ON REDUCTIONISM:

ONTOLOGICAL, SCIENTIFIC, AND BIOLOGICAL ISSUES IN REDUCTIONISM

by

CARL EUGENE MILLER III

(Under the Direction of Scott A. Kleiner)

ABSTRACT

The subject of the present work is the currently fashionable, scientific view of reality, which I shall denominate reductive materialism. According to this metaphysical theory, the only ontologically real objects are the micro-particles described by physics; the macro-objects of ordinary experience, such as organisms, artifacts, and natural things, are nothing more than collections of these interacting micro-particles.

Reductive materialism is a metaphysical theory about what exists, about what should be included in an ontological inventory of real things. In discussing this theory, I provide a conceptual analysis of ontological reduction, showing what it means to say that an upper-level object is 'nothing more than' a collection of interacting constituent particles, and I argue that the reducibility of upper-level objects to their constituent particles determines whether the laws and theories of the special sciences can in principle be reduced to the laws and theories of physics and chemistry. Thus, the critical question in the debates between reductionists and anti-reductionists is the nature of the objects of experience. Are these upper-level objects real things?

In chapter one, I introduce and examine general epistemological issues in belief formation and evaluation. In chapter two, I discuss ontological reductionism, and a materialist alternative, ontological emergentism, according to which the interaction of material particles can, in some instances, inexplicably give rise to novel ontological objects. I evaluate six objections to the possibility of ontological emergence, and conclude that it remains a viable metaphysical position. In chapter three, I consider scientific reductionism. I argue that several well-known arguments against the reduction of scientific theories are unsuccessful, and that ontological reductionism actually entails scientific reductionism. In chapter four, I take up the 'gene's eye' view of nature, championed by Richard Dawkins, which attempts to reduce organisms to a competition among their genes. I argue that the reduction fails. And finally, in chapter five, I conclude that objections to reductive materialism ultimately spring from its denial of the reality of upper-level objects. Therefore, the problem confronting the anti-reductionist is to supply a metaphysical theory in which upper-level objects can find their reality.

INDEX WORDS: Reduction, Reductionism, Emergence, Emergentism, Supervenience, Selfish genes, Explanation, Skepticism, Materialism, Unit of selection, Physicalism, Epistemology, Metaphysics, Philosophy of science

ON REDUCTIONISM:

ONTOLOGICAL, SCIENTIFIC, AND BIOLOGICAL ISSUES IN REDUCTIONISM

by

CARL EUGENE MILLER III

B.A., Appalachian State University, 1991

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

ON REDUCTIONISM:

ONTOLOGICAL, SCIENTIFIC, AND BIOLOGICAL ISSUES IN REDUCTIONISM

by

CARL EUGENE MILLER III

Major Professor: Scott A. Kleiner

Committee: Robert G. Burton
William L. Power
Richard D. Winfield

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

Affectionately Inscribed

to

Sallie and Carl E. Miller, Jr.

&

Elizabeth and Carl E. Miller, Sr.

ACKNOWLEDGEMENTS

It is customary at the beginning of any book having pretensions to scholarship for the author to perform two tasks: to thank those people who were sources of encouragement and ideas for the work, and to absolve them of any responsibility for the errors and absurdities that it contains. The second of these should be unnecessary, for it would be perverse to hold anyone other than me responsible for the content of these pages, and the first of these can be only inadequately completed, since it is impracticable to list here all the many people who have helped in their own ways to bring me to the place where I could complete this project. But there are several people who are especially deserving of my gratitude, and whom it is a duty and joy to acknowledge.

I wish to thank Scott Kleiner, for patiently reading several drafts of this work, and offering invaluable suggestions for its improvement; Robert Burton, William Power, and Richard Winfield, for serving on my committee; Charles Cross, for helping me to navigate the treacherous shoals of academic bureaucracy; Steve, Carolyn, and Claire Marie Moore, for welcoming me into their home one summer when I was beginning work on this; Jim and Donna Trieschmann, for the constant love and support they have shown me; Wes Yonamine and Mike Hübler, for discussing philosophy with me, and helping me to find my way, though they might disagree about whether I have actually found it; and my parents and grandparents, to whom this work is dedicated, for instilling in me the value of advanced education, though they themselves did not complete college. Finally, I thank my dear wife, Catherine, who has never known me when I was not working on this dissertation. Because of her love, unfailing encouragement, and gentle prodding, it is at length completed; and though the work itself may not be any better for her influence, I most certainly am.

TABLE OF CONTENTS

CHAPTER 1
EPISTEMOLOGICAL PRINCIPLES

The work that follows comprising an evaluation of some aspects of reductive materialism, this first chapter is devoted to a consideration of the standards that I shall employ in that evaluation. The philosopher must not treat his readers as the magician treats his audience, by concealing from them the methods by which he practices his craft. As Charles Sanders Peirce writes, in an essay describing his own methods, '[t]he reader has a right to know how the author's opinions were formed.'[1] Any epistemological and metaphysical assumptions that inform his discussion should be duly acknowledged. I am persuaded that many of the seemingly intractable controversies that exist among philosophers can be traced directly to differences in their fundamental presuppositions. So I shall accordingly attempt to set forth here my own beliefs with respect to the evaluation of theories.

Skepticism

To begin with, I am interested in the extent to which we may ascertain the truth of a theory; that is, whether or not the theory describes faithfully those features of reality that it purports to describe. Other philosophers, of course, having other interests, may be less concerned with the truth of a theory, than with, for example, its empirical adequacy or its success in generating accurate predictions. They may have reasons for discounting considerations of truth. But it is not a question of there being a *right* epistemological approach to take in regard to a theory; it is rather that each of us seeks to answer different questions, according to his interests and commitments.

[1] Peirce (1955) 1, 'Concerning the Author.'

A concern for truth, however, requires that one consider how far we may go in meeting the challenges of those philosophers who are skeptical of the possibility of knowledge. The peculiar question we confront here can be stated briefly: How can one determine whether or not a theory is true? The skeptic provides reasons to deny that we can ever reach such a determination. But even if the skeptic is correct that the high standards of knowledge can not be met, and let me anticipate the conclusion of this section by allowing that I believe he is correct, we may nevertheless inquire into the criteria that one uses in deciding whether or not one shall *believe* that a particular theory is true. This is what I take Quine to be doing, at least in part, in his attempt to naturalize epistemology, and to reduce our evaluative criteria to psychology.[2] As Quine and others have recognized, there is a fundamental distinction between establishing the truth of a theory, and believing that theory because it satisfies criteria that one accepts, and *trusts* to be truth-conducive.

But I am getting ahead of myself. Before discussing the epistemological consequences of making this concession to skepticism, I need to explain why I make it. Skepticism about the possibility of knowledge has been defended with a bewildering variety of arguments, since the very beginning of western philosophy. I shall consider only three such arguments: the problem of criteria, the problem of induction, and the problem of underdetermination.[3] These are sufficient for my purposes.

Plato provides perhaps the clearest and most concise statement of the problem of criteria in the history of philosophy, with his introduction of what has come to be called Meno's paradox, named for the character who introduces this 'trick argument' against Soc-

---

[2] See Quine (1969), especially ch. 3, 'Epistemology Naturalized,' and Quine (1953).

[3] We can conveniently distinguish two types of skepticism. In the first type, the skeptic raises, with respect to any proposed knowledge claim, the possibility that the believer is deceived, and challenges him to show that he is not in fact deceived in his belief. Examples of such skeptical arguments include the possibility that one's belief is actually the product of insanity, hallucination, dreaming, or an evil deceiver, or that one is really a brain in a vat, 'kept alive in a nurturing liquid and subjected to hallucinations that falsely convey the impression of a normal life.' Audi (2003) 300. The second type of skeptical argument, with which I shall be primarily concerned, does not require the believer to prove that he is not mistaken, but instead requires that he demonstrate that his method of belief-formation, his standard of epistemic justification, 'constitutes a path to truth.' BonJour (1985) 8. Accordingly, the skeptic tries to show that the believer is unable to demonstrate the truth-conduciveness of his preferred standards of justification.

rates.[4] According to Meno, it is pointless to pursue knowledge, because we have no way of knowing when we have discovered the truth that we are seeking. Either we know something or we do not know it; but it is unnecessary to seek that which we already know, and unavailing to seek knowledge that we do not possess, because 'if you do not already know the thing that you are endeavoring in your inquiry to discover, then you will be unable to recognize it if it appears.'[5] Without some way of recognizing the truth when we encounter it, the search for knowledge can meet with no success.

What the argument shows is that the pursuit of knowledge, the evaluation of theories, requires the availability of criteria by which we can distinguish true beliefs from false ones, because, as Laurence BonJour suggests, there is no 'immediate and unproblematic access' to true beliefs.[6] These evaluative criteria, then, provide the link between belief and truth. But in removing our epistemological attention from beliefs to criteria of truth, it seems that we have just transferred the problem that confronts us.

We need some truth-conducive criteria, some standards of justification, to validate our beliefs as true, but how shall we validate these standards and criteria themselves? We must establish some connection between the chosen criteria and truth. But the pursuit of truth-conducive criteria will not avoid Meno's paradox, because unless we have some means of identifying truth-conducive standards, we shall not know what standards to look for, nor know when we have found them. The appeal to a metajustification of our criteria of truth would merely transfer the problem once more, and the metajustification would stand in need of its own validation.

We may distinguish five types of justification that a belief or a criterion of truth can possess. It may be either unjustified, circularly justified, justified by some other standard,

---

[4] Plato, *Meno*, 80d-e.
[5] White (1976) 43.
[6] BonJour (1985) 7. James (2000) 207 notes that '[n]o concrete test of what is really true has ever been agreed upon.' 'The Will to Believe,' Sect. VI.

justified by an infinite epistemic chain, or self-justified. Unfortunately, none of these options really succeeds in diminishing the force of Meno's objection.

