the context I have described.[7]

Oyama's criticism of gene-centrism, then, does not rest on a universal argument that gene-centrism is conceptually muddled or that there is no proper way to draw the dichotomizing contrasts that geneticists frequently invoke. Her position seems consistent (perhaps she would say this is an understatement) with one of the central ideas I want to advance in this paper, that science is a pluralistic enterprise, that there is no single "first place." Scientists have a variety of interests and a distinction might be properly drawn with respect to some interests and not with respect to others. This is not just an issue of different conceptual frameworks applying to different parts of the world. Rather, for one and the same causal system, different interests will lead scientists to properly draw different distinctions and different causal conclusions that need not be integrated with one another.

Oyama introduces her account of the nature-nurture dichotomy as follows:

> An important question in developmental studies, if not *the* important question, pertains to the ontogenetic origin of organismic form and function, including the form of the mind. Traditionally, answers have either focused on a set of basic structures supposed to be transmitted in the genetic material, or else on the contingencies of individual experience. Thus, nature-nurture battles are ostensibly about the *allocation of causal responsibility* for development either to the genes or to the environment. The motivating concern, however, often appears to be with the notion of *limits,* rather than actual details of causation. What people finally seem to be speculating about is the limits on possible phenotypic variation and change. (2000, 99; emphasis in original)

When Oyama writes "*the* important question," she *seems* to be entertaining the possibility that studying development might come down to answering one fundamental kind of question. I doubt that there is a single important question, a single interest to which all other interests in development will be subsumed. When Oyama writes "the motivating concern," she seems to mean "the" with respect to the nature/nurture controversy. If so, then she makes a plausible point that drawing the nature/nurture divide might not be helpful for those with the interests of determining limits.[8] But I suspect philosophers reading such passages are prone to overlook the possibility that there are important interests in science that are well served by distinguishing between genes and environment and by putting genes in the forefront of scientific study.

Philosophers are particularly susceptible to this kind of oversight because analyses in philosophy often proceed by trying to identify *the* point, or *the* question, or *the* first place. I want to claim that with respect to areas of scientific inquiry such as developmental genetics there is no essential point, no single motivating question, and no first place. Philosophers of biology should abandon the quest for the essentials and conduct our philosophical study by identifying a possible plurality of *particular* interests and determining to what extent the concepts, theories, and methods might serve those interests. This pluralistic endeavor can help identify what interests a particular approach to science will serve and what interests it will not serve. An important question for the pluralist is, what epistemic interests might practicing scientists have for distinguishing genes (and DNA) and centering much of their research on these elements?

The Widespread Misinterpretation of Gene-Centrism

The popular myths about why genes are at the center of biologists' attention are closely associated with a misinterpretation of gene-centrism. The common interpretation is often advanced in newspapers and magazines. For example, in a newspaper article from the *New York Times* titled "Reading the Book of Life: A Historical Quest," genes are characterized as follows: "*Genes:* These sections of DNA instruct the cell to make proteins which perform all the body's essential tasks, like digestion, and determine physical features, like eye color" (June 27, 2000, A21). Such popular statements are as common as they are erroneous. Genes don't issue instructions to make proteins, and genes don't, as far as geneticists know, "determine" physical features like eye color. Differences in genes can cause differences in eye color in organisms with uniform genetic backgrounds living in similar environments. But differences in environments were also shown by geneticists to cause differences in eye color. Hence, it is at best misleading to state that genes "determine" eye color. One might as well say that environmental factors determine physical features like eye color and that transcription factors instruct cells to make proteins that perform all the body's essential tasks.

This misinterpretation can be found in the writings of scientific leaders as well: "Over the next ten years, as a consequence of the advance of our biological knowledge, we will arrive at new understandings. We will understand *deeply* how we are assembled, dictated by our genetic information" (Gilbert 1992, 96). A problem with using information metaphors like "direct" or "dictate" is that they make it difficult to be clear that the processes being investigated and modeled with the gene-centric (or DNA-centric) approach are highly interactive. Gilbert tries to address this issue by remarking, "genetic information does *not* dictate everything about us. We are not slaves of that information" (96). But it is difficult to draw precise distinctions with vague concepts of information. It is better to interpret gene-centrism in terms of the concepts of causality than concepts of information. It is more informative to say that gene differences can cause phenotypic differences in certain contexts than to say genes contain the information for phenotypic differences.[9] I will sketch an interpretation of gene-centrism that does not rest on the information metaphors.

Toward a Pluralist Interpretation of Gene-Centrism

My interpretation of gene-centrism, that is, my interpretation of the widespread practice of placing genes at the center of scientific attention, challenges two epistemological assumptions underlying the common misinterpretation of gene-centrism. According to the misinterpretation, genes are at the center of attention because biologists have established a theory that shows genes are the fundamental causal agents determining the developmental and metabolic processes within organisms. This misinterpretation implies that practicing the science involves answering questions about development and function by filling out this theory, and that this is accomplished by identifying the role the genetic determinants play in the particular biological process under investigation.

The first epistemological assumption is that coherent bodies of scientific knowledge are organized around central theories and inquiry involves filling out the theory. I believe gene-centrism is better understood as a *scientific approach* for investigation, manipulation, and modeling than as a theory or framework for comprehensive explanations. The second assumption is that the aim of science is to establish a single, integrated scientific theory or approach, at least for any given domain of phenomena. I will argue that gene-centrism should be interpreted from the tempered perspective of the *pluralist stance,* which holds open the possibility that causal factors emphasized in one adequate scientific theory or approach are not necessarily more "fundamental" than causal factors left in the background. Pluralism is open to the possibility that there could be a multiplicity of adequate theories or approaches, each of which emphasizes a different set of causal factors. Furthermore, according to this view, there is no reason to think that there must be a neutral perspective (or framework or language) in which all important causal factors can be integrated to yield a single, unified account of a complicated natural system.

Gene-centrism can be understood as a general *scientific approach* for investigating and modeling a broad range of biological processes. As such, it includes practical knowledge about various procedures, descriptive knowledge about the makeup and causal regularities of model organisms, and evaluative knowledge assessing the utility of procedures, materials, and ideas for further research. The functioning of this knowledge is structured not just by patterns of explanatory reasoning, but also by strategies for investigation (see Waters 2004a for a detailed illustration of what constitutes a scientific approach). Understanding gene-centrism as an approach centered on a set of open-ended strategies for investigating a broad range of biological phenomena, rather than as an explanatory enterprise centered on filling out the details of a central theory, makes it possible to entertain the idea that genes are at the center of attention because of their investigative utility, not because of their alleged explanatory power.

Oyama sometimes treats gene-centrism as a concrete approach for investigating biological phenomena, as she does in the following conciliatory passage:

> There is still a place for the "provisional single-mindedness" mentioned in chapter 3, the tactical decision to focus on one level for a particular purpose. It is important to recognize that such an approach temporarily delimits the context for the investigation; eventually the questions and findings must be recontextualized. There is, after all, no research that does not limit the scope of entities and variables studied. (2000, 4)

Oyama is open to the idea that it is appropriate for scientists to emphasize some causal elements at the cost of obscuring others. But this passage raises an important issue: why *must* questions and findings from an approach that emphasizes a subset of causal elements be recontextualized? The demand for recontextualization presumes that the goal is to integrate the knowledge gained from single-minded approaches into a comprehensive, grand theory that includes all causal elements. This is monism.

Oyama is a pluralist of sorts. She acknowledges the possibility that single-mindedness, that is, the adoption of approaches that obscure some features of a process in order to make other features salient, is a permanent, inherent feature of science. This leaves her open to the possibility that the best science can offer with respect to experimentation, even in the long run, is a plurality of investigative approaches, each of which makes it possible to investigate certain aspects of the phenomena, but none of which makes it possible to investigate all aspects of interest at once. I push pluralism further. Pluralists should be open to the possibility that there could be a plurality of ways to "recontextualize" the results of a single-minded approach *and* the possibility that different approaches will yield various explanations that cannot be integrated into a single, comprehensive explanatory framework. Perhaps there will never be a grand theory of development that integrates all the causal factors of interest into a single, useful perspective. The demand that the questions and findings of the gene-centric approach be recontextualized, presumably to DST, is a demand of monism.

Gene-centrism, if understood in terms of pluralism, does not presuppose grand claims about the causal primacy of genes (or DNA). Gene-centrism is consistent with the possibility that alternative approaches might yield viable explanatory models that emphasized causal elements obscured by the gene-centered approach. For example, with respect to human behavior, the gene-centered approach is but one of a number of viable approaches, some of which put genes into the background in order to make environmental elements more salient (see Longino, chapter 6 in this volume). In some other cases, viable alternatives have not been developed. For example, investigation into early development of model organisms such as *C. elegans* is dominated by gene-centered research. If it is bad policy to place so many investigative resources into the gene-centered (or DNA-centered) research, it is because more resources should be devoted to investigating aspects of the phenomena (or potential technologies) obscured by gene-centrism.[10]

Interpreting gene-centrism as one of a plurality of possible approaches leaves open the possibility that genes are the center of attention not so much because of their explanatory value, but because of their investigative utility. I will argue that this is indeed the case, that genes provide a unique entry point for investigating and modeling a broad range of biological processes.

What Are Genes? What Do Genes Do? Why Are They Central to Research?