First, an unjustified belief or criterion, having no established relation to truth, remains entrammeled in skepticism; such a belief may be said to be 'anchored in sand.'[7] Second, attempts at circular justification, including the vast justificatory networks characteristic of coherence theories, ultimately fare little better, because it would need to be shown that the circle or network of mutually supporting beliefs corresponds with reality. The connection between coherence and truth would still remain to be established, presumably on grounds other than coherence, lest one should beg the question against the skeptic.[8] Since presumably there can 'be a vast number of equally coherent systems of beliefs that are mutually incompatible, . . . why should my having one of these coherent systems provide any reason to think *my* beliefs, rather than someone with one of the "opposing" systems, are justified or represent knowledge?'[9]

Third, appeals to some other standard necessarily raise the question of whence that standard receives its justification. Fourth, with respect to the infinite series of justifying standards, among other problems '[i]t is doubtful that, given our psychological make-up, we *can* know, or even believe, infinitely many things.'[10] And fifth, as regards a foundational, self-justified standard, what Robert Audi calls a 'directly justified belief,'[11] the skeptic might properly ask what reason there is to suppose that such a foundational belief is true.[12] William Alston concludes, after a lengthy discussion of the trustworthiness of our beliefs and the challenge of skepticism, that 'with respect to even those sources of belief of which we are

[7] Audi (2003) 188.
[8] BonJour (1985) ch. 8, argues, by means of an argument to the best explanation, that the continued coherence of a system of beliefs is evidence of its truth.
[9] Audi (2003) 199.
[10] Ibid., 189.
[11] Ibid., 194.
[12] Audi argues that such foundational beliefs are not necessarily true; they are presumptively true but fallible.

normally the most confident we have no sufficient noncircular reason for taking them to be reliable.'[13]

But he and a number of other prominent contemporary epistemologists, including Alvin Goldman, Fred Dretske, Robert Audi, and David Armstrong, have argued that the skeptical conclusion can be avoided, or at least minimized, by reconsidering the requirements of knowledge. Skepticism arises, they contend, because we have no way of determining whether our basic beliefs, or our accepted criteria of truth, are justified. But what matters for knowledge, in their view, is whether the beliefs or standards in question *are* justified, 'whether or not anyone realizes in any way that this is so.'[14] According to this view of justification, which has come to be denominated *externalism*, our criteria of justification must correspond with reality, in order for any belief based on those criteria to be properly known, but the fact of that correspondence need not be available to the knower. The externalist is concerned with 'the comparative *reliability* of the [belief-causing] processes.'[15] As long as the belief was in fact caused by an epistemically reliable process, the belief constitutes knowledge. Thus, externalism can permit attribution of knowledge to a person who can provide no justification.

Externalism is a relatively recent development in epistemology. Indeed, in BonJour's estimation, 'until very recent times, no serious philosopher of knowledge would have dreamed of suggesting that a person's beliefs might be epistemically justified merely in virtue of facts or relations that are external to his subjective conception.'[16] But, as BonJour also recognizes,[17] it is difficult to argue against externalism without begging the question against that view, although he does introduce numerous 'intuitive' counter-examples.

But BonJour's efforts may be unnecessary. The externalist has merely changed the conditions of what is to count as knowledge. He no longer includes among the require-

[13] Alston (1991) 146.
[14] BonJour (1985) 33.
[15] Alvin Goldman, 1979, *Justification and Knowledge*, ed. G. S. Pappas (Dordrecht: D. Reidel), emphasis added.
[16] BonJour (1985) 36
[17] Ibid., 37.

ments of knowledge, as most philosophers traditionally have done, that a person be aware that his standards of justification are truth-conducive; it is sufficient that the standards accurately and reliably identify true beliefs. But the externalist is free to define knowledge in any way that he chooses: 'Words,' wrote Samuel Johnson, 'are the daughters of men.' Therefore, it is no criticism of the externalist position, as BonJour seems to believe, to show that externalism is incompatible with tradition, since the externalist has expressly disavowed the traditional understanding. The externalist does not bind himself to customary usage. And I do not see that there is any argument that can prove that he should.

The main virtue of the externalist position is 'how neatly [it] deals with traditional and modern forms of skepticism.'[18] By redefining knowledge, the externalist can resist the force of skepticism. If my beliefs are produced by an evil deceiver, or if I am in fact a brain suspended in a vat, 'then [my beliefs] are false and I am ignorant. On the other hand, if the beliefs are true and produced in the appropriate way, then I do know.'[19]

Notwithstanding this advantage of externalism, I must admit that the position holds no appeal for me. It seems to me that unless we have some means of distinguishing the reliable, truth-producing processes from the unreliable variety, we can claim only that the skeptic has not defeated our claims to knowledge; but likewise, we have not thereby shown that we do have any knowledge, because we have not shown that our belief-producing processes are epistemically reliable, only that *if* they are reliable, then we have knowledge. As Keith Lehrer puts it, '[o]ne requires information about whether the received information is trustworthy or not, and lacking such information, one falls short of knowledge.'[20] The goal of epistemology, I take it, is to distinguish true beliefs from false ones, to identify epistemic criteria that track truth, and that provide a 'mediating link between our subjective starting point and our objective goal,'[21] but I do not find that externalism achieves that goal. It is

[18] Keith Lehrer, 1990, *Theory of Knowledge* (Westview Press: Boulder, Colorado).
[19] Ibid.
[20] Ibid.
[21] BonJour (1985) 7.

little consolation to be told that *if* my belief-producing processes are reliable and non-deceptive then I have knowledge. What I want to know is whether they *are* reliable and non-deceptive. But I leave others to derive what consolation they can from this position.

The second skeptical argument, the problem of induction, is typically credited to David Hume.[22] What he pointed out, in its simplest form, is that there is no logically valid way to move inductively from what we know to what we do not know. This is a significant conclusion. It means that, absent some assumption of uniformity, we are not justified in drawing inductive generalizations about the world, if we understand justification as being an indicator of truth. Logically speaking, the conclusion of an inductive argument does not follow necessarily from its premises, so it is possible that the premises are true and yet the conclusion drawn from them is false. An inductive conclusion is ampliative, in the 'sense of going beyond the evidence in hand.'[23] For all we know, for all that is contained in the premises, the conclusion of any inductive argument might be false; therefore, we can not be said to know the conclusion, even if we are persuaded that the premises are true.

The problem is most acute in our attempts to discover the so-called laws of nature, and to derive from them accurate predictions about the future. A law of nature describes a relation between events that is distinguished from a coincidental correlation. In a lawful relation, one occurrence is caused by another. But the problem arises when we attempt to determine whether or not some observed pattern of events expresses a causal relation. Though we have observed the pattern, perhaps in great detail, can we ever validly conclude that it is more than a coincidence, or that the causal relation will continue into the future? Hume answers that we can not ever establish that there is an underlying cause, nor that the pattern will persist.

A causal relation, according to Hume, requires a necessary connection between the cause and effect, which is simply another way of saying that a causal relation is not merely

[22] See Hume (1975) Sect. IV.
[23] Rescher (1980) 6.

an accidental correlation. But the persistent correlation of events, which Hume calls their 'constant conjunction,' does not reveal any necessary connection between them; the putative cause and its effect are separate events, so we can not discover any necessity in their being joined together. Consequently, at most we can say that in our experience certain events have always occurred together; we are not logically entitled to suppose that the relation between them will continue to obtain.

A law of nature is also a statement about a uniformity or regularity of nature. We are willing to make predictions about the future course of natural events because we assume that the patterns and processes that we have discovered in the past will continue to hold in the future. The extrapolation from past to future must assume, in some respect, the uniformity of nature. According to Hume, induction must 'proceed upon that principle, *that instances of which we have had no experience, must resemble those, of which we have had experience, and that the course of nature continues always uniformly the same*.'[24] But the problem of induction suggests that such an extrapolation is invalid, for two reasons. First, the relation between the past and the future is not discoverable in the events of the past; there is nothing in what we have observed to occur that indicates that future events will follow a similar pattern. And second, even if we could establish a causal connection between events, we could not demonstrate that the causal connection would continue to operate in the future. As Hume puts it, the 'appeal to past experience decides nothing in the present case; . . . [it] can never prove, that the same [causal] power must continue in the same object or collection of sensible qualities.'[25]

The problem of induction seems to be an insuperable barrier to knowledge. We can not determine whether or not our lawful generalizations about nature, and its future course, are true, because the evidence does not entail the conclusion that we draw from it. However, given this skeptical conclusion, there are several options available to us. First, we may

---

[24] Hume (1985) 137, Book I, Part III, Sect. VI, emphasis in the original.
[25] Ibid., 139.

adopt a pragmatic attitude[26] towards induction, an admittedly question-begging approach,[27] and continue to use it because it has proved to be reliable in the past; second, we may follow Hume in recognizing our inductive practices as unavoidable habits of thought; or third, we may embrace the efforts of Karl Popper to replace scientific induction with falsification. But I do not see that we have much hope of escaping the meshes of skepticism with regard to induction.

The third skeptical problem of underdetermination arises in connection with the formation of theories and hypotheses. According to Peter Lipton, 'to say that an outcome is underdetermined is to say that some information about initial conditions and rules or principles does not guarantee a unique outcome.'[28] It is customary, in discussions of underdetermination, to point out that the data points plotted on a graph do not determine the curve that passes through them, since many, even an infinite number of curves, will pass through all of the points.[29] Similarly, a historian or police detective confronted with various established 'facts' must somehow construct a theory or explanation that 'fits' them satisfactorily. However, the facts are compatible with other theories and explanations that the historian or detective does not find particularly satisfactory. There are numerous equivalent theories that cover all the available evidence.

And since 'scientists never have all the relevant data,'[30] their theories are always underdetermined. This means that, since 'there is no general algorithm that could take [the

---

[26] Rescher (1980) calls this the 'pragmatic approach to justifying induction,' because inductive arguments are necessary to our survival in 'a difficult and often hostile environment' (p. 6), and 'we must face the fact that, in the circumstances, this sort of [defense of inductive] argumentation is the strongest that can reasonable be asked for, because it is the strongest that can be had' (p. 53).

[27] Although Rescher (1980) allows that inductive generalizations about nature go beyond experience, which he calls the 'vexing problem of rationalizing the cognitively crucial step from particular experiences to the reasoned acceptance of empirical generalizations' (p. 51), he argues that nonetheless '[e]xperience is the only route to the destination in these matters of generality; it is the only game in town' (p. 52). 'A rational animal,' he writes, 'facing the uncertain future must triangulate from past experience: it must make its way in a difficult world by its cognitive wits. Its reliance on experience is natural and inevitable, and thereby also legitimate, seeing that no rationally superior alternative lies to hand' (p.53).

[28] Lipton (1991) 6.

[29] E.g., Lipton (1991) 6, and Harré (1992) 44.

[30] Lipton (1991) 88.

scientist] from data to a hypothesis,'[31] the actual choice of a theory, the one that the scientist actually believes, is ultimately determined by considerations and principles other than the data that must be accommodated. But if there are other 'equivalent' theories, including innumerable theories that no one has even imagined, and outlandish theories that no one would entertain, then how can anyone be sure that the chosen theory is the correct one? Among the infinite number of curves that will fit the data points, how can we determine which one is true?

I think we must concede to the skeptic that we can not make such a determination. Any criteria we use to choose the true theory would have to be truth-conducive, but as I discussed under the problem of criteria, we have rather narrow prospects of establishing that any criteria are truth-conducive. However, it is clear that, notwithstanding the concession to skepticism, we do nonetheless appeal to various criteria in coming to believe one theory or hypothesis over another. Meeting these criteria renders a theory more satisfying to us than one that does not meet them, or meets them to a lesser degree. Hence, of all the curves that can be overlaid on a set of data points, we are likely to choose one that satisfies to a greater degree, however that is measured, our epistemic criteria, including such virtues as parsimony, predictability, and fruitfulness; and the researcher, trying to account for a body of facts, accepts the explanation that seems most plausible, given the principles of selection that he acknowledges. In section four of this chapter, I shall consider some of the criteria that influence our choice of theories.

But before I conclude this section on skepticism, and consider in the next the nature of beliefs, it will be useful to summarize my conclusions regarding these three skeptical problems. Most of us are concerned that our beliefs about the world be to some extent true, but how do we establish that they are true? This it is rôle of justification, to act as 'a *means* to truth.'[32] Since beliefs do not reveal their own truth or falsity,[33] we must hold them up and

---

[31] Ibid.
[32] BonJour (1985) 7.

compare them, so to speak, with truth-conducive standards. Being justified, then, means that a belief satisfies standards that warrant its truth. But if the foregoing arguments for skepticism are correct, then we can not hope to provide a justification for our beliefs that ensures that they are true, inasmuch as we have no way of determining whether the standards that we use for this purpose are in fact truth-conducive.

These criteria of justification are necessary because our theories and lawful generalizations about the world are underdetermined by the evidence, and as Lipton observes, 'there are possible worlds where the principles [of justification] we use . . . fail us.'[34] In order to defeat the skeptic, we should need to show that in the actual world our accepted standards do not in fact lead us to false conclusions. But I can see no possible way of satisfying the thoroughgoing skeptic upon this point, without begging the critical questions. BonJour admits that 'not all [varieties of skepticism] can be successfully answered,' [35] and Alston concludes that in our attempts to establish the reliability of our basic sources of beliefs 'epistemic circularity thus would appear to be inescapable.'[36]

If it is not possible to provide the requisite justification of our beliefs, is there any way to avoid the need to provide it? BonJour identifies three approaches to which many philosophers have been attracted.[37] First, one can simply change what it means for a belief to be justified, or to be true. The externalists, as we have seen, remove the requirement that one be aware of the truth-conduciveness of one's criteria of justification. Likewise, pragmatists such as Peirce and William James change the meaning of truth from correspondence with objective reality, to the view that truth is that conclusion which we accept in the long run; 'pragmatism faces forward to the future.'[38] According to James, 'the "absolutely" true, meaning what no further experience will ever alter, is that ideal vanishing-point towards

---

[33] '[N]o bell in us tolls to let us know for certain when truth is in our grasp.' James (2000) 218, 'The Will to Believe,' Sect. X.
[34] Lipton (1991) 9.
[35] BonJour (1985) 14.
[36] Alston (1991) 147.
[37] BonJour (1991) 11.
[38] James (2000) 99, *Pragmatism*, 'Pragmatism's Conception of Truth.'

which we imagine that all our temporary truths will someday converge.'[39] Adopting a verificationist criterion of truth, James identifies truth with what we *accept* as the truth, in the long run, when there is no more evidence to take into account that would cause us to reconsider our theories. In a similar vein, Peirce writes that '[t]he opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth.'[40] As Hilary Putnam explains, James and Peirce 'declare[their] allegiance to a notion of truth *defined in terms of ultimate consensus*.'[41]

Second, one can appeal to common sense in identifying what one knows. In his famous essay, 'A Defense of Common Sense,' G. E. Moore enumerates 'a set of propositions, every one of which (in my own opinion) I *know*, with certainty, to be true.'[42] These include the propositions that he has a body, which was born in the past, and has existed continuously ever since, while undergoing some changes. But on what grounds does he claim to know these 'truths'? What justifies his certainty? Moore does not provide any such justification, other than the claim that these beliefs are part of the 'Common Sense view of the world,'[43] because he apparently does not think it necessary to do so. 'How absurd it would be to suggest that I did not know [these things], but only believed [them], and that perhaps [they were] not the case!'[44] Remember, justification is that which transforms a mere belief, so to speak, into knowledge; it connects our beliefs with truth. But it is pointless to try to connect knowledge with truth, for what is known *is* true: if we know some proposition, then no further justification is needful. Moore simply affirms, or assumes, from the outset that these propositions, these 'obvious truisms,'[45] are known to himself, and to the majority of human beings as well. Among our stock of beliefs, we recognize some as constituting

[39] Ibid., 98.
[40] Peirce (1955) 38, 'How to Make Our Ideas Clear.'
[41] Putnam (1997) 169, 'James's Theory of Truth,' emphasis in the original.
[42] Moore (1959) 32, 'A Defense of Common Sense,' emphasis in the original.
[43] Ibid., 45.
[44] Ibid., 146, 'Proof of an External World.'
[45] Ibid., 32, 'A Defense of Common Sense.'

knowledge, and consequently we do not need to trouble ourselves to find justification for them, until, we might say with Peirce, there arises some doubt about the matter.

And third, some philosophers, including Richard Rorty and Michael Williams, have declined to engage the problems of skepticism. Rorty, for one, has concluded that one should simply 'refuse to take [skepticism] seriously.'[46] But by adopting a version of epistemological relativism, according to which our belief systems may not be said to mirror the structure of reality, Rorty would seem to have taken skepticism very seriously indeed. A similar dismissive approach is adopted by Audi, who, without subsiding into relativism, attempts to shift the burden of proof to the skeptic. For Audi, 'a belief can constitute knowledge without being infallible,'[47] and 'requir[ing] that a belief can be knowledge only if . . . it can be conclusively shown to be true would . . . beg the question against the commonsense view that a belief can constitute knowledge without being infallible.'[48] Since the skeptic has not produced any argument to show that knowledge must be infallible, we should not be greatly vexed by skeptical arguments that call our knowledge into question. Audi thereby succeeds in securing an impregnable defense for our knowledge claims, but at the expense of making a fatuous argument. It is quite impossible for anyone to prove that knowledge requires infallibility, since what we mean by knowledge is determined by what we choose to mean by knowledge. But Audi's general point seems to be that we are epistemically entitled, under certain situations, to claim knowledge for our beliefs, unless the skeptic can advance good reasons to show that we are in fact deceived; it is not sufficient that we *might* be

[46] Richard Rorty, 1979, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press), quoted in BonJour (1991) 13.
[47] Audi (1988) 144.
[48] Audi (2003) 306. Audi insists that the skeptic, by requiring that knowledge be infallible, begs the question against the commonsense view, but it is not clear why one could not assert, with equal propriety, that the commonsense view begs the question against the skeptic! In *Belief, Justification, and Knowledge* and *Epistemology*, Audi defers his discussion of skepticism till the last chapter. Part of his reason for doing this, I suspect, is that it gives the commonsense view command of the field, and thus allows Audi to allege that unless the skeptic can defend his infallibility requirement for knowledge, the commonsense view must prevail by default.

deceived. A person can claim to have knowledge, even without 'knowing, of every possible explanation of how [his] belief could be false, that this explanation is incorrect.'[49]

But even if Audi meets the skeptical challenge by disclaiming the need to show that he is not deceived in his beliefs,[50] what positive reasons does he provide to think that his foundational beliefs are true? Audi peppers his text with italicized epistemic principles. For example, the *visual experience principle* tells us that 'when, on the basis of an apparently normal visual experience[,] . . . one believes something of the kind the experience seems to show[,] . . . normally this belief is justified';[51] and the *visual knowledge principle* states that 'at least normally, if we see that a thing . . . has a property[,] . . . we (visually) know that it has it.'[52] But why should we accept these principles? What positive reason does Audi provide for his claims that perception is 'normally' truth-conducive?

Audi insulates his account of knowledge from the need to answer such questions. First, as have seen, Audi repudiates the requirement that the believer prove that he is not deceived, and devolves upon the skeptic the burden of showing that the believer *is* deceived. Thus, we can claim knowledge for our perceptual beliefs because we have 'no reason to doubt that these perceptual beliefs are commonly justified or that, quite often, they are true and constitute knowledge.'[53] Second, as an externalist, Audi denies that having knowledge means being able to give a justification of that knowledge. If the perceptual beliefs are reliably produced, that is, if perception is in fact a truth-conducive source of beliefs, then such perceptual beliefs normally count as knowledge, even if the perceiver can provide no justification of his beliefs. And in default of skeptical evidence to the contrary, we have 'no reason to doubt that perception is a rich and basic source of both knowledge and justification.'[54]

---

[49] Audi (2003) 308.
[50] Though Audi may perhaps rebut those skeptical arguments that challenge the believer to prove that he is not deceived in his beliefs (see footnote 3 for the distinction between the two types of skeptical arguments), he must nevertheless offer some positive defense of any claims to knowledge.
[51] Audi (2003) 28.
[52] Ibid.
[53] Ibid., 29.
[54] Ibid., 51.