Critics sometimes argue that there are no such things as genes, that "gene" doesn't pick out anything special, or that genes are whatever biologists want to call a gene (e.g., Burian 1986; Kitcher 1992; and Fogle 2000). Hence, I need to begin by clarifying how biologists conceive of genes before explaining why genes are central to biological research. Although "gene" is used differently by different biologists, and even by the same biologists in different situations, contemporary reasoning involving genes nevertheless often presumes one or another of three models. One model of the gene is classical and dates back to the Morgan school of genetics. Another is molecular and is integrated with contemporary understanding of the system of reactions that leads from DNA to the synthesis of RNA molecules and polypeptides. A third model, which I will not discuss here, involves conceiving molecular genes in terms of evolution.[11]

Centering on Classically Conceived Genes

The classical concept (or model) of the gene is typically applied by contemporary investigators when they lack a molecular specification of a gene (or in contexts where gross phenotypic differences are of practical importance). Classical genes are identified by their phenotypic effects in particular contexts, not by their molecular makeup. When contemporary biologists refer to a gene associated with a gross phenotypic characteristic (e.g., purple eye color), they typically are referring to the element of DNA, the difference in which is causing the phenotypic difference (e.g., in eye color) among individuals in a particular population under particular environmental conditions.[12] It is important to emphasize that this does not necessarily mean that the function of the gene is to produce eye color or that the gene is "directing" the production of eye color. The phrase "gene for" is misleading when used with respect to the classical model. Classically identified genes are usually named after one of the quirks caused when mutations interfere with their typical range of functioning, not after their functions. The more careful writings by classical geneticists (e.g., Sturtevant in his doctoral dissertation) make clear that producing purple eyes was not the function of the purple allele, and producing red eyes was not the function of its wild-type alternate.

Although gene concepts can be traced back further, what I'm calling the classical model is grounded in the classical genetics of the Morgan school. Asking why the Morgan school put genes at the center of their research is instructive. One might think genes were central because they helped classical geneticists explain a broad range of phenomena. But this answer is not satisfactory because even though genes figured into explanations of inheritance patterns, they did not figure into explanations of anything else of significance (e.g., development). Another answer is that although genes didn't account for much in concrete explanations, geneticists had a grand theory, and according to that theory genes played a determining role. The answer that genes had tremendous explanatory potential (rather than current explanatory power) is more plausible, but it is still not convincing. After all, classical geneticists had no idea what genes did. They were convinced genes were important (hence the plausibility), but they did not have anything amounting to a theory about what genes did, which the following speculation on the part of Morgan makes clear:

> Suppose, for instance, to take perhaps an extreme case, all the genes are instrumental in producing each organ of the body. This may only mean that they all produce chemical substances essential for the normal course of development. If now one gene is changed so that it produces some substance different from that which it produced before, the end-result may be affected, and if the change affects one organ predominantly it may appear that one gene alone has produced this effect. In a strictly causal sense this is true, but the effect is produced in conjunction with all the other genes. (Morgan 1926, 306)

What epistemic interests did classical genetics have for distinguishing genes and centering much of their research on these elements? It wasn't because they had a grand theory of development and wanted to fill out the details. The Morgan school didn't have a grand theory at all. At best they had the vague idea that organs are developed through the culmination of a long series of processes, and genes somehow act on the steps along the way. But this was just a vague idea. They did have a more precise explanatory theory (transmission theory), but the theory was narrow and could do little more than explain patterns of inheritance in carefully orchestrated experiments performed on a few model organisms. Then why was so much of their research centered on genes?

Genes not only seemed to be involved in a wide range of processes, but differences in genes caused differences in a wide range of processes. Even if genes turned out to be minor players, the tight causal connection between gene differences and differences in basic biological processes (under controlled conditions) offered a means for manipulating those processes. And, as every physiologist knows, one way to learn about how a biological system works is to manipulate the underlying elements and processes. Classical geneticists advanced a strategy, called *the genetic approach,* for using genes to study biological processes (Waters 2004a). The strategy involved identifying gene mutations that interfered with processes of interest and then using the techniques of genetics to manipulate those processes. This strategy was successfully applied to learn a great deal about chromosomal mechanics; it was far less successful at shedding light on other biological processes. Nevertheless, geneticists hoped that the manipulation of genes would shed light on other biological processes as well (e.g., gene action, mutation, speciation, etc.).

To summarize the historical case: the gene concept of classical genetics did not specify what genes are made of or what genes do. Rather, it related differences in genes to differences in phenotypes in certain contexts. The causal relation between gene differences and phenotypic differences in these contexts nevertheless enabled biologists to explain various inheritance patterns. More important, it provided them with an entry point for studying chromosomal mechanics, for mapping genes, and for investigating a wide range of biological processes. The reason classical genes were central to research was because identifying genes provided a basis for conducting research into these biological processes. It was the *investigative reach,* not the *explanatory scope* that motivated biologists to put genes at the center of so many investigations.

Although biologists now have a gene model that identifies what genes are made of and what they do (see the next section), the classical model of the gene still plays an important role in biological research. There are many, many investigative contexts in which genes are identified by the phenotypic effects brought about by differences in genes or regulatory regions. When newspaper headlines announce that biologists have discovered a gene for X, the underlying research often involves the identification of classical genes via phenotypic effects (and hence the terminology "gene for X" is often used in a very misleading way). The genetic approach, which can be employed without knowing the specific molecular identity of the gene, has remained an important strategy for studying biological processes. In fact, this approach was utilized in tandem with the physical approaches of biochemistry to investigate the mechanisms of a wide range of important biological processes, including the syntheses of DNA and RNA. For our purposes, it is important to keep in mind that this research is not dependent on the monistic presumption that genes are more fundamental than other causal elements participating in these mechanisms. Rather, it is based on the strategic idea that it is often easier to investigate mechanisms by manipulating genes than by manipulating the other elements.

Centering on Genes Conceived at the Molecular Level

Reconceiving genes at the molecular level provided the conceptual basis for developing powerful new methods of investigation. Biologists have managed to reverse the approach of classical genetics. Instead of starting with phenotypic differences and trying to work their way back to gene differences, biologists can now start with genes, and interfere with gene function directly to bring about phenotypic differences.[13] But the fundamental strategy of genetics remains the same: utilize the handle provided by genes to manipulate and investigate basic biological processes. It is largely in the context of implementing this strategy that molecular genes take center stage in the study of what goes on in model organisms such as yeast, worms, and fruit flies.

I will argue that the legitimacy of putting molecular genes at the center of biological research does not depend on the idea that genes or DNA have a greater causal role in the overall development of an organism than do other elements. In fact, it is not clear what it would mean to say that genes and DNA have a greater causal role. According to the popular interpretation, it means that genes contain the "information," provide the "blueprint," or "direct" the processes. But the information metaphors are vague, and it is not clear how they could be made more precise or translated into the terminology of causation. What can be made clear is that genes have a distinctive causal role and that this role provides biologists with an epistemological entry for investigating a wide range of basic processes within organisms, including developmental processes. I will support this claim by explaining how genes are conceived at the molecular level and clarifying the distinctive causal role that genes play. I will then show that although this role justifies privileging genes in only a relatively narrow range of explanatory contexts, it nevertheless warrants centering attention on genes in a wide variety of investigative situations. This challenges the monistic assumption that centering so much attention on genes would be justified only if genes are depicted as the fundamental causal agents in the single, comprehensive theory representing the development and functioning of organisms.

The molecular concept (or model) of genes includes a specification of what genes are made of and what genes do (i.e., what roles they play). Molecular genes are segments of DNA that determine linear sequences in molecules produced through a special chain of processes.[14] The chain of processes, which I call DNA expression, can produce a succession of products with linear structures: RNA, processed RNA, and polypeptides. The first molecule produced in DNA expression is RNA (unprocessed and unedited). RNA is made up of a string of nucleotides. A portion of DNA, also consisting of a linear sequence of nucleotides, serves as a template for the synthesis of RNA. This synthesis results in the production of an RNA molecule whose linear sequence of nucleotides corresponds to the linear sequence of nucleotides in the DNA segment.[15] Many different segments of DNA serve as templates, and hence a variety of RNA molecules is produced. The linear structure of any particular RNA molecule corresponds to the linear structure of the DNA segment that served as the template in the production of that particular RNA molecule. This DNA segment is the gene *for the RNA molecule*. Hence, for any particular unprocessed RNA molecule among a set of unprocessed RNA molecules with different linear structures, we can say that its linear structure was *causally determined* by the DNA segment, that is, the gene, because it is the difference in the nucleotide sequences in genes that caused the differences in linear structures among the unprocessed RNA molecules. This is true even though the gene was only one of many players (and a passive player at that) in the production of the RNA molecule.

The expression of DNA segments often involves additional processes. Many RNA molecules serve as templates in the synthesis of polypeptides.[16] The linear sequence of amino acids in a polypeptide is determined by the linear sequence of nucleotides in the RNA template in the same sense in which the linear sequence of nucleotides in the RNA template is determined by the linear sequence of nucleotides in genes. Hence, DNA segments can be said to determine the linear sequence of amino acids in the syntheses of different polypeptides. These segments are genes *for* *polypeptides.* It is worth stressing that these segments, the genes *for polypeptides,* play a distinctive determining role because the differences in the linear structures among different polypeptides synthesized in a cell or cell structure result *for the most part* from actual differences in the linear structures of genes (DNA segments), not from differences in the many other causal agents essential for the synthesis of the polypeptide. (I will discuss complications relevant to the *for the most part* qualifier later in this section, but I will note that this qualifier does not need to be added for the case of genes for RNA or polypeptides in prokaryotes or for the case of genes for unprocessed RNA in eukaryotes.)

Readers might wonder why I am willing to say that molecular genes "determine" linear structures when I deny that classical genes determine the phenotypic traits with which they are associated. I have argued that the classical genes are not truly genes *for,* that phrases such as "gene for purple eye color" are misleading. But I claim that molecular genes truly are genes *for.* Why? My reasoning is related to the reasoning about the *Achillea millefolium* experiment discussed earlier. In the case of the height of *Achillea millefolium* plants, I concluded that causal claims could be made about sets of plants, but not about individual plants. We can say genotypic differences cause phenotypic differences in height among plants in a set, but we cannot say that a genotype causes height in the case of an individual plant. Likewise, we can say molecular genes determine linear structure in the context where different molecular genes are playing causal roles in the syntheses of molecules with different linear structures, but we cannot say that genes cause the linear structure of a single molecule. If polypeptide synthesis had occurred only once in the history of the universe, there would be no reason to say that DNA played a distinctive role in determining the structure of the single polypeptide. One could say tRNA played just as distinctive a role in determining the structure. Likewise, if, contrary to the actual situation in living cells, every molecule synthesized via DNA expression had the same linear structure, then biologists could not reasonably claim that DNA played a distinctive role in determining that structure. It is only in the context where molecules of different linear structures are synthesized that singling out the causal role of particular elements makes any sense. And the reason for singling out genes is that they are difference makers.