I remain unpersuaded, and somewhat disappointed, by Audi's attempt to '*rebut* skepticism.'[55] Audi's epistemic foundations seem to be nothing more than articles of unacknowledged faith, from which Audi will not be driven unless the skeptic can show that such foundations are mistaken.[56] 'People,' George Santayana has wisely observed, 'are not naturally skeptics.'[57] But we typically expect that, of all people, philosophers will adopt the critical stance towards their first principles. Although I agree with Santayana that 'a philosopher to-day would be ridiculous and negligible who had not strained his dogmas thought the utmost rigours of skepticism,'[58] I do not fault Audi, in James' words, for his 'faith that truth exists, and that our minds can find it,'[59] for I, too, accept 'the lawfulness of voluntarily adopted faith.'[60] I do think Audi is culpable, however, for his unwillingness to admit that faith stands at the foundation of his system, because such obstinacy strikes me as a disingenuous attempt to appropriate more rationality for his beliefs than they warrant. Far better, it seems to me, for the philosopher to 'push skepticism as far as [he] logically can, and endeavour to clear [his] mind of illusion,'[61] and to insist on 'the right to believe at [his] own risk any hypothesis that is live enough to tempt [his] will.'[62]

---

[55] Ibid., 317.

[56] Audi (2003) 324 himself addresses this point in the last paragraph of *Epistemology*: 'Perhaps viewing knowledge, justification, and rationality in the way I have might be thought to be an article of epistemological faith. I do not think it is; but the difficulty of determining whether it is partly an article of unverifiable faith, or can be established by cogent argument, or is more than the former yet less of the latter, is some testimony to the depth and complexity of skeptical problems.'

This denial of faith is misleading in at least two ways. First, it is impossible to provide an argument, cogent or otherwise, for the truthfulness of the foundational principles of a foundationalist epistemology, without those foundations ceasing to be the foundations at all. If the truth-conduciveness of the foundational principles, such as the veridicality of normal perception, is established on other grounds, that is, if we provide reasons for thinking that perception is reliable, then it is those grounds that become the foundations of the system. Second, the appeal to faith at the heart of foundationalism has nothing whatever to do with the 'depth and complexity of skeptical arguments'; it has to do rather with the very nature of foundationalist epistemology, which depends on belief in the truthfulness of principles for which no proof is available. Foundationalism tries to escape skepticism by positing non-inferential basic beliefs as the ground of our derivative, non-basic beliefs, but those basic beliefs are supported only by the sturdy faith that, at least in these principles, we have grasped the truth.

[57] Santayana (1923) 11-12.

[58] Ibid., 9.

[59] James (2000) 205, 'The Will to Believe,' Sect. V.

[60] Ibid., 198, Introduction.

[61] Santayana (1923) 10.

[62] James (2000) 217, 'The Will to Believe,' Sect. X.

Therefore, although these three solutions to skepticism have found a number of adherents, particularly among recent philosophers, I do not choose to follow any of these strategies in responding to the problems of skepticism. I have no desire to preserve artificially the truthfulness of my beliefs, either by redefining truth, by dogmatically asserting that my beliefs are known to be true, or by dismissing the force of skepticism. I prefer instead to acknowledge candidly the inescapable force of skepticism, and to follow Quine in considering how we actually go about forming and evaluating beliefs, whether or not we are justified in our epistemic practices.

Beliefs as Commitments

Following on the skeptical conclusions of the previous section, it seems clear that each of us holds a vast assortment of beliefs, despite our inability to justify them against the objections of the skeptic. It does not seem possible for us ever to succeed in establishing that any of our beliefs is true. But nonetheless we are committed to our beliefs with varying degrees of confidence. We have 'faith' in their correctness. And by faith, I mean that we take a chance on our beliefs, so to speak, and accept the risk of their being wrong.

Consequently, I reject the popular view that 'faith' is a component exclusively of religious discourse, and of supernatural speculation, and that the proper employment of the scientific method can somehow obviate the need for faith. For faith, as I am using the term, is simply one's willingness to believe that some proposition is correct, even though one is not finally justified in that belief. And the foregoing skeptical arguments, which I find irresistible, deny that one is ever so justified. Therefore, since faith is an element of every belief, the scientist can avoid appeals to faith only by refusing to hold any beliefs.

A common objection to this account of faith is represented in the celebrated passage of W. K. Clifford, that 'it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence.'[63] Clifford seems to denigrate any appeal to faith. But whether or not there is any sense in what he says depends upon what he means by insufficient evi-

[63] W. K. Clifford, *The Ethics of Belief.*

dence. If he intends that the evidence or justification for a belief should be sufficient to put it beyond the challenge of skepticism, then it would seem that all beliefs, with the possible exception of those about our immediate states of consciousness,[64] are insufficiently supported, and that it is wrong always, everywhere, and for anyone to believe anything. But if, on the other hand, he merely requires that the evidence be sufficient to the believer, then beliefs accepted on faith must be admitted as legitimate. And if the skeptic is correct, as I believe he is, that all standards of justification are finally subjective, then Clifford must either deny the legitimacy of beliefs, or admit those that appeal to faith.

In acknowledging the fact that our beliefs are based on our faith in their correctness, J. J. C. Smart has said that 'the best we can do is to choose the best warranted hypothesis . . . in the *hope* that warranted assertability is a pointer towards truth. The pointer may lead us astray, since the universe may always trick us. Still, we do the best we can.'[65] We accept certain epistemic criteria as being indicative of truth, and we are inclined to accept beliefs that satisfy those criteria. Smart carefully distinguishes our 'best explanation,' which is one that 'fits best into our web of beliefs,'[66] from the true explanation, since, he recognizes, 'truth differs from coherence.'[67]

Peirce is correct, I think, when he writes that we 'seek for a belief that we shall *think* to be true.'[68] The pursuit of truth thus amounts to the quest for beliefs in whose truth we repose great confidence, and once this confidence has been achieved, and our opinions have been settled, 'we are entirely satisfied, whether the belief be true or false.'[69] Only when a genuine doubt intrudes upon one's beliefs does one feel a need to reform them. It is this 'ir-

---

[64] For instance, James (2000) 207 writes that '[t]here is but one indefectibly certain truth, and that is the truth that pyrrhonistic skepticism itself leaves standing,—the truth that the present phenomenon of consciousness exists.' 'The Will to Believe,' Sect. VI.
[65] Smart (1990) 10, emphasis added.
[66] Ibid., 9.
[67] Ibid., 7.
[68] Peirce (1955) 11, 'The Fixation of Belief,' emphasis in the original.
[69] Ibid., 10.

ritation of doubt [that] is the only immediate motive for the struggle to attain belief.'[70] But, it seems clear to me, whether or not one experiences a genuine doubt with respect to one's system of beliefs, and therefore desires to remove it, is ultimately a subjective matter. We accept those beliefs that satisfy our individual epistemic standards, and we evaluate beliefs according to personal criteria.

Socrates recognized this. Accordingly, he conducted his philosophical discussions by way of the *elenchus*, the eponymous Socratic method. In order to prevail upon another person to reëvaluate his beliefs, one must discover, through close questioning, precisely what the other person believes and what evaluative criteria he accepts. It will avail the philosopher nothing if he advances arguments whose premises or conclusion his audience finds incredible, or if, as James puts it, the philosopher asks someone to believe something that 'makes no electric connection with [his] nature.'[71]

But despite the skeptic's seemingly irrefutable case against the justification of our beliefs, we continue to have beliefs, we come to doubt the correctness of beliefs held previously, and we reform our belief systems in response to these doubts. As Lipton recognizes, 'even if our inferences [are] unjustifiable, one still might be interested in saying how they work.'[72] How do we make epistemological decisions? How do we decide which theories or hypotheses to accept, and which ones to reject? What principles influence the reformation of our belief systems? What, asks Lipton, is 'the black box mechanism that governs our inductive practices'?[73] These questions are quite independent of those concerning the justification of our beliefs. We can seek, with Peirce, to identify the principles that dispose us to *think* that a belief is true, without attending to the question of whether those beliefs really are

---

[70] Ibid.

[71] James (2000) 199, 'The Will to Believe,' I. Hume (1975) made a similar point in his famous remark, concerning Bishop Berkeley's arguments against the existence of matter, 'that they admit of no answer and produce no conviction.' Footnote, Sect. XII, Part I.

[72] Lipton (1991) 14.

[73] Ibid., 133.

true. And if one has concluded that the justification of our beliefs can not be acquired, then the better to address the question of what guides our epistemological choices.

In seeking to discover these guiding principles, we must find either that each of us appeals to different criteria in evaluating a belief, or that there is some general agreement concerning the epistemic traits of a sound theory. I think that the latter more closely describes the epistemic situation in which we find ourselves. Most of us share generally similar standards of evaluation; most people, for instance, consider it a virtue of a theory if it should agree with other beliefs that one holds, and not contradict them, and if it should correctly predict some future course of events. But it would be a mistake to suppose that the *application* of these principles does not vary significantly among individuals. For example, while there may be wide agreement that one's beliefs should cohere with one another, there exists nevertheless considerable disagreement about how much incoherence one should countenance within a set of beliefs. And how one chooses to reform one's beliefs, in order to remove that incoherence, is similarly subject to considerable divergence.

In the balance of this chapter, I propose first to consider the different 'types' of beliefs, and second to investigate the epistemic criteria we employ in evaluating beliefs. But given my concession to skepticism that we have no way of recognizing truth-conducive epistemic standards, the most I can actually do is suggest the standards that I find persuasive, and hope that others acknowledge the same principles of evaluation. However, there are no guarantees. If someone is willing to admit into his system of beliefs a higher level of incoherence than I am, or if someone disdains altogether the need for logical consistency, I do not see how I could demonstrate that he is in error. Our individual beliefs may be irremediably subjective, but, as Smart suggested, 'we do the best we can.'[74]

A System of Beliefs

How do beliefs differ from one another? Do beliefs fall 'naturally' into different categories? Are there different 'kinds' of beliefs? Having traditionally answered these last two

[74] Smart (1990) 10.

questions in the affirmative, philosophers have sought to identify the sundry categories into which beliefs can be divided. We have been told, for instance, that there is a distinction in kind between synthetic and analytic beliefs, between observable and non-observable phenomena, between empiricism and rationalism, and between science and pseudo-science, which may include anything from metaphysics to magic and astrology. These distinctions are supposed to be significant and non-arbitrary. Philosophers place great stock in them.