DNA expression can be extremely complicated. In multicellular organisms, differences in the linear structures among polypeptides and processed RNA in different tissues and at different stages of development are often determined, not just by differences in genes, but also by differences in other causal elements (such as splicing agents). Hence, the distinctive determining role of genes is not always unique with respect to determining differences in processed RNAs and polypeptides synthesized in different tissues and at different developmental stages. Other entities also play the role of difference maker. But biologists still believe the determining role of genes is unique with respect to differences among linear structures of unprocessed RNA molecules. It is important to keep in mind that the complications mentioned here place limits on the explanatory power of genes and DNA, but they do not affect the conclusion I draw about why molecular genes are at the center of attention in so much biological research.

The complications of RNA processing do not render the molecular gene concept problematic. The gene concept has appeared hopelessly sloppy to many philosophers of biology (Rosenberg 1985; Burian 1986; Kitcher 1992) and at least a few biologists (Portin 1993; Fogle 2000). But appearances are deceptive. The gene concept is incredibly versatile, and once a context is fully specified, its application is remarkably precise (Waters 2000). Consider, for example, the complications of RNA splicing. In eukaryotes, portions of RNA molecules (called *introns*) are frequently removed, and the remaining sections *(exons)* are reconnected before the molecules are employed as templates in polypeptide synthesis. In such cases, precision can be achieved by distinguishing the gene *for the unprocessed RNA molecule* from the gene *for the polypeptide*. The gene for the unprocessed RNA molecule includes the parts of the DNA segment corresponding to introns as well as parts corresponding to exons. The gene for the polypeptide does not include the sections of DNA corresponding to the introns (the regions spliced out of the RNA molecule before polypeptide synthesis). Molecular biologists call genes of the latter kind "split genes" because the sequence of DNA nucleotides corresponding to the sequence of amino acids in the polypeptide is interrupted by nucleotides that correspond to introns. The gene concept is versatile because it can be applied to identify genes for linear sequences at different stages of DNA expression. The concept is precise because once the linear sequence, the molecule, and the relevant stage of DNA expression are specified, the molecular identity of the relevant gene is fixed. Alternative splicing does not render the concept imprecise because the concept can be used to distinguish a molecular gene for one spliced product from the molecular genes for alternatively spliced products.

Biologists' attention is often directed toward genes because genes are the difference makers. Most causal agents required for DNA expression are uniform in a particular cell or cell structure at a given time. A difference in one of the many other causal elements (e.g., a functional difference in one of the tRNA molecules) would result in changing all the polypeptides synthesized in the cell or cell structure, not just in changing a subset. Hence, differences among the polypeptides synthesized in a particular cell or cell structure *for the most part* result from differences in genes. Exceptions involve cases of differential RNA splicing and editing. When differential RNA splicing occurs within the same cell structure at the same time, then differences in the linear sequences among *these* polypeptides, that is, the polypeptides resulting from differential processing of the same kind of unprocessed RNA, could be said to be caused by differences in splicing factors rather than differences in DNA.[17] It would still technically be true that different split genes were involved. Nevertheless, it is perfectly reasonable to say about such cases that the splicing factors, as well as molecular genes, play a distinctive causal role in determining differences in the linear structure of the polypeptides in the cell. Hence, the distinctive role for genes is not always and necessarily unique.

Molecular genes are central to biological investigation for the same reason classically identified genes are central: they provide handles for manipulating biological processes. Molecular genes are actual and potential difference makers. By manipulating molecular genes, scientists can tinker with a wide variety of biological processes ranging from gene regulation to pathways for neutralizing toxins. Bernard investigated the functioning of the *milieu intérieur* by manipulating nerves of the sympathetic system. In similar fashion, biologists studying development in *C. elegans* or memory in *Aplysia* or mice aim to understand how these processes work by manipulating genes. Investigators need not assume that genes are the central causal agents in these processes. But experimenters can manipulate the agents that are central by manipulating genes. For example, systems biologists are investigating metabolic pathways that involve recurrent patterns of interaction. The immediate causal agents in these pathways are metabolites, not genes or DNA. An important strategy for learning about the patterns is to manipulate metabolites by interfering with underlying genes. The strategy does not depend on the idea that genes direct metabolism or any grand theory of systems biology. It is based on the strategic premise underlying the genetic approach: one can investigate biological processes by manipulating genes.

Gene-centered sciences are changing. Now regulatory molecules have joined genes at the center of attention. It is important to note that things also stay the same; regulatory molecules are at the center of attention in part for the same reason genes are. In addition to explaining differences, regulatory molecules also provide molecular biologists with a means to manipulate basic biological processes within organisms. And how do biologists manipulate regulatory molecules? Often, by manipulating genes. To learn about a network of genes, biologists alter genes (or the activity of genes) that determine linear sequences in regulatory molecules that affect the functioning of these molecules, which in turn affects the function of the genes these molecules regulate. It appears that as attention turns to higher-level processes, genes will still provide an important means for investigation. The staying power of gene-centrism stems not from the alleged power to explain all processes of interest directly in terms of genes and gene regulation, but in its incredible utility for investigating biological processes.

There are, of course, a number of molecular biologists who claim more. They believe that identifying the cascades of reactions emanating from gene transcription will explain the processes of development, metabolism, life cycles, environmental response, and so on. Some geneticists talk as if once they understand the causal role of every individual gene they will understand everything that happens within the organism. Some systems biologists talk as if once they understand all the networks of genetic regulation they will understand the fundamental basis for all life processes. Proponents of DST are right to criticize such explanatory claims. There is more to life than the regulated synthesis of linear molecules. But DST provides the wrong diagnosis and remedy. Scientists should not avoid gene-centered investigations on the grounds that gene-based explanations tend to obscure important causal elements. It is not the gene-centric approach but the interpretation of gene centrism, and the underlying epistemological assumptions, that should be discarded.

We should be skeptical about any claim, whether made by gene-centrists or by proponents of DST, to the effect that a single kind of account or perspective will provide a comprehensive framework for understanding everything of importance in a process as complicated as development. But this implies that we should also not dismiss an approach because it leaves out of its forefront elements that are causally important. Gene-centrism does obscure, its explanatory power is limited, but it nevertheless explains a lot about the synthesis of DNA, RNA, and polypeptides. According to this explanation, genes play the causal role of difference makers.[18] Because genes have this distinctive role, they can bring about differences in a very broad range of biological processes at different levels of organization. And although this distinctive causal role does not give genes the explanatory power often attributed to them, it does make genes important tools for scientific investigation. Pluralism, and the idea that science is organized around the activity of pursuing investigative strategies rather than filling out grand theories, provides an epistemological basis for interpreting gene-centrism and understanding why an approach with limited explanatory scope can nevertheless have such a dominant role in scientific practice.

Conclusion

Gene-centrism is not mistaken in the first place. Its viability does not rest on the naive view that genes are the fundamental causal determinants of everything that happens within organisms. When people talk as if gene-centered biology is premised on this view, they are misinterpreting the science. This misinterpretation is plausible because of mistaken epistemological views, widespread even among philosophers of science, about the nature and structure of scientific knowledge.

Philosophers of science have stressed one kind of epistemic interest to the practical exclusion of all other interests, and that is the interest that should be explained. My interpretation of gene-centrism is based on the idea that there is another epistemic interest, the interest to investigate, which often plays a more fundamental role in shaping a science. I have argued that it is this kind of interest that accounts for the dominance of gene-centrism throughout so many of the biological sciences. Placing genes at the center of attention has served a variety of investigative interests of many biologists starting with the Morgan school's investigation of meiosis and continuing through today's investigations of processes of development and metabolism, and phenomena of life cycles and environmental adaptability.

My interpretation of gene-centrism is thoroughly pluralistic. It does not assume that centering attention on genes is the only approach for investigating biological phenomena such as development or even that the findings of different investigative methods must be recontextualized into a grand genic-centric theory (or any theory for that matter). So much scientific attention is centered on genes because genes provide a useful entry for investigating a very broad variety of biological phenomena. The kind of grand theory associated with gene-centrism (e.g., that genes are the fundamental determinants of development) helps scientists garner resources to pursue their research (e.g., it creates excitement that helps attract funding), but the research can be defended while remaining totally agnostic about whether the gene-centric theory, or any single theory or theoretical perspective, could possibly explain the variety of developmental phenomena of potential interest.

Readers might complain that I've been too generous to Morgan and subsequent gene-centrists. After all, Morgan was not tolerant of diversity. The passage quoted from Morgan's *Theory of the Gene* (1926) indicates that although he admitted to knowing nothing about what individual genes did, he nevertheless assumed that it was the interaction of genes, not the interaction of genes and other stuff, that produced the phenotypic effects.

And the Morgan school is well known to have staunchly objected to alternative ideas about inheritance, such as cytoplasmic inheritance. But that doesn't mean *we,* as philosophers or historians, shouldn't adopt the pluralist stance when we try to understand the strengths and the limitations of their science. Likewise, I think we should be pluralists when we interpret contemporary gene-centered science (or for that matter, DST). There is room in contemporary science for multiple approaches (provided that a case can be made for each).[19] Some think the danger is not getting *the* right comprehensive picture of development. I think the danger is thinking that *the* right picture exists and that if a picture doesn't get it all in or get everything right, then it should be rejected. This ideal leads critics to overlook the possibility that centering research on genes (and DNA) is *a* viable approach for contemporary science, but more important, it leads advocates of gene-centrism to deny its limitations.