Analytic beliefs, for example, are purported to be discoverable by the light of pure reason, and therefore to be true in every possible world, whereas synthetic beliefs, concerning contingent states of affairs, are not necessarily true. Hume embraced a similar distinction, later denominated 'Hume's fork,' between demonstrable 'relations of ideas,' and contingent 'matters of fact.'[75] Moreover, the 'facts' discovered by observation are supposed to be the empirical ground upon which rational speculation about the unobservable world is founded. And science, allegedly pursuing its unique method of data collection and experimental verification, is widely thought to generate beliefs that differ materially from, and enjoy a greater claim to knowledge than, those reached through non-scientific pursuits. It is common in texts on the philosophy of science, for instance, to propose some distinction between scientific and philosophical beliefs, often with the implication that the latter are less respectable, and might be done away with without much loss to our knowledge. Biologist Edward O. Wilson is one of the legion who subscribe to such a position. In his book on the nature of scientific knowledge, Wilson writes that '[w]e have the common goal of turning as much philosophy as possible into science.'[76] He seems to believe that philosophical speculation is that activity with which we amuse ourselves until science provides understanding, and thereby transforms the unknown into the known. His view is widely shared.

I reject this view, because I am persuaded that our ordinary division of beliefs into kinds, such as empirical or rational beliefs, and scientific or philosophical beliefs, is arbitrary

[75] Hume (1975) Sect. IV, Part I.
[76] Wilson (1998) 12.

and subjective. I think we shall come nearer to a correct understanding of our beliefs, if we understand the difference between them as one of degree and not one of kind. This conclusion derives from the skeptical results we reached in the first section of this chapter. If it is the case, as I have urged, that none of our beliefs can finally be justified, and that we hold them on faith, then our beliefs differ only in the strength of our commitment to them, and in our willingness to accept them as true.

In the end, perhaps all we can say about a belief is that, for reasons we consider compelling, we find it *believable*. Of course, we find some possible beliefs more credible than others, and many of us may be more inclined to accept 'scientific' beliefs, than those based on 'superstition.' But the distinction between beliefs is not based on any established access to truth; the epistemological quest to show that our beliefs are true, or likely to be true, has fully come a cropper. Indeed, there seems no way, even in principle, for us ever to succeed in establishing the truth of a belief. So we are left with subjective distinctions between beliefs, based on our individual credulity.

This is a position similar, though not identical, to that advocated by Quine in his classic article, 'Two Dogmas of Empiricism.' In that article, it will be remembered, Quine is concerned in part to show what he perceives to be the consequences for epistemology of abandoning the traditional distinction between analytic and synthetic truths.[77] The effects of abandoning this 'dogma' are two-fold: 'a blurring of the supposed boundary between speculative metaphysics and natural science . . . [and] a shift towards pragmatism.'[78] In other words, Quine rejects Hume's fork: there are no necessary 'truths of reason . . . which could

[77] For Quine, the inadequacy of the analytic/synthetic distinction is based on our inability to attribute any significance to the concept of analyticity. Whether an analytic belief is defined as one that is 'true by virtue of meanings and independently of fact,' (Quine (1953) 21, 'Two Dogmas of Empiricism,' §1), or one whose terms are interchangeably synonymous, Quine shows that we have not so much defined the term, as exchanged it for another term that stands in need of its own signification:

> Analyticity at first seemed most naturally definable by appeal to a realm of meanings. On refinement, the appeal to meanings gave way to an appeal to synonymy or definition. But definition turned out to be a will-o'-the-wisp, and synonymy turned out to be best understood only by dint of a prior appeal to analyticity itself. So we are back at the problem of analyticity. (Ibid., 32, §4.)

[78] Quine (1953) 20, 'Two Dogmas of Empiricism,' Introduction.

not possibly be false.'[79] Rather, 'the totality of our so-called knowledge or beliefs . . . is a man-made fabric which impinges on experience only along the edges.'[80] Consequently, all of our beliefs are revisable, including even the laws of logic.

I think Quine is correct to insist that all of our beliefs are in principle revisable, and to abandon the supposed distinction between analytic and synthetic truths, but I think that he does not go far enough in rejecting other dubious distinctions. He writes of the 'conflict [of our beliefs] with experience . . . [which] occasions readjustments in the interior of the field' of our beliefs.[81] But as Smart has pointed out, in epistemology '[t]here is no bed-rock,'[82] and this includes experience, which, for Quine, represents the 'boundary conditions'[83] of our beliefs. At times, Quine seems to suppose an absolute distinction between our set of beliefs and the experiences or observations that give rise to the reëvaluation and re-adjustment of those beliefs. But such a distinction is arbitrary, for if all beliefs are revisable, then our observational beliefs are likewise subject to revision, and thus can not be the absolute test of all our other beliefs.

We have different degrees of confidence in the correctness of individual beliefs, which means a greater or lesser willingness to revise them. Some of our beliefs are subjectively more 'recalcitrant' than others, to borrow Quine's term for experience, and thus more resistant to revision, while others we readily abandon or alter if we discover a conflict in our set of beliefs. For example, we are frequently more willing to surrender a memorial belief about some event in our past, than a belief about what is occurring at the present moment. But these recalcitrant beliefs are not to be identified with the indubitable or self-evident beliefs at the base of some foundationalist epistemologies, for no matter how recalcitrant a belief appears to be, it always remains revisable. Its truth is not established by our strong commitment to it, and, of course, our commitment to any belief is always subject to change.

---

[79] Ibid., §1.
[80] Ibid., 42, §6.
[81] Ibid.
[82] Smart (1990) 7.
[83] Quine (1953) 42, 'Two Dogmas of Empiricism,' §6.

Hence, we do not prove that a belief is true; rather, we find that the evidence—other beliefs that we hold—disposes us to accept the belief in question. Some of these supporting beliefs, as I have suggested, are more subjectively compelling than others, and thus we have more confidence in the beliefs based upon them. According to Pierre Duhem, 'the validity of [one's] conclusion is [only] as great as the validity of [one's] confidence' in the truth of the other beliefs that support that conclusion.[84] Our beliefs are based on our faith or confidence in their truth or validity, not on whether or not they actually are true.

But why do we find some beliefs compelling, and why are we disposed to have faith in the truth of certain beliefs? The short answer to this question must be, I think, that a person accepts a belief because he is the sort of person who is persuaded by such a belief. As Peirce writes, 'that which determines us, from given premises, to draw one inference rather than another, is some habit of mind, whether it be constitutional or acquired.'[85] James agrees that each of us has an 'individual way of just seeing and feeling the total push and pressure of the cosmos,'[86] which he denominates a person's 'temperament.' James is less concerned to explain why we have the particular temperament that we do, than to convince us that this temperament, or mindset, influences our individual response to the arguments and hypotheses proposed to our consideration. James has perceptively characterized the whole 'history of philosophy [as] to a great extent that of a certain clash of human temperaments,' even though philosophers are wont to deny and 'sink the fact of [their] temperament.'[87] A person's temperament 'loads the evidence for him one way or the other' because each of us '*trusts* his temperament,' and supposes that his peculiar way of 'see[ing] things' is the correct way, while being 'dissatisfied with any opposite way of seeing them.'[88]

What Peirce and James are suggesting here, is that the experiences and principles, from which we derive our theories and views of reality, do not by themselves determine

[84] Duhem (1954) 185.
[85] Peirce (1955) 'The Fixation of Belief,' 8.
[86] James (2000) 7, *Pragmatism*, 'The Present Dilemma in Philosophy.'
[87] Ibid., 8.
[88] Ibid., 9.

how we shall interpret and incorporate those experiences and principles. There is an ineliminable element of subjectivity in any interpretation, and in our willingness even to consider a possible belief. In 'The Will to Believe,' James distinguishes between 'live' and 'dead' hypotheses. 'A live hypothesis is one which appeals as a real possibility to him to whom it is proposed,' while a dead hypothesis 'refuses to scintillate with any credibility at all.'[89] To put it another way, whether or not we trust that a belief is true, or are willing even to treat it as a possibility, depends upon whether or not, with respect to the belief in question, we have, as it were, a trusting nature. We evaluate a belief the same way we judge the character of someone we meet, in 'the simple adjectives of praise or dispraise.' With any proffered belief, we 'measure [it] against the total character of the universe as we feel it.'[90]

Incidentally, some philosophers, while being receptive to the idea that beliefs are shaped by temperament, have attempted to explain why we have the particular temperaments that we do. John Dewey, for instance, seems to have been one of the first to conclude that, since human beings are products of evolution by natural selection, our traits, including the ways we think and feel, must be 'honed by evolutionary considerations.'[91] According to this conception of 'evolutionary epistemology,'[92] as it has come to be called, we can 'account for the characteristics of cognitive mechanisms in animals and humans by a straightforward extension of the biological theory of evolution to those aspects or traits of animals that are the biological substrates of cognitive activity,' such as their brains.[93] We are disposed to accept certain beliefs, and to find certain hypotheses plausible, because, in evolutionary terms, the possession of such cognitive dispositions has aided our survival and reproduction. On this view, our temperaments are biological adaptations.[94]

---

[89] James (2000) 199, 'The Will to Believe,' Sect. I.
[90] James (2000) 21, *Pragmatism*, 'The Present Dilemma in Philosophy.'
[91] Bradie (1994) 453.
[92] For an excellent collection of articles discussing issues in evolutionary epistemology, see Radnitzky and Bartley (1987).
[93] Ibid., 454.
[94] The claim that our cognitive abilities are evolutionary adaptations has been developed into separate arguments for realism and for the reliability of our sense perceptions. First, without the supposition of an external

But whatever the source or cause of our temperaments, if I am correct that beliefs are distinguished merely by the strength of our commitment to them, then many of the traditional distinctions between beliefs must be abandoned, including any absolute distinction between scientific and non-scientific knowledge claims, or between the claims of science and those of religion or metaphysics. The popular distinction between the scientific method on one hand, and metaphysical speculation on the other, finally collapses, because there is no justified test for establishing which of the claims generated by science or metaphysics is correct. And if we are unable to distinguish beliefs based upon truth, then any distinction that we impose upon our beliefs must be founded upon our personal, social, and pragmatic interests. Not all beliefs are equally appealing to us.

Nonetheless, many scientists and philosophers have been convinced, as A. F. Chalmers writes in his popular book on the nature of science, 'that there is something special about science and its methods. The naming of some claim or line of reasoning or piece of research "scientific" is done in a way that is intended to imply some kind of merit or special kind of reliability.'[95] For instance, historian of science Roger French has characterized modern science as 'objective' and 'non-religious'; and '[i]n being objective, passionless, crea-

---

world, the remarkable agreement between human perceptions, and our common beliefs about the world, must remain 'an inexplicable miracle, a piece of "pre-established harmony."' Bartley (1987) 39. There must be a common world to explain our common perceptions. And second, if our cognitive abilities are in fact evolutionary adaptations, then they must be at least approximately true, because false beliefs and unreliable faculties are maladaptive. Our success as a species can best be explained by the supposition of the truth of our beliefs.

These arguments are vulnerable on a number of points. First, it seems unlikely that the realist could prove the existence of the other observers, and banish the possibility of solipsism, in order for his argument for realism to go through. Second, it is inconceivable that the realist could discover the origin of our ideas of the world, whether they proceed, for example, from Berkeley's God or from real physical objects, simply from the fact of our common experiences of the world. Third, the argument for the reliability of our senses and beliefs is circular, in that it must rely upon the truthfulness of those senses and beliefs in order to establish the facts about evolution. And fourth, it does not follow from the fact that our cognitive faculties are adaptations that they must be true. The most that natural selection seems to require is that our beliefs be true in the instrumentalist sense that true beliefs are those that lead to successful predictions about those aspects of the world that interest us. It is our ability to predict future events, to have some inkling of what to expect in terms of experience, that increases our survival and reproductive fitness. But whether or not our beliefs also happen to correspond with ultimate reality and to unobservable entities, matters not a whit to our biological success.

Perhaps an appropriate last word on the topic is Peirce's: 'If investigation cannot be regarded as proving that there are Real things, it at least does not lead to a contrary conclusion . . . and not having any doubt, nor believing that anybody else whom I could influence has, it would be the merest babble for me to say more about it.' Peirce (1955) 18-19.

[95] Chalmers (1982) xv.

torless, [science] alone produces tangible truth, which in modern society is given privileged status.'[96] But Chalmers is correct, I think, to reject this view of science, because '[t]here is just no method that enables scientific theories to be proven true or even probably true.'[97]

Aside from arguments 'based on a detailed analysis of the history of science and modern scientific theories,'[98] the chief argument for the 'claim that scientific theories cannot be conclusively proved or disproved,'[99] as Duhem and Quine have shown, is that every scientific claim is embedded in a vast network of mutually supporting, interlocking beliefs, none of which has individually been proved to be true. In order to verify or falsify a scientific belief, or any belief for that matter, the belief must be evaluated, as Quine writes, 'in isolation from its fellows.'[100] But if Quine is correct in his suggestion, borrowed from Duhem, that 'our statements about the external world face the tribunal of sense experience not individually but only as a corporate body,'[101] then there is no way to subject individual beliefs to critical evaluation. Chalmers puts the point simply: 'Observation statements must be made in the language of some theory, however vague.'[102] There are no naked observations or theories; they all depend on other beliefs that we hold, which other beliefs in turn depend on still others. It would be an interesting and useful enterprise, which must impress this point admirably, if some prodigious polymath should examine a well-established scientific claim, such as the age of the earth, or the distance to a neighboring galaxy, and attempt to extract from that claim *all* the assumptions and scientific theories upon which it rests, including basic metaphysical presuppositions about the world.

The logical positivists' verificationism, and Karl Popper's falsificationism, must be severely tempered, if not abandoned, in light of this interdependence of scientific claims. Such theories of science suppose, as Duhem writes, 'that each one of the hypotheses em-

---

[96] French (1994) xi.
[97] Chalmers (1982) xvi.
[98] Ibid., xvi-xvii.
[99] Ibid., xvi.
[100] Quine (1953) 41, 'Two Dogmas of Empiricism,' §5.
[101] Ibid.
[102] Chalmers (1982) 28.

ployed in [science] can be taken in isolation, checked by experiment, then when many varied tests have established its validity, given a definitive place in the system of [science].' But science, he continues, 'is not a machine which lets itself be taken apart; we cannot try each piece in isolation . . . [Rather p]hysical science is a system that must be taken as a whole.' [103]

Consequently, one can never subject an isolated hypothesis to experimental test. Instead, we test a scientific belief by showing its compatibility or incompatibility with the 'corporate body' of other beliefs that we hold. As Lipton has written, 'the route from theory to observation is long and often obscure. The only way to tie the theory deductively to the observational consequences will be with an elaborate and purpose-built set of auxiliary statements,…approximations, idealizations, and *ceteris paribus* clauses needed to make observational contact.'[104]

In order to verify or falsify a hypothesis, according to the traditional understanding of the scientific method, we test the implications of that hypothesis by means of a well-designed experiment. But deriving a testable prediction of a hypothesis, as Lipton has suggested, involves the use of a great many auxiliary assumptions, in each of which the scientist will have various degrees of confidence. Therefore, the failure of a hypothesis to produce the expected experimental results does not show, notwithstanding Popper's well-known work, that the hypothesis is false. According to Duhem, 'if the predicted phenomenon is not produced, not only is the proposition questioned at fault, but so is the whole theoretical scaffolding' from which the prediction derives; 'the only thing the experiment teaches us is that among the propositions used to predict the phenomenon and to establish whether it would be produced, there is at least one error; but where this error lies is just what it does not tell us.'[105] The hypothesis itself might be false, but for all we know the error may lie in any of the host of auxiliary assumptions that are logically connected with the hypothesis.

[103] Duhem (1954) 187.
[104] Lipton (1991) 147.
[105] Duhem (1954) 185.

So how are we to flush out the error? As Quine recognizes, 'the total field [of beliefs] is so underdetermined . . . that there is much latitude of choice as to what statements to reëvaluate in the light of any single contrary [experiment].'[106] We have much freedom of choice in selecting which beliefs to retain, and which to dispense with; in making the choice, according to Duhem, we must finally 'leave to our sagacity the burden of *guessing*.'[107] The importance of this point can not be overstated, as an antidote to any uncritical belief in the uniqueness of the scientific method: the searcher for truth must ineluctably bear 'the burden of guessing' which of his beliefs is true. The acceptance of any belief 'implies in general an *act of faith* in a whole group of theories.'[108] And this analysis does not change, whether the belief derives from scientist's laboratory, or from the metaphysician's airy speculations.

Thus, the attempt by many scientists and philosophers, including Karl Popper, to distinguish science from metaphysics is ultimately fruitless, for, as Quine recognizes, 'in point of epistemological footing the physical objects and the gods differ only in degree and not in kind.'[109] We posit the existence of such entities because we find, individually and culturally, that the world simply makes more sense to us when it is organized in certain ways. As James would have it, each of us '*trusts* his temperament . . . [and] believes in any representation of the universe that . . . suit[s] it.'[110] Our acts of epistemological faith, our commitments to the truth of particular beliefs, are expressions of that temperament.

This view of 'science [as] a continuation of common sense'[111] rightly denies any fundamental distinction between the scientific method of pursuing truth, and that engaged in by the metaphysicians and the theologians. For Quine, the theoretical posits of science and metaphysics are merely 'myths' that differ only 'in the degree to which they expedite our

[106] Quine (1953) 42, 'Two Dogmas of Empiricism,' §6.
[107] Duhem (1954) 211, emphasis added.
[108] Ibid., 238, emphasis added.
[109] Quine (1953) 44, 'Two Dogmas of Empiricism,' §6.
[110] James (2000) 9, *Pragmatism*, 'The Present Dilemma in Philosophy,' emphasis in the original.
[111] Quine (1953) 45, 'Two Dogmas of Empiricism,' §6.

dealings with sense experiences.'[112] This is a pragmatic view, which sees 'science as a tool, ultimately, for predicting future experience in the light of past experience.'[113] From this instrumentalist view, truth as correspondence with objective reality simply drops out, and in James' hands, the concept of truth itself is redefined as what '*we can assimilate, validate, corroborate and verify*.'[114] However, I see no reason to follow James in this redefinition, nor to suppose that we are indifferent to whether our beliefs make epistemological contact with objective reality.[115]

A desire for the successful prediction of future experiences is but *one* of our motives for investigating reality. There is also, I maintain, the desire that our beliefs should be objectively true, irrespective of their instrumental value, and the correspondence of those beliefs with sense experience is *one* factor in fortifying our confidence in their truth. That sense experience brings us into relation with a mind-independent world is, for most of us, I think, one of our firmest articles of epistemological *faith*, and consequently, we are frequently disposed to evaluate other hypotheses in terms of their coherence with sense experience. But because this belief in the truthfulness of sense experience does not itself logically follow from sense experience, it is a mistake to suppose that sense experience provides the only grounds for evaluating our beliefs. Other principles, which philosophers are wont to call *a priori*, also influence our choice of beliefs. Thus, I agree with Quine that '[o]ntological questions,

[112] Ibid.
[113] Ibid., 44.
[114] James (2000) 89, *Pragmatism*, 'Pragmatism's Conception of Truth,' emphasis in the original.
[115] While I agree with the instrumentalist position that a theory may be useful, and generate empirically adequate predictions, without being objectively true, I can not agree with those who insist that we dispense with the objective conception of truth. If beliefs are commitments, then one may be committed to the *truth* of a theory as well as to its empirical adequacy. The instrumentalist is correct, of course, to point out that the truth of a theory goes beyond the empirical evidence; and one is certainly free, and within one's epistemological rights, so to speak, to refrain from taking that step. But to require that everyone else likewise forbear to believe anything beyond what the instrumentalist finds acceptable, is to insist dogmatically that others share his commitments, and his aversions to the risk of being mistaken. We choose our own commitments, and take our own epistemological chances.

[which depend for their answers on such *a priori* principles,] are on a par with questions of natural science.'[116]

The Reformation of a System of Beliefs

We have seen that a system of beliefs may be confronted with a problem, as when a scientific theory generates a false prediction, or when one's metaphysical worldview comes into conflict with some principle that one finds compelling. In such cases, it is necessary to have some means of reforming one's system of beliefs, to remove the conflict, to account for the false prediction. In this section, I want to consider some of the evaluative criteria that we use in determining what we shall believe.