Notes

I thank participants in the "Workshop on Scientific Pluralism" organized by the Minnesota Center for Philosophy of Science in 2002 for offering valuable feedback on a draft of this paper. Susan Oyama kindly provided comments on an earlier draft and pointed out that she is not the kind of monist I erroneously claimed her to be. Evelyn Keller read a draft of this paper and offered detailed and helpful suggestions. Of course I did not follow all the advice I received, and any remaining shortcomings of the essay should be attributed to me. The workshop was supported by NSF grant 0125682, and the final stages of my work on this paper by NSF grant 0323591.

1. As biologists are learning more about the nature of regulation, the issue is quickly becoming one of "DNA-centrism" in many gene-centric sciences, including development. The arguments discussed in this chapter apply to both gene-centrism and DNA-centrism.

2. Examples of the kind of philosophical critiques I have in mind include Griffiths and Knight 1998, Moss 2003, and Robert 2004. Robert advances a grand theory of development different from DST, but his account of what's wrong with gene-centrism is in line with criticisms offered by DSTers.

3. Oyama pointed this out to me (personal communication) when I mistakenly claimed otherwise in a draft of this paper.

4. The main issue I'm dealing with here concerns how biologists distinguish genes (or DNA) from the *cellular* environment of genes. But some of the examples I discuss involve distinguishing genes plus nongenetic elements from elements in the *organism*'s environment. The term "environment" is ambiguous, and failure to notice this ambiguity has obscured the pluralistic nature of evolutionary biology (see Waters 1991, 2005). But I don't believe that ambiguity is the problem here. I avoid the nature/nurture terminology because, as Keller (2001) persuasively argues, there is much more to nature than genes and DNA.

5. One might object that, given the procedure of cloning in this example (cutting the original plant into pieces), the difference might be caused by random cytoplasmic factors,

not genetic factors. This is true. But it doesn't show that the distinction between genes and environment is mistaken in this case. It just shows that the conclusion is mistaken. The difference would be caused not by differences in the genes (or in the organisms' environment), but rather by differences in nongenetic factors within the organisms.

6. The authors also provide an example in which the genotype-phenotype relation is qualitatively robust among the range of environments examined (see Suzuki, Griffiths, and Lewontin 1981, 16, 19).

7. There seems to be a tension in Oyama's views on this point. At times it appears that Oyama is offering a categorical argument against dichotomizing in general, but her acknowledgment of the coherence of the kind of reasoning I have detailed here indicates that she does not take her argument to be decisive against dichotomizing in general.

8. Plausible, but by no means decisive. It might be that drawing a contrast between genetics and environment can help scientists shed light on possible limits. For instance, it might show that certain limits do not exist. Genetic analysis might lead to knowledge about underlying mechanisms (of which genes take part) that can be influenced by manipulating environmental factors.

9. Some biologists prefer information talk to gene talk (e.g., James Shapiro) and are trying to construct an account of development in terms of information rather than genes. As a pluralist, I would not rule out the future viability of such approaches, but the insightful arguments of Oyama and others in the DST movement have convinced me that those who wish to model development in terms of information will require a new account of information.

10. As I will explain in the next section, the investigative reach of the gene-centered (DNA-centered) approach is applied to investigate all sorts of biological processes in which genes are peripherally involved. Hence, although gene-centered explanatory models obscure important elements, the investigative strategies centered on genes obscure much less.

11. The target is moving. In 1994, I claimed there were two important concepts or models of the gene, in 2000 I said there were at least two important concepts. Now I believe there are at least three. Identifying genes shared by organisms of different species has become part of the regular practice of molecular biology. The underlying gene concept is molecular, but different from the one that I call the molecular concept of the gene. One might think of the classical concept as being "top down" (gene identity determined from phenotypic differences; causal chains traced backward toward DNA), the molecular concept as "bottom up" (gene identity determined by molecular level activity; causal relationships traced toward higher-level processes), and the third concept as "comparative" (gene identity determined by similarity of molecular structure and molecular-level activity across different species; causal relationships traced toward higher-level processes). The latter sense is evolutionary in that biologists believe that DNA sequences from different species are the same gene because they descended from a common ancestral gene.

12. When the molecular makeup of the difference and gene are discovered, and the situation is reconceived in terms of the molecular gene concept (which I discuss later), the conceptual situation can become more complicated. Sometimes, geneticists learn that the difference used to identify the classical gene is caused by a difference in a bit of DNA that regulates a molecular gene and that the regulatory region is located outside the molecular gene. So, for example, many genes identified classically by the Morgan workers were identified via mutations caused by p-element insertions upstream from the molecular gene. Pluralists need not be troubled by the facts that (1) the different conceptual models divide the chromosome differently, and (2) there is no useful way to combine the divisions into a single, integrated account.

13. This is sometimes achieved by changing the gene, sometimes by interfering with the regulation of the gene, sometimes by introducing elements that eliminate gene transcripts.

14. I believe my analysis gives a clear account of the reasoning of molecular biologists, but perhaps I should be credited with adding rigor that is not implicit in their thinking (or at least their use of language). Paul Griffiths and Karola Stotz are conducting poll-based research into this kind of question in their Representing Genes project. My fundamental goal is to provide an account of how gene-centered sciences ought to be understood in order to appreciate their epistemic virtues and limitations (with respect to epistemic ideals such as truth, explanatory power, investigative utility, etc.). Whether biologists understand their science in this way is an open question that might be pursued through historical analysis, ethnographic investigation, or poll-based research. Depending on the results of such research, it might be necessary to say genes *ought to be* conceived according to my analysis. I explore these meta-issues in Waters 2004b.

15. A careful reading of advanced textbooks of molecular biology shows that the process is understood in causal terms, without essential appeal to notions of information.

16. Again, I will not go into the details of the synthesis, but examination of advanced texts reveals that the process of polypeptide synthesis is also understood in causal terms, without essential appeal to notions of information. In this case, the linear sequences of nucleotides serve as templates via the intervention of tRNA molecules.

17. As Evelyn Keller and Karola Stotz emphasize, recent research is shedding additional light on the significance of RNA processing and editing. It has already become evident that differential splicing of RNA is an important factor causing differences in linear sequences of polypeptides across different tissues and developmental stages. Such processes can be important for differences within one and the same cell in the same stage of development.

18. One might object to my claim that genes are the primary difference makers. If we trace the reactions back a few steps, we will see there are lots of difference makers. For example, there are hosts of regulatory molecules that determine which segments of DNA are active and which are not. So these molecules also have a distinctive causal role. But it is a different causal role. And besides, these agents are at the center of attention in gene-centered biology as well.

19. An implication of this essay is that DST cannot be justified on the basis of a simple disjunctive syllogism (either gene-centrism or DST, or not gene-centrism therefore DST). It needs to be advanced on its merits as an approach for practicing science.

References

Burian, R. M. 1986. “On Conceptual Change in Biology: The Case of the Gene.” In *Evolution at a Crossroads,* ed. D. J. Depew and B. H. Weber, 21–42. Cambridge, Mass.: MIT Press.

Fogle, T. 2000. “The Dissolution of Protein Coding Genes in Molecular Biology.” In *The Concept of the Gene in Development and Evolution, Historical and Epistemological Perspectives,* ed. S. Oyama, R. Gray, and P. Griffiths, 3–25. Cambridge: Cambridge University Press.

Gilbert, W. 1992. “A Vision of the Grail.” In *The Code of Codes, Scientific and Social Issues in the Human Genome Project,* ed. D. J. Kevles and L. Hood, 83–97. Cambridge, Mass.: Harvard University Press.

Griffiths, P. E., and R. D. Knight. 1998. “What Is the Developmentalist Challenge?” *Philosophy of Science* 65: 253–58.

Keller, E. F. 2001. “Beyond the Gene but beneath the Skin.” In *Cycles of Contingency,* ed. Susan Oyama, P. E. Griffiths, and R. D. Gray, 299–312. Cambridge, Mass.: Bradford/MIT Press.

Kitcher, P. 1992. “Gene: Current Usages.” In *Keywords in Evolutionary Biology,* ed. E. F. Keller and E. Lloyd, 128–31. Cambridge, Mass.: Harvard University Press.

Morgan, T. H. 1926. *The Theory of the Gene.* New Haven, Conn.: Yale University Press.

Moss, L. 2003. *What Genes Can’t Do.* Cambridge, Mass.: MIT Press.

Oyama, S. 2000. *Evolution’s Eye.* Durham, N.C.: Duke University Press.

Oyama, S., P. E. Griffiths, and R. D. Gray. 2001. “Introduction: What Is Developmental Systems Theory?” In *Cycles of Contingency,* ed. Oyama, Griffiths, and Gray, 1–12. Cambridge, Mass.: Bradford/MIT Press.

Portin, P. 1993. “The Concept of the Gene: Short History and Present Status.” *Quarterly Review of Biology* 68: 173–223.

Robert, J. S. 2004. *Embryology, Epigenesis, and Evolution: Taking Development Seriously.* Cambridge: Cambridge University Press.

Rosenberg, A. 1985. *The Structure of Biological Science.* Cambridge: Cambridge University Press.

Sober, E. 1988. “What Is Evolutionary Altruism?” *Canadian Journal of Philosophy* 14: 75–99.

Suzuki, D. T., A. J. F. Griffiths, and R. C. Lewontin. 1981. *An Introduction to Genetic Analysis.* New York: W. H. Freeman.

Waters, C. K. 1991. “Tempered Realism about the Force of Selection.” *Philosophy of Science* 58: 553–73.

———. 1994. “Genes Made Molecular.” *Philosophy of Science* 61: 163–85.

———. 2000. “Molecules Made Biological.” *Revue Internationale de Philosophie* 4, no. 214: 539–64.

———. 2004a. “What Was Classical Genetics?” *Studies in History and Philosophy of Science* 35: 783–809.

———. 2004b. “What Concept Analysis Should Be (and Why Competing Philosophical Analyses of Gene Concepts Cannot Be Tested by Polling Scientists).” *Studies in History and Philosophy of the Life Sciences* 26: 29–58.