I begin by observing that most of us are rather selective about the hypotheses we regard even as *candidates* for incorporation into our systems of beliefs. As we have seen, James makes this same point by way of a distinction between live and dead hypotheses, between, that is, a hypothesis that 'appeals [to us] as a real possibility,' and one that enjoys no 'credibility at all.'[117] The 'deadness and liveness in an hypothesis are not intrinsic properties, but relations to the individual thinker.'[118] Accordingly, a person may change in relation to a proposed belief, such that a belief that formerly made no 'electric connection' with him, may come to seem among the most obvious truths in the world. A belief is more at home, as it were, in some belief systems than others, and thus, a person whose belief system changes over time may become more receptive to a belief that he previously could not accept. As Lipton writes, in deciding what to believe, 'we often start from a group of plausible candidates, and then consider which of these is the best, rather than selecting directly from the vast pool of possible [beliefs].'[119] Thus, we employ two 'epistemic filter[s:] . . . one that

[116] Quine (1953) 45, 'Two Dogmas of Empiricism,' §6.
[117] James (2000) 199, 'The Will to Believe,' I.
[118] Ibid.
[119] Lipton (1991) 61.

selects the plausible candidates, and a second that selects from among them.'[120] These plausible candidates, Lipton agrees,[121] are tantamount to James' live hypotheses.

But most of these plausible, live hypotheses turning out not to be such 'happy guesses,'[122] they conflict with other beliefs that one holds, thus requiring some decision about how to eliminate the conflict. This can be accomplished in a variety of ways. One can reject the proposed hypothesis, one can abandon a belief that one has hitherto held, one can propose another belief to resolve the discrepancy, or one can accept both beliefs as true, trusting that their apparent incompatibility will eventually be removed.

Just as Plato tries to understand morality in the individual by considering how it manifests itself in the *polis*, Thomas Kuhn can be read as describing this process of reforming one's belief system by considering how scientific communities have reformed their theories throughout history. For Kuhn, the process of science has a distinctly revolutionary character. The practice of what he calls 'normal science' 'is predicated on the assumption that the scientific community knows what the world is like.'[123] In normal science, scientists share basic theoretical assumptions and experimental techniques for testing hypotheses and evaluating results; this collection of shared assumptions, the dominant 'paradigm,' is transmitted to students of the discipline, and by and large goes unexamined and unquestioned.

But the practice of normal science inevitably encounters difficulties and experimental results that seem to contradict its theories, as when 'a normal problem, one that ought to be solvable by known rules and procedures, resists the reiterated onslaught of the ablest members of the group within whose competence it falls.'[124] This is equivalent to the conflicts we find within the belief systems of an individual, for whom a sense experience or a compelling explanatory hypothesis may seem to falsify some other belief that he holds. And the periods

[120] Ibid.
[121] Ibid.
[122] Ibid., 89, quoting Carl Hempel, 1966, *The Philosophy of Natural Science* (Englewood Cliffs, NJ: Prentice Hall), 15.
[123] Kuhn (1970) 5.
[124] Ibid.

of normal science, when scientists resist changes to their general presuppositions, correspond to our individual tendencies, as Peirce puts it, to 'cling tenaciously, not merely to believing, but to believing just what we do believe.'[125]

These theoretical conflicts and difficulties giving rise to dissatisfactions with the paradigm, eventually culminate in a scientific 'revolution,' when the background commitments of the paradigm are exchanged for different, incompatible assumptions. At some point, scientists and philosophers simply lose patience with the inability of their theories to solve problems. Kuhn could give no final explanation for why these revolutions occur, since, as Duhem recognizes, 'logic does not determine with strict precision the time when an inadequate hypothesis should give way to a more fruitful assumption.'[126] But 'clearly [our basic] commitments may evolve.'[127] For example, Kuhn points out that Copernicus initiated a revolution in astronomy despite the fact that '[no] fundamental astronomical discovery, no new sort of astronomical observation, [had] persuaded [him] of ancient astronomy's inadequacy or of the necessity for change.'[128] But, as Duhem argues, it is precisely 'by resolutely carrying out a reform among the propositions declared untouchable by common consent [that one] will accomplish the work of a genius who opens a new career for a theory.'[129]

For Kuhn the new paradigm is incommensurate with the old, since it is the foundational assumptions that are overturned and assailed by the new paradigm. The scientific community rejects 'one time-honored scientific theory in favor of another incompatible with it.'[130] Duhem agrees: there is the 'possibility of lengthy quarrels between the adherents of

---

[125] Peirce (1955) 10.
[126] Duhem (1954) 218.
[127] Kinoshita (1990) 306.
[128] Kuhn (1985) 132.
[129] Duhem (1954) 211.
[130] Kuhn (1970) 6.

an old system and the partisans of a new doctrine, each camp claiming to have good sense on its side, each party finding the reasons of the adversary inadequate.'[131]

However, as I have suggested, human beings seem to be constitutionally loath to part with their beliefs. Even during revolutions we tend to retain much. As Smart observes, '[w]e can recognize Kuhnian revolution without jettisoning much of the cumulative or "Whig" conception of scientific advance.'[132] In his earlier work, *The Copernican Revolution*, Kuhn himself allows that '[e]ach new scientific theory preserves a hard core of the knowledge provided by its predecessor and adds to it.'[133]

This notion that we tend to be conservative in resisting change to our beliefs has been recognized by numerous thinkers. Quine writes that 'our natural tendency [is] to disturb the total system as little as possible.'[134] Duhem adds that when confronted with the choice of which of conflicting principles to abandon, we choose the 'alternative . . . [that] does not make us lose anything of the [scientific] terrain already conquered.'[135]

But, in the end, I think, there are simply no absolute guidelines to determine how to reform a system of beliefs in conflict, since any such guidelines would require of us what we can not accomplish; namely, identifying which of the conflicting principles is more accurate. Perhaps we can do no more than appeal to Duhem's hand-waving principle of 'good sense': 'The day arrives when good sense comes out so clearly in favor of one of the two sides that the other side gives up the struggle even though pure logic would not forbid its continuation.'[136] Likewise, in certain situations, James preaches the 'lawfulness of voluntarily adopted faith.'[137] And Smart concludes that '[h]ow beliefs should be accepted, rejected, strengthened or weakened is a matter for the dynamics of belief.'[138]

---

131 Duhem (1954) 217.
132 Smart (1990) 10.
133 Kuhn (1985) 3.
134 Quine (1953) 44, 'Two Dogmas of Empiricism,' §6.
135 Duhem (1954) 211.
136 Ibid., 218.
137 James (2000) 'The Will to Believe,' I, 198.
138 Smart (1990) 8.

But the study of epistemology is not in fact hopeless, because, it turns out, our epistemic *practices* are not irremediably subjective. There are common and wide-spread ways of evaluating beliefs to which most of us, in our daily lives, faithfully subscribe. A number of accounts have been proposed to explain this agreement about epistemological criteria. According to philosopher of religion Richard Swinburne, '[w]e come to our investigations of the causes of phenomena equipped with an *a priori* apparatus of general understanding of what is evidence for what, and within that of what is evidence for this being the explanation of that.'[139] If we did not possess such *a priori* principles, Swinburne argues, we should never be able to discover an explanation, since 'we need to know what an explanation looks like before detailed *a posteriori* investigation will reveal when we have got one.'[140] This is similar to Plato's argument in the *Phaedo*[141] that one must possess the innate idea of 'the equal,' in order to recognize that particular objects are equal, for otherwise the experience of equal things should never alert one to their equality.

For Kuhn, there must be 'field-specific principles that determine the actual judgments' that scientists make, in order to explain exactly why the scientists working within a particular paradigm 'are in broad agreement about which problems to work on, how to attack them, and what counts as solving them.'[142] These principles being passed on to young scientists through their education, influence what they regard as good, normative science. And for the evolutionary epistemologists, to give one last illustration, a rough agreement among the fundamental epistemological principles that most of us accept is precisely what we should expect to find, because such principles are nothing more than successful rules for surviving in the world, and individuals who do not share them tend to have lower rates of survival. There is agreement because the dissenters have been selectively eliminated.

[139] Swinburne (1990) 182.
[140] Ibid.
[141] 74-75b.
[142] Lipton (1991) 7.

But however we come to possess our evaluative criteria, whether they are innate or learned, whether they are the products of evolution or of society, we can never be sure that they pick out true beliefs, or even that a belief that satisfies them is more likely to be true than one that does not. For since '[w]e cannot directly compare hypothesis with fact,' Smart writes, '[t]he best we can do is to choose the best warranted hypothesis . . . in the *hope* that warranted assertability is a pointer towards truth.'[143] Unfortunately, these 'criteria for "best" do not include "true",' so Smart identifies the 'best' hypothesis with 'the one which survives' in an individual's or a community's continually modified 'web of beliefs.'[144] Lipton agrees that because our inferences about the world are 'underdetermined by the evidence . . . we must use additional rules or principles' to reach a proper evaluation of our beliefs.[145]

But what are these rules or principles that we use to evaluate our beliefs? Lipton calls them 'aesthetic considerations,'[146] in order to distinguish the characteristics of a belief that cause it to be appealing to us, from those characteristics that render it true, or likely to be true. Although we can not identify the properties of a belief that point towards truth, what we can say is that certain beliefs are congenial to our ways of thinking.

Kinoshita lists a number of 'familiar criteria [of acceptability,] such as testability, universality, simplicity, observational adequacy, consistency, predictive power, fruitfulness, and so on.'[147] Others have compiled different lists. Alfred North Whitehead, for example, requires that a speculative 'philosophical scheme should be coherent, logical, and, in respect to its interpretation, applicable and adequate,'[148] while Edward O. Wilson finds that consil-

---

[143] Smart (1990) 10, emphasis added.
[144] Ibid.
[145] Lipton (1991) 8.
[146] Ibid., 68.
[147] Kinoshita (1990) 309.
[148] Whitehead (1981) 191-192.

ience, a concept borrowed from William Whewell, is 'the key to unification [of knowledge],' and a test of a theory's truth to boot.[149]

But in addition to identifying our evaluative criteria, it is important to 'specify how each is to be weighed, individually and relative to one another.'[150] For instance, a theory may yield successful predictions and cohere roughly with observation, but yet strike one as unacceptably 'inelegant.' This point is clearly illustrated by the Copernican system vis-à-vis the ancient Ptolemaic theory of the geocentric universe. Although the 'full [Copernican] system was little if any less cumbersome than Ptolemy's had been[, b]oth employ[ing] over thirty circles . . . [n]or could the two systems be distinguished by their accuracy,'[151] Copernicus' system had 'aesthetic advantages'[152] over Ptolemy's, including its ability to explain the planets' retrograde motions without appeal to epicycles. However, the Copernican system also predicted that 'each of the stars should seem slightly to change its position [relative to the observer] . . . during the course of a year.'[153] But this stellar parallax could not be observed with the instruments available at the time. Nevertheless, Copernicus' system was preferred over Ptolemy's by 'those astronomers who valued qualitative neatness far more than quantitative accuracy.'[154] That is, they had a temperamental preference for, and thus gave greater weight to, the virtues of simplicity and elegance, over predictive power and observational accuracy.