———. 2005. “Why Genic and Multilevel Selection Theories Are Here to Stay.” *Philosophy of Science* 72: 311–33.

10
Disciplinary Pluralism for Science Studies

Borrowed Knowledge and Scientific Pluralism

This essay is part of a larger project which examines the phenomenon of "borrowed knowledge." It is within the context of this larger project that I seek to outline and defend a pluralistic approach within science studies. What do I mean by borrowed knowledge? Consider, for example, that the intense scientific interest in nonlinear dynamics (popularly known as chaos theory) has inspired a number of attempts to draw broad implications for areas usually considered far removed from the physical sciences. In addition to some relatively uncontroversial applications of chaos theory to economics, we find endeavors to unearth strange attractors in postmodern literature, and to resolve longstanding legal or even theological quandaries by appeal to nonlinear dynamics. Some of these attempts merit serious attention, some are harmless speculation, and some are simply misleading.

Two brief examples should serve to illustrate what is meant by borrowed knowledge. In the field of legal theory, Glenn Harlan Reynolds has called on chaos theory to steer between conservative legal scholars and the radical proponents of critical legal studies. Where the former seek to understand the law in terms of exact predictions, the latter view the legal system as an indeterminate arena of random power plays. Reynolds suggests that by examining the patterns of decisions made by the Supreme Court in terms of chaos theory, we can find a way beyond both rigidity and nihilism: "Despite this unpredictability, the actions of the Supreme Court are not random. Just as there is structure within chaos, so there is pattern of sorts within the actions of the Court—pattern that itself reflects recursion and sensitivity to initial conditions, and that exists on both large and small scales" (1991, 114). Elsewhere, I detail how in this instance borrowed knowledge serves as a valuable antidote to previous importations from the natural sciences (Kellert 2001).

A less happy example comes from the field of literary theory, where Alexander Argyros has sought to use chaos theory to uphold the aesthetic superiority of traditional narrative, among other things. He proposes that "the most robust and complex things in the universe, including human beings and their theories, works of art, and social structures, are best understood . . . as chaotic systems" (1991, 6). Because he speculates that chaotic systems lead to increased complexity, and complexity yields progress, Argyros feels confident that chaos theory can ground evaluative claims about both social and textual organization. Yet the borrowing here, besides making inaccurate claims about a preference for complexity within chaotic systems, relies on poorly developed analogies between physical systems and texts (see Kellert 1996 for a fuller discussion).

The phenomenon of people taking chaos theory and using it in other disciplines raises a number of questions, including: Why do people borrow knowledge, and how do they go about it? What do they hope to accomplish by borrowing knowledge, and what do they actually accomplish? When does it work well, and when does it work badly? In addressing these questions, I contend that approaches from a multiplicity of disciplines provide useful and legitimate means for understanding and evaluating the phenomenon of borrowing. By turns, we may have to ask questions about the social structures of academic work, about the way language works, about the nature of questions about values. Rather than going about addressing these questions, my task here is to articulate and defend a methodological approach toward answering such questions, an approach characterized by disciplinary pluralism.

Confusion may arise because my object of study, borrowing, is itself a phenomenon that takes place within other peoples' research process. So when we ask, "How does borrowing take place?" and "When does it work well?" we are undertaking a methodological inquiry. But I am not asking those questions here; they are the primary subject matter of my larger project. In other words, here we will be asking not about borrowing but rather about how to go about answering questions about borrowing.

Disciplinary pluralism invites the use of techniques from multiple disciplines to understand the subject matter of an investigation. These techniques might be used by one investigator, by one team of researchers, or by a number of individuals or teams. The "division of cognitive labor" often provides a valuable research strategy, so particular investigators or teams might each pursue their own disciplinary approach. These disciplines may combine, cooperate, compete, or remain aloof from one another—a genuine pluralism does not require any particular form of interaction between disciplinary approaches. In fact, a fully pluralistic approach will admit the possibility that for some questions the number of disciplinary approaches required is equal to exactly one. To borrow a technical term, we might label a nonimperialist monodisciplinary approach, which remained open to the possibility of other disciplines being relevant, as a degenerate case of disciplinary pluralism (on this point, see also Nozick 1981, 644).

The argument for disciplinary pluralism in science studies draws much inspiration from pluralistic positions within the sciences themselves, including the positions outlined in this volume. One view of these scientific pluralisms would characterize them as rejecting the following two assumptions:

> Monistic Assumption 1: All theories (or all models, or all causal factors, or all experimental approaches, or all interpretive frameworks for mathematical formalisms) should be able to be joined into one comprehensive account (or approach, or framework).
>
> Monistic Assumption 2: Any theory (or model, etc.) that is incomplete or partial is deficient, and needs to be made part of a more comprehensive account (or approach, or etc.).

Scientific pluralism, according to this formulation, suggests that many of our best knowledge-making practices have shown us that we ought not to accept these two monistic assumptions. Rather, we should proceed with our inquiry in such a way as to leave open the question as to whether comprehensiveness is achievable or desirable. In slogan form, pluralism claims, "Comprehensiveness is not a cognitive ideal, nor a cognitive virtue." But pluralists need not take the extra step toward claiming that comprehensiveness is a cognitive vice. Hankering after unification may sometimes lead some researchers in a helpful or progressive direction, and sometimes it may not. But such a methodological monism certainly does not qualify as a motivational necessity for practicing science, and it should be left behind in our efforts to understand the practice of science as well.

What Is a Discipline?

In this discussion, I will be using a relatively broad sense of what counts as a discipline, following William Bechtel's overview of how disciplines can be identified and distinguished: by their objects of study (domains, phenomena, model systems), by their cognitive features and tools (problems, theories, techniques), or by their social structure (turf, professional organizations, genealogies of training, journals) (1987, 297). Using this broad framework, we can see that other characterizations of disciplines focus on one or more of these three criteria. Darden and Maull (1977, 44) and Joseph Kockelmans (1979, 127) consider objects of study and cognitive tools but exclude considerations of social structure. Moti Nissani, on the other hand, defines a discipline almost entirely in terms of social structure as "any comparatively self-contained and isolated domain of human experience which possesses its own community of experts" (1997, 203). While Stephen Toulmin includes all three features of a discipline, he reserves the term for those systematic endeavors with a clear agreement on central problems and ways to solve them, characterizing atomic physics and law as disciplines but excluding philosophy (1972, 145). I will be talking about disciplines in a broader sense in which they are knowledge-producing enterprises with some shared problems, some overlapping cognitive tools, and some shared social structure. But I consider philosophy and sociology disciplines, and there is certainly no set of shared techniques that all philosophers, for example, share. Consider a Heideggerian philosopher such as Edward Casey (1997) and an analytic philosopher of science such as John Earman (1989) discussing space, for example. There is very little that they have in common in terms of style of argumentation or overlapping references.

In some sense, then, pluralism needs no defense and requires no one to advocate for it. The current state of knowledge production in most disciplines is in fact rife with a diversity of methods and marked by opportunism. In fields such as science studies, as in area studies or women's studies, multidisciplinarity is already firmly emplaced. But while disciplines may be fluid and multiple in practice, this is far from their professed ideology. All too often we operate with a picture of unbridgeable differences in objects of study or technique, and we conceive of interdisciplinary interaction as the cooperation of these essentially isolated endeavors. This defense of disciplinary pluralism will therefore need to focus on the question of the relationship between disciplinary methods of inquiry. Disciplinary pluralism along the dimension of social structure will not be the focus of this discussion—I will not seek to defend the multiplicity of academic departments, journals, or degree-granting programs, for instance. Neither will I seek to defend a pluralism along the dimension of domains of study—the ontological question of whether different fields have irreducibly different objects of inquiry will have to be left aside. Of course, we may have metaphysical inclinations that guide us to consider some disciplines inappropriate for some subject matters. Many will consider theological inquiry to be unnecessary in botany, while others will confidently assert that number theory calls for exactly one disciplinary approach. But the question of which discipline or disciplines are appropriate ought not be settled by prior metaphysical intuitions about the intrinsic nature of the subject matter. The question of which disciplines to use is an empirical, contingent matter, where our ontological convictions play the role of starting points rather than definitive conclusions.

Disciplinary Pluralism as a Variety of Interdisciplinarity

In articulating and defending pluralism with regard to disciplinary approaches, it will be helpful to clarify the distinctions between interdisciplinary efforts narrowly conceived and the broad range of activities that are called interdisciplinary. A variety of terminology finds use here, but most of it aligns in a way that helps us make a few useful distinctions. First, there are interdisciplinary efforts in the narrow sense of making a new discipline between two existing ones, perhaps leaving the two original disciplines unchanged. In the sciences, Darden and Maull call such new disciplines as biophysiology "interfield theories" (1977). Multidisciplinary work, such as that found in area studies, involves the juxtaposition of two or more disciplines; the different perspectives of the different disciplines are cumulative but not highly interactive, so there is little mutual change, combination, or integration. The term "cross-disciplinary" provides the most apt term for the borrowing of knowledge from one field in order to assist the endeavors of another discipline. Finally, transdisciplinary approaches are more comprehensive, looking for unity in an overarching synthesis in the grand and sweeping manner of Marxism, systems theory, sociobiology, and so on (see Klein 1990 and Kockelmans 1979).

Klein highlights a contrast between what she calls the instrumental and the synoptic views of interdisciplinarity. The former view considers interdisciplinary efforts as a practical matter for solving problems and is likely to be satisfied with multidisciplinary approaches, while the latter is motivated by a philosophical commitment to coherence and unification and longs for transdisciplinary synthesis (1990, 42). Klein goes so far as to say that "all interdisciplinary activities are rooted in the ideas of unity and synthesis, evoking a common epistemology of convergence" (11). But pluralism about disciplinary approaches rejects any presumption that the synoptic view is exclusively correct. Pluralism does not reject transdisciplinary efforts out of hand, but rather refuses to assume that unity is always possible or synthesis always useful. Pluralism about disciplinary approaches does not presume that all disciplines are commensurable and ultimately unifiable, but neither does it presume that each discipline is necessarily isolated and incommensurable with all others. Genuine pluralism would hold that for any two disciplines, it is a contingent matter and an open question whether or not they can or should be united or crossbred.

Such questions must be answered on a case-by-case basis, and the answers depend on the usefulness of interdisciplinary efforts for addressing the problem at hand. For some problems, a multidisciplinary approach of cooperation but not integration is the best. For other problems, the creation of a new discipline that combines two or more others may give rise to a useful approach. And in still other situations, what is most helpful is the unification of a number of disciplines that are then subsumed in a transdisciplinary synthesis. But pluralists hold that there is no good reason to presume that isolation, cooperation, crossbreeding, or synthesis is good in all cases.

A preliminary case for disciplinary pluralism in the examination of borrowed knowledge could begin by sketching some general advantages to an interdisciplinary approach. As Bechtel points out, sometimes there is a fruitful interchange between practitioners of disciplines who realize that they have interpreted the same phenomenon in radically different ways (1987, 299). In fact, Moti Nissani has cataloged a long list of advantages of interdisciplinarity, which includes bringing about creative breakthroughs, correcting disciplinary oversights and blind spots, addressing topics that fall through the cracks because they do not fit into any established discipline, and addressing complex, practical, real world problems that require us to move beyond the tunnel vision of experts (1997, 204–8). A book-length survey of theories of interdisciplinarity by Julie Thompson Klein concludes: "[C]utting across all these theories is one recurring idea. Interdisciplinarity is a means of solving problems and answering questions that cannot be satisfactorily addressed using single methods or approaches" (1990, 196).

Such general praise for interdisciplinary approaches is all well and good, but I would like to focus on the need for interdisciplinarity in investigating scientific knowledge, that is, within the field of science studies. And here we find a welter of recent declarations that in order to understand scientific knowledge we must use more than one disciplinary approach. The argument often proceeds from the point that science is an activity of great complexity; as Ronald Giere puts it, "[Science is] at least as complex as the reality it investigates. This great complexity implies, I think, that it is impossible to obtain an adequate overall picture of science from any one disciplinary perspective. Different perspectives highlight different aspects while ignoring others" (1999, 28). Two of these different aspects are the sociological feature of competing interests and the epistemological feature of experimental data. Giere conceives of these as complementary factors in accounts of scientific theory choice: "In some cases experimental data may strongly influence theory choice; in other cases political commitments or professional interests might be dominant. Most cases are mixed"

(61). Giere advocates a multidisciplinary approach, saying that different perspectives in science studies need to be integrated and their practitioners should collaborate (63). But there are different ways of integrating perspectives, and we need to be clear that what is called for is not always the creation of new, hybrid disciplines or grand unified theories.

Alison Wylie has spoken of a growing consensus that "each of the existing science studies disciplines is inherently limited, taken on its own. Indeed, given the complex and multidimensional nature of scientific enterprises—a feature of science that is inescapable when you attend to its details—it is simply implausible that the sciences could be effectively understood in strictly philosophical, or sociological, or historical terms" (1994, 394). And the point finds support from the chemist Henry Bauer (1990, 113), the philosopher David Stump (1992, 458–59), and the sociologist Andrew Pickering (cited in Wylie 1994). Despite this impressive consensus, one may still ask, why does complexity require interdisciplinarity? After all, chaos theory itself has shown us that complicated behavior may be the result of a simple underlying dynamic. But the complexity of science lies not only in its behavior—its "output" of successive theories, for instance—but also in the multiplicity of factors at work in its dynamic processes and in the multiplicity of questions that can be asked about them. I will return to this question later, in the course of articulating an image of multidisciplinary science studies that goes beyond the conception of multiple perspectives.

If we should use more than one disciplinary approach in investigating scientific knowledge, how much more important will it be if we broaden our scope to include a consideration of how scientific knowledge is used in other fields! For example, when legal scholar Glenn Harlan Reynolds (1991) claims that the pattern of decisions by the U.S. Supreme Court should be thought of as a strange attractor, we may ask a number of different questions. We may want to know why he has chosen to borrow concepts from chaos theory, or what effect his argument has on his audience, or what implications this argument would have for the practice of law. We may even want to know whether and in what respect the pattern of Court decisions is in fact like a strange attractor. To answer these questions, we will need the resources of such disciplines or subdisciplines as the rhetoric of inquiry, the sociology of knowledge, the philosophy of law, and the psychology of metaphor. These different approaches interact and inform one another, but cannot be integrated into one general scheme for the examination of borrowing. Part of the reason they are nonintegrable is that different approaches to the phenomenon of borrowing constitute different objects of study: legal scholarship is a practice that generates knowledge, but it is also a discourse that generates money, power, prestige, and policy recommendations. I will return to the issue of the incompatibility of disciplinary approaches when I develop a new image for interdisciplinarity.

Despite the arguments I have given here, some people dissent from an enthusiastic endorsement of interdisciplinarity, and I will address two concerns that arise from the use of multiple approaches. First, interdisciplinarity can become indiscriminate or exclusive when it becomes an end in itself. This danger is raised by Bechtel (1987, 298) as well as Bengt Hansson (1999, 340) and Bauer (1990, 113). It is important to recognize that fruitful connections between disciplines depend on the existence of methods of inquiry developed within particular disciplinary contexts. So interdisciplinarity can be problematic if it is all that we do and is seen as more important than the pursuit of any particular discipline. But this is not a position I advocate. Pluralism in disciplinary approaches means to recognize and affirm the use of multiple approaches when they are appropriate. But sometimes a single approach is called for, and more is not necessarily better. As David Stump counsels, as with science, we should test our methods and see which ones work. Some will probably be unproductive for some purposes. Different disciplines have different perspectives to offer and may all be useful at different times for different purposes (1992, 459).

A second concern comes from an opposite direction, which is the danger of dilettantism, a point raised by Nissani (1997, 212) and Bauer (1990, 113), among others. After all, there is a reason why serious fields of inquiry are called disciplines—their practice requires time, dedication, and, indeed, discipline. The demands of rigorous and specialized scholarship make it exceedingly difficult to engage responsibly with more than one discipline. Dabblers may easily be misled by superficial resemblances when they are not acquainted with the technical details wherein so much of the real effort lies. In response to this concern, I would say, examine particular instances of interdisciplinarity work. Are they superficial dabbling? Look and see. Perhaps interdisciplinary efforts have a greater percentage of superficiality, but I have found no systematic or even anecdotal evidence of this. Surely narrow monodisciplinary efforts are not free from the risk of superficiality. There is nothing about interdisciplinarity that necessitates a lack of seriousness. The difficulty of becoming conversant with more than one field can be outweighed by the new insights that can be gained.

If we step back for a moment, we can see a parallel to a general point of this project. Interdisciplinarity is neither always good nor always bad. It is neither automatically an end in itself, worth pursuing in all cases to the exclusion of disciplinary effort, nor is it always simple superficial dilettantism. It can be useful or useless, as the case may be. This is true at the meta level of how to go about examining borrowing, but it is also true at the level of borrowing itself. "Dilettante" is merely a term of abuse for bad scholarship that deviates from accepted turf boundaries. Bad scholarhip is the problem; interdisciplinarity itself carries no special risk. Or rather, if it does carry a special risk, this is just because the existing structures for rigorous criticism of interdisciplinary work are presently weaker. And it is precisely this weakness that my project seeks to remedy, by creating some tools for the rigorous critique of cross-disciplinary work.

Disciplinary Pluralism as Cross-Training

Debates about interdisciplinarity often use metaphorical imagery, so it will serve us well to examine some of these metaphors and seek one that helps to highlight a pluralistic approach. Metaphors matter for more than just explicating or illustrating a point because they highlight or conceal particular features that are important. To clarify how disciplinary pluralism contrasts with other accounts of interdisciplinarity, I will utilize the metaphor of cross-training in sports. The idea behind cross-training is that one can improve one's performance in a chosen sport by practicing other activities. In the eighties it was not unheard of for wide receivers to take ballet lessons to learn grace and agility. Kareem Abdul-Jabbar is said to have taken up karate to improve his performance at basketball, and ice skater Elvis Stojko is reported to have studied kung fu to improve his speed and balance. Hockey players may practice gymnastics to improve agility, and swimmers may study yoga to improve concentration. Different sports help us develop different skills, and an exclusive focus on one sport may even become counterproductive. There may be a number of different sports that would contribute to one's goals, although one does not have to practice every sport, and cross-training does not necessarily work for everyone.

The contrast between the image of cross-training and other metaphors for interdisciplinarity will help to articulate the nature of disciplinary pluralism. The other metaphorical images to be considered will be nations, tiles, and languages. Julie Thompson Klein, in her comprehensive survey of metaphors for disciplinary relationships, points out that the dominant image is drawn from geopolitics, with disciplines conceived of as nations. Hence, we find much talk of boundaries, borders, and frontiers; turf, territory, and no-man's-land; expeditions, trade, and migration; balkanization, protectionism, and autonomy; and nationalism, tribalism, ethnocentrism, and imperialism (1990, 77; see also Gieryn 1999). She also mentions the images of fish scales, honeycombs, and perspective slices of a solid, and these are what I will refer to as "tiles"—disciplines conceived of as separate pieces of knowledge that map onto contiguous parts of the world (1990, 81–82). We can also find other images such as rivalries between siblings, systems with feedback, markets with competition and entrepreneurs, and organisms with hybrid vigor, cross-fertilization, and symbiosis, but I will have less to say about these (1990, 80). Instead, I will deal with the metaphor of disciplines as languages, each describing the world differently.

Looking first at disciplines as nations, Donald T. Campbell provides an early example of this image when he identifies the "ethnocentrism of disciplines" as "the symptoms of tribalism or nationalism or ingroup partisanship in the internal and external relations of university departments, national scientific organizations, and academic disciplines" (1969, 328). A pluralist can make sense of the nation metaphor by saying that certain questions, like adjustments in income tax rates, are matters of internal politics to be dealt with within one sovereign nation. Such a situation would correspond to the fact that monodisciplinary approaches are sometimes appropriate. Other matters require international cooperation, as when there is a lake that straddles a border. Still other matters may require the forging of alliances or the creation of transnational bodies such as the UN, and there are even some people who call for dissolving all nations into bodies such as the EU or a world government. These situations correspond to varying forms of cross-, multi-, and transdisciplinary endeavors, each appropriate for different types of inquiry.[1] However, the nation metaphor paints a very static picture, seeing disciplines as organized in space rather than as ongoing human activities. Because of our contemporary geopolitical conception of the nation-state, the image of disciplines as nations highlights the characterization of disciplines in terms of competing social structures (turf) and, to some extent, their objects of study (domains). But this metaphor diverts attention away from the aspect of disciplines as collections of cognitive tools for active investigation, while the image of cross-training avoids this limitation.

We can find the image of disciplines as tiles in Ronald Giere's argument for multiple disciplinary perspectives in science studies, where he says, "the only adequate overall picture will be collages of pictures from various perspectives" (1999, 28). Speaking of disciplinary approaches as "perspectives" or "views" conjures the image of a preexisting object with several aspects. Different disciplines may examine different aspects, or they may examine one and the same aspect from different directions. Sociology will look at funding patterns and networks of training, for instance, but rhetoric will not. Yet this defense of multidisciplinarity works best for situations where a number of independent and clearly demarcated causal factors are at work, with each perspective identifying one factor. Certainly the "tiles" may overlap in some areas, or have porous boundaries, but the conceptions of partial perspectives or contributing factors promise an easy compatibility. In more tangled situations, multiple perspectives may be necessary, but the views provided may not all fit together or lie flat, as discussed by Longino (chapter 6 in this volume). It is not clear that we can parcel things out tidily, for example, attributing 60 percent of an episode of theory change to the evidence and 40 percent to social interests. Again, the tiling image of multiple visual perspectives limits us to a static picture, while the metaphor of cross-training encourages us to think of active endeavors that may not be able to be practiced simultaneously. You cannot do yoga while swimming (at least, not for long), and it is similarly difficult to maintain a sociologist's methodological relativism while engaging in epistemological evaluation. Rhetoric and epistemology both look at arguments, but engage with them in very different ways. Different explanatory factors fail to line up neatly here, in a manner akin to the account given by Fehr (chapter 8 in this volume). Also note that the metaphor of cross-training is not especially well suited for those who might view the goal of science studies as the construction of a comprehensive representation of science. While perfectly happy with the way this particular image moves us away from a monistic account of science, genuine pluralists will nonetheless welcome a plurality of metaphors and a plurality of conceptions of the goal of science studies.

Campbell's image of comprehensive knowledge involves a fish-scale pattern: "a continuous texture of narrow specialties which overlap with other narrow specialties. Due to the ethnocentrism of disciplines, what we get instead is a redundant piling up of highly similar specialties, leaving interdisciplinary gaps" (1969, 328). This description evokes an ideal image of slightly overlapping shapes, tiling the terrain much more effectively than our current situation where clusters of highly overlapping fields leave broad spaces between clusters. But such an image suggests that each piece of the world has one discipline (or at most a small number of disciplines) appropriate for studying it. Here we find a very weak form of pluralism, which says that for any particular phenomenon there is one discipline that works best for it, or else some combination of disciplines each of which precisely accounts for some distinct parcel of the phenomenon. A thoroughgoing disciplinary pluralism goes deeper and suggests that sometimes the perspectives don't fit nicely together on the same plane: they overlap or conflict or cannot both be held at the same time, and yet you need both of them. This disciplinary pluralism shares much in common with the philosophical pluralism advocated by Robert Nozick, which allows for a multiplicity of admissible views. The multiplicity does not entail relativism, however, because not all views are admissible, and the admissible views can be ranked on their merits: “Yet the first ranked view is not completely adequate all by itself; what it omits or distorts or puts out of focus cannot be added compatibly, but must be brought out and highlighted by another incompatible view, itself (even more) inadequate alone” (1981, 22).

The third image to consider is that of disciplines as languages. Bauer uses this metaphor to point out that just as languages have different grammars and not merely different vocabularies, disciplines not only deal with different facts but have different ways of doing things (1990, 112). So this image, unlike that of disciplines as nations or tiles, draws attention to disciplines as modes of human activity, especially if we view a language as a tool and not merely a collection of vocabulary and grammar. Bauer suggests that some languages are good for some purposes and not others, and uses this point to argue for the need for multidisciplinary approaches in science studies (113). Now perhaps some languages are especially suited to certain tasks—for instance, some claim that the ability to easily coin new terms by concatenation facilitates philosophizing in German. But a number of mismatches detract from the usefulness of the language metaphor for conceiving of interdisciplinarity. It is not clear that there are many situations in which one needs to know a number of different languages in order to solve a problem or express an idea. And Bauer himself raises other difficulties: what could “interlingual” speaking possibly be? While we lend words, we shouldn’t mix grammars. And while dialects (subdisciplines) grow up to become new languages, deliberately created “transdisciplinary” artificial languages fail (114).

Yet we should not dismiss the metaphor of disciplines as languages too quickly, because some interesting research on cross-disciplinary work has made use of the notion of pidgins. Peter Galison describes the way physicists and mathematicians with different specialties created a pidgin language to enable them to communicate about the Monte Carlo procedure.[2] Eventually, the pidgin became a full-fledged creole, a language of its own (1996, 152–53). Steve Fuller describes the different ways of conceiving of the relationships between disciplines in linguistic terms: bilingualism (corresponding to Kuhnian incommensurability) is discrete and holistic, requiring one to switch back and forth rather than translate. On the other hand, there are trade languages that can break down barriers (1996, 171). The image of languages helps us avoid static spatial pictures of interdisciplinary relationships and provides a valuable metaphor for the creation of new intermediate disciplines. But the sports metaphor not only highlights activity but helps illustrate the fruitful interactions between different activities. While we sometimes borrow words, we rarely find it helpful to use several languages to try to say the same thing.

Challenges to Disciplinary Pluralism

Pluralism as cross-training is opposed both to the notion that there is one true, best sport for everyone and to the notion that each sport is intrinsically separate and has nothing to do with any other. These two points of view also find expression in the idiom of the metaphor of languages as the ideal of a single universal, adamic, perfect language and the view of absolute untranslatability. In the idiom of nations, these positions go by the familiar names of imperialism and isolationism, and if we consider tiles, they would correspond to the views that one tile is enough for the whole universe and that each tile is strictly separate from its neighbors with no overlap or vagueness at the boundary. While the metaphor of cross-training provides a salutary image for displaying the virtues of disciplinary pluralism, the language of sport unfortunately lacks convenient terms for the vices corresponding to imperialism and isolationism. But we can certainly imagine fanatical gymnasts insisting that their sport is the foundation of all others and subsumes them, or separatist golfers insisting on a pure and single-minded devotion to their game.

Dealing with them in turn, we first confront the view that one discipline is ultimately sufficient for all forms of inquiry. Here we find an extreme form of reductionism, or what Nancy Cartwright calls fundamentalism (1999, 25). Despite the occasional real successes of reductionist approaches, interactions between disciplines encompass a much greater variety than simply reduction to one ultimate unifying field of inquiry. Darden and Maull point this out in science (1977, 60) and Giere insists that in science studies no single theoretical account can be adequate (1999, 63). And yet the sociologist Andrew Pickering talks about the "collapse" of all disciplines into an antidisciplinary synthesis for science studies (1994, 416). Oddly enough, imperialist transdisciplinary talk can come either from the direction of universalizing reductionism or from advocates of radical social constructivism.

Just because the disciplines are historically contingent, multiple, fractured, and blurry does not make them unreal or mean that they all ought to collapse all their domains and methods into one. Although I noted above that all disciplinary research contains an element of openness, opportunism, and even plurality, nonetheless rough distinctions between objects of study and techniques of inquiry apply. Even if the socially constructed boundaries between academic departments were to wither away, why should we think that one kind of training, method, or approach will always suffice? In philosophy, translating things into first-order logic sometimes works, but surely it will not solve every problem we face. Pluralism rejects both universalizing reductionism and bland totalizing collapse into one undifferentiated endeavor.

I conclude by indicating some of the connections between challenges to disciplinary pluralism and other topics in the philosophy of science and in science studies. Opponents of disciplinary imperialism who wish to defend the contingent integrity of disciplines face the opposite temptation of falling into isolationism—the view that each discipline is utterly separate. Those who want to be good at football, this view claims, have nothing to gain from pursuing boxing or ballet. In contrast to this view, pluralism holds out the notion of mutually helpful cooperation. And we see this clearly when Giere says that science studies "must draw on knowledge from many disciplines, including some of the sciences it studies" (1999, 29). The name for this kind of view in philosophical endeavors is naturalism, specifically the kind of normative naturalism suggested by Larry Laudan, Ronald Giere, and Harold Brown.[3] Challenges to naturalism are among the main challenges to a pluralistic approach in philosophy—an approach holding out the possibility of fruitful interactions between the various disciplines that examine scientific knowledge. Another challenge, related to the issues surrounding naturalism, is the purported distinction between the context of discovery and the context of justification. This distinction would seek to isolate epistemology from empirical disciplines such as history and sociology by drawing a sharp line between descriptive and normative questions and insisting that these endeavors are not only different, but separate. Against all those who would claim that each discipline is like a self-contained athletic endeavor, or that any one discipline is like a fundamental, all-encompassing sport, disciplinary pluralism modestly insists that sometimes, at least, there is more than one way to train.

Notes

Many thanks to Michael Reynolds, Helen Longino, and Terry Kent for reading earlier drafts of this material and offering helpful suggestions. Thanks to Geoffrey Gorham for his very helpful response during the Minnesota Center for the Philosophy of Science "Workshop on Scientific Pluralism" in 2002, which led to several important clarifications. Special thanks to Sara Mack, both for crucial encouragement and for suggesting the metaphor of cross-training.

1. The analogue to interfield theories, i.e., the creation of a new nation between two existing ones without changing the original two, is a mismatch for this image.

2. A technique for estimating the solution to an intractable problem by simulating a number of trials with random numbers.

3. It is important to note that the naturalism spoken of here should not be confused with overreaching claims that all phenomena can be fully accounted for by the natural sciences. That view, another kind of imperialism, ought to be called not naturalism but scientism.

References

Argyros, A. J. 1991. *A Blessed Rage for Order*. Ann Arbor: University of Michigan Press.

Bauer, H. H. 1990. "Barriers against Interdisciplinarity: Implications for Studies of Science, Technology, and Society (STS)." *Science, Technology, & Human Values* 15: 105–19.

Bechtel, W. 1987. "Psycholinguistics as a Case of Cross-Disciplinary Research: Symposium Introduction." *Synthese* 72: 293–311.

Brown, H. I. 1988. "Normative Epistemology and Naturalized Epistemology." *Inquiry* 31: 53–78.

Campbell, D. T. 1969. "Ethnocentrism of Disciplines and the Fish-Scale Model of Omniscience." In *Interdisciplinary Relationships in the Social Sciences,* ed. M. Sherif and C. W. Sherif, 328–48. Chicago: Aldine.

Cartwright, N. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Casey, E. S. 1997. *The Fate of Place: A Philosophical History*. Berkeley: University of California Press.

Darden, L., and N. Maull. 1977. "Interfield Theories." *Philosophy of Science* 44: 43–64.

Earman, J. 1989. *World Enough and Space-Time: Absolute versus Relational Theories of Space and Time*. Cambridge, Mass.: MIT Press.

Fuller, S. 1996. "Talking Metaphysical Turkey about Epistemological Chicken, and the Poop on Pidgins." In *The Disunity of Science: Boundaries, Concepts, and Power,* ed. P. Galison and D. Stump, 170–86. Stanford, Calif.: Stanford University Press.

Galison, P. 1996. "Computer Simulations and the Trading Zone." In *The Disunity of Science: Boundaries, Contexts, and Power,* ed. P. Galison and D. Stump, 118–57. Stanford, Calif.: Stanford University Press.

Giere, R. N. 1999. *Science without Laws*. Chicago: University of Chicago Press.

Gieryn, T. F. 1999. *Cultural Boundaries of Science: Credibility on the Line*. Chicago: University of Chicago Press.

Hansson, B. 1999. "Interdisciplinarity: For What Purpose?" *Policy Sciences* 32: 339–43.

Kellert, S. H. 1996. "Science and Literature and Philosophy: The Case of Chaos Theory and Deconstruction." *Configurations* 2: 215–32.

———. 2001. "Extrascientific Uses of Physics: The Case of Nonlinear Dynamics and Legal Theory." *Philosophy of Science* (Proceedings) 68: S455–66.

Klein, J. T. 1990. *Interdisciplinarity: History, Theory, and Practice*. Detroit: Wayne State University Press.

Kockelmans, J. J. 1979. "Why Interdisciplinarity?" In *Interdisciplinarity and Higher Education,* ed. J. J. Kockelmans, 123–60. University Park: Pennsylvania State University Press.

Kuhn, T. 1977. "Objectivity, Value Judgement, and Theory Choice." In *The Essential Tension: Selected Studies in Scientific Tradition and Change,* 320–39. Chicago: University of Chicago Press.

Laudan, L. 1990. "Normative Naturalism." *Philosophy of Science* 57: 44–59.

Nissani, M. 1997. "Ten Cheers for Interdisciplinarity: The Case for Interdisciplinary Knowledge and Research." *Social Science Journal* 34: 201–16.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge, Mass.: Harvard University Press.

Pickering, A. 1994. "After Representation: Science Studies in the Performative Idiom." *Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association,* ed. D. Hull, M. Forbes, and R. M. Burian, 2: 413–19. East Lansing, Mich.: Philosophy of Science Association.

Reynolds, G. H. 1991. “Chaos and the Court.” *Columbia Law Review* 91: 110–17.

Stump, D. 1992. “Naturalized Philosophy of Science with a Plurality of Methods.” *Philosophy of Science* 59: 456–60.

Toulmin, S. E. 1972. *Human Understanding.* Vol. 1. Princeton, N.J.: Princeton University Press.

Wylie, A. 1994. “Discourse, Practice, Context: From HPS to Interdisciplinary Science Studies.” *Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association,* ed. D. Hull, M. Forbes, and R. M. Burian, 2: 393–95. East Lansing, Mich.: Philosophy of Science Association.

Acknowledgments

We would like to thank our friends and colleagues who contributed to the research that led to this book. The Minnesota Center for Philosophy of Science conducted a workshop on scientific pluralism, October 10–13, 2002, during which the authors discussed one another's drafts of the chapters included in this book. In addition to the authors, we gratefully acknowledge the contributions of Paul Magee, Michael Root, Arthur Fine, and Geoffrey Gorham, whose participation in the workshop and constructive commentaries on the works in progress led to a number of improvements. We would also like to thank our anonymous colleagues, including NSF referees for their sound advice about the workshop and book referees for assessments that helped us make important revisions in the final stages of the project.

We would like to acknowledge our enormous debt to Steve Lelchuk, who provided administrative assistance and guidance, beginning with the grant application process, continuing through the workshop activities, revision period, and review process, and concluding with the final editing and publication tasks. We are also indebted to Kathryn Plaisance for scrutinizing the page proofs and working with Steve on the index. Finally, we would like to take this opportunity to express our gratitude to the editors and staff at the University of Minnesota Press.

This research was supported by National Science Foundation grant no. 0135682, Pragmatic C Software, the Minnesota Center for Philosophy of Science, and the College of Liberal Arts of the University of Minnesota.

Stephen Kellert
Helen E. Longino
C. Kenneth Waters

This page intentionally left blank

Contributors

John L. Bell is professor of philosophy at the University of Western Ontario and was formerly a reader in mathematical logic at the London School of Economics at the University of London. He is the author of eight books and some seventy papers in mathematics, mathematical logic, the philosophy of mathematics, and the foundations of physics.

Michael Dickson is professor of philosophy at the University of South Carolina. He works in history and philosophy of science, especially physics. He edits the journal *Philosophy of Science.*

Carla Fehr is associate professor of philosophy at Iowa State University. Her research and teaching explore the philosophy of science, philosophy of biology, and feminist philosophy.

Ronald N. Giere is professor of philosophy and a former director of the Center for Philosophy of Science at the University of Minnesota. He is author of *Understanding Scientific Reasoning, Explaining Science: A Cognitive Approach, Science without Laws,* and *Scientific Perspectives,* and editor of *Cognitive Models of Science* and *Origins of Logical Empiricism* (Minnesota, 1996). He is a past president of the Philosophy of Science Association and a member of the editorial board of the journal *Philosophy of Science.* His current research focuses on the perspectival nature of scientific knowledge and on scientific cognition as a form of distributed cognition.

Geoffrey Hellman has been professor at Indiana University and the University of Minnesota. He has worked primarily in the philosophy of mathematics and in the philosophy of science, especially quantum mechanics. He is author of *Mathematics without Numbers: Towards Modal-structural Interpretation* and coeditor (with Richard Healey) of *Quantum Measurement: Beyond Paradox* (Minnesota, 1998).

Stephen H. Kellert is professor of philosophy at Hamline University and fellow of the Minnesota Center for Philosophy of Science. He is author of *In the Wake of Chaos: Unpredictable Order in Dynamical Systems,* and he is working on a book about the ways nonlinear dynamics, or chaos theory, has been used (and abused) in the fields of economics, law, and literary theory.

Helen E. Longino is professor of philosophy at Stanford University. Her books include *Science as Social Knowledge, The Fate of Knowledge, Feminism and Science* (coedited with Evelyn Fox Keller), and *Women, Gender, and Science* (coedited with Sally Gregory Kohlstedt). She is working on a book-length comparative analysis of approaches in the sciences of behavior.

Alan W. Richardson is professor of philosophy at the University of British Columbia. He has written many essays on logical empiricism and is author of *Carnap's Construction of the World.* He has coedited two volumes in the Minnesota Studies in the Philosophy of Science series, *Origins of Logical Empiricism* (with Ronald N. Giere) and *Logical Empiricism in North America* (with Gary L. Hardcastle). He is now working on a book project, *Logical Empiricism as Scientific Philosophy.*

C. Wade Savage is professor of philosophy at the University of Minnesota and a member of the Minnesota Center for Philosophy of Science. He has been director of the center and was codirector of its institute on consensus in philosophy of science. He is author of *The Measurement of Sensation: A Study of Perceptual Psychophysics,* as well as editor of *Perception and Cognition: Issues in the Foundations of Psychology* and senior editor of *Rereading Russell: Essays on Bertrand Russell's Metaphysics and Epistemology,* both in the Minnesota Studies in the Philosophy of Science series. He has written often on philosophy of science and philosophy of psychology.

Esther-Mirjam Sent is professor of economic theory and policy at the University of Nijmegen in the Netherlands. Her book *The Evolving Rationality of Rational Expectations: An Assessment of Thomas Sargent's Achievements* was awarded the 1999 Gunnar Myrdal Prize of the European Association for Evolutionary Political Economy. She is editor (with Philip Mirowski) of *Science Bought and Sold: Essays in the Economics of Science.* Her research interests include the history and philosophy of economics as well as the economics of science.

C. Kenneth Waters is associate professor of philosophy and director of the Minnesota Center for Philosophy of Science at the University of Minnesota. He has written articles on the nature of scientific knowledge; pluralism and reductionism in biology; and the conceptual, investigative, and explanatory basis of genetics and molecular biology.

This page intentionally left blank

Index