So our choice of beliefs is based on our choice of evaluative criteria, and the weights we give to those criteria. But if I hope to make a persuasive case for this central theme of these pages, that all of our beliefs rest ultimately upon our temperamental preferences, and thus can not be justified, in the sense of being shown to be true, I must trade 'glittering generalities' for a specific examination of a few of these evaluative criteria.

---

[149] Wilson (1998) 8.
[150] Kinoshita (1990) 309-310.
[151] Kuhn (1985) 169.
[152] Ibid., 172.
[153] Ibid., 163.
[154] Ibid., 172.

Logical consistency

Some there may be who affect on occasion Walt Whitman's extravagant nonchalance towards contradiction: 'Do I contradict myself? / Very well then I contradict myself, / (I am large, I contain multitudes.)'[155] Others may venture on occasion the Emersonian opinion that '[a] foolish consistency is the hobgoblin of little minds, adored by little statesmen and philosophers and divines. With consistency a great mind has simply nothing to do.'[156] But it is not clear that anyone is genuinely prepared to disclaim the need for logical consistency in his thinking.

However, Leibniz surely overstates matters when he concludes that the principle of 'contradiction, in virtue of which we judge false that which involves a contradiction,'[157] is one of the two foundations of our reasoning.[158] Better to acknowledge that logical consistency is *one of many* principles that we use in evaluating hypotheses, because most of us recognize that we often accept beliefs that we can not square with one another, and if our earlier skeptical arguments are cogent, then the conclusions of those arguments apply with equal force to the validity or truthfulness of the principles of logic.

Having identified no certain criteria of truthfulness, we are woefully unable to satisfy the skeptic's importunate demands to be told how anyone can be sure that the principles of logic are conducive to truth, or, what is the same thing, how we can know that truth must be logical. 'How are we to know, in a given case,' writes Bertrand Russell, 'that our belief is not erroneous?' And he rightly concludes that '[t]his is a question of the very greatest difficulty, to which no completely satisfactory answer is possible.'[159]

---

[155] Walt Whitman, *Leaves of Grass*, 'Song of Myself,' §51.
[156] Emerson (1992) 138, 'Self Reliance.'
[157] Leibniz, *Monadology*, §31.
[158] For Leibniz, the other foundation of reasoning is 'that of sufficient reason, in virtue of which we hold that there can be no fact real or existing . . . unless there be a sufficient reason, why it should be so and not otherwise.' Ibid., §32.
[159] Russell (1971) 119.

Logical consistency does not seem to be a necessary presupposition of thought, nor an indispensable requirement of a satisfactory theory or worldview. But consistency *is* a virtue. By this I mean, and mean only, that most of us highly value consistency among beliefs.

Parsimony

Parsimony or simplicity is typically invoked to help distinguish between empirically and explanatorily equivalent theories. 'The usual philosophical picture,' according to Elliott Sober, concerns 'several hypotheses that are each consistent with the evidence, or explain it equally well . . . Parsimony is then invoked as a further consideration'[160] to guide the choice between them. Philosophers commonly point out that of parsimony is important to attempts to fit a curve to a finite number of data points on a graph. The points 'underdetermine the curve, since there are many curves that would pass through those points,'[161] but the principle of parsimony tells us that, in choosing between curves, 'smooth curves are better than bumpy ones.'[162]

However, 'parsimony, in and of itself, cannot make one hypothesis more plausible than another.'[163] We have no epistemic assurance that a simpler theory has a greater likelihood of truth than a less simple one, so it is rather surprising to find a number of influential philosophers, including epistemologist and philosopher of religion Richard Swinburne, concluding that '[a]mong theories able equally well to yield the data . . . the simplest is that most likely to be true.'[164] The principle is far from self-evident, and it is not clear at any event that we are ever, except in contrived over-simplified scenarios, confronted with two theories that 'equally well . . . yield the data.' The 'actual difficulty is often to come up with even one theory that fits the observed facts.'[165] And for that matter, '[w]hy should *simple*

[160] Sober (1990a) 84-85.
[161] Lipton (1991) 6.
[162] Sober (1990a) 90.
[163] Ibid., 77.
[164] Swinburne (1990) 188.
[165] Lipton (1991) 15.

worlds . . . have any greater likelihood than infinitely complex ones?'[166] And surely it would be reckless to assume, as a methodological principle, that we just happen to inhabit a world best described by simple principles.

One suspects that there is more than a little ambiguity in this use of 'simple.' The principle of parsimony does not operate in isolation from other methodological considerations. In practice, the scientist or philosopher is never faced with two otherwise empirically equal theories, which he must choose between solely in terms of their relative simplicity. The simplest theory is not always, and perhaps not even usually, preferable. 'In many fields, at particular moments in their histories, scientists quite cheerfully postulate new entities in order to account for new empirical discoveries rather than making other theoretical accommodations equally compatible with the data in question.'[167]

This suggests that theoretical decisions reflect an individual's beliefs about a theory's plausibility, in light of his theoretical and methodological commitments, rather than simply a conclusion regarding the theory's simplicity. 'Thus, judgments of simplicity and parsimony are . . . theory-dependent,'[168] and '[o]nly because of a set of background assumptions does parsimony connect with plausibility in a particular research problem.'[169]

In some contexts, and I mean within certain frameworks of assumptions, one may regard a theory as more plausible than any of the available alternatives, though it contains more entities, coincidences, axioms, or mathematical complexity than they do. While in other contexts, the more reasonable theory may also enjoy the advantage of being simpler than the rejected theories. It is the collection of one's other beliefs that determines which of the proposed theories to accept, and not some methodological maxim that 'simpler is better,' however one wishes to define simpler. My sense, then, is that the adjective 'parsimonious' typically functions rather as a term of approval, than as a criterion of truth. Identifying

[166] Clarke (1990) 202.
[167] Boyd (1995) 374.
[168] Ibid.
[169] Sober (1990a) 77.

a theory as parsimonious is little more than a way of expressing one's satisfaction with that theory. Likewise, 'extravagant,' 'speculative,' and 'inelegant' describe theories that one finds implausible, because those theories are either incompatible with one's stock of beliefs, or inadequately supported by them.

Coherence

Coherence comes in two varieties. It is both a theory of justification, and a theory of truth. [170] It is with the first of these that I am concerned here, the use of coherence as a means of justifying one's beliefs. Coherence certainly seems to be one of our most important reasons for believing that a proposition is true—in the correspondence sense of truth, as I shall hereafter understand the term—even though we have no means of justifying our faith in coherence as a sign of truth. In the main, we think that a theory is true because it is supported by, or coheres with, other beliefs that we think are true, even though there may be some basic beliefs that we hold for no other reason than that we are naturally or temperamentally disposed to find them compelling. For example, Descartes observes in his *Meditations* that we all have by nature 'a very great inclination to believe that [sense data] are conveyed to [us] by corporeal objects.'[171] The truth of such principles, as the existence of the

[170] In the coherence theory of truth, 'truth is simply *identified* with coherence.' BonJour (1985) 88. The coherence of a set of beliefs is not merely an indicator of their truth, as it is in the coherence theory of justification, but rather it is what we *mean* by truth, since truth is redefined as coherence itself. This view is opposed to the correspondence theory of truth, according to which 'truth consists in some form of correspondence between belief and fact.' Russell (1971) 121.

There are at least two virtues of the theory. First, it avoids the intractable problem of trying to establish an epistemological link or relation between what we believe and what we assume to exist apart from the mind. Second, it provides a means of overcoming skepticism, by severing truth from any relation with the inaccessible external world. On this view, the only relations that matter for truth are those internal relations that obtain between beliefs, and that are directly accessible to introspection.

But this view of truth is also vulnerable to a couple of criticisms, which undermines its plausibility for me. First, since 'there is no reason to suppose that only *one* coherent body of beliefs is possible' (Russell (1971) 122), the coherence theory of truth implies the possibility of a multitude of incompatible 'true' theories, and thus leads relentlessly to relativism. Second, this view of truth fails to accomplish what most of us want a theory to accomplish, namely, to describe the world as it really, objectively is. But however we define truth, we may still properly ask whether the coherence of a set of beliefs gives us reason to suppose that those beliefs correctly describe and correspond with reality.

[171] Descartes (1970) 191, *Meditations*, VI.

external world, 'is evident to us, [although] . . . they themselves, or at least some of them, are incapable of demonstration.'[172]

Coherence can not justify our beliefs, nor show that they are true, or even likely to be true, so it is well to regard coherence not as a theory of justification at all, as it is usually presented, but as a criterion of theory selection. We choose to believe a theory or proposition because it is coherent with our other beliefs, while our conviction that coherence itself is truth conducive is merely an article of faith. BonJour concludes, in his work on coherence, 'that something like coherence is indispensable to any nonskeptical epistemological position which is even *prima facie* adequate.'[173] Since even most forms of epistemological foundationalism 'employ the notion of coherence in their total epistemological accounts,'[174] the practical difference between foundationalism and coherence concerns the extent to which someone is willing to hold beliefs based on intuition, rather than on the epistemic support provided by other beliefs.

But coherence, for all the esteem in which it is held as an evaluative criteria, 'seems to be subject to a number of crushing objections.'[175] It is instructive to bear these in mind, as a necessary corrective to the prevalent claim that the coherence of a set of beliefs is a sign of the truthfulness of those beliefs.

The first difficulty with coherence as an evaluative criterion is the problem of specifying the nature of coherence. What does it mean for a set of beliefs to be coherent? Beyond the 'intuitive[recognition that] coherence is a matter of how well a body of beliefs "hangs together,"' based on 'the various sorts of inferential, evidential, and explanatory relations which obtain among the various members of a system of beliefs,' [176] '[i]t turns out to be very

[172] Russell (1971) 112.
[173] BonJour (1985) 94.
[174] Ibid.
[175] Ibid., 25.
[176] Ibid., 93.

difficult to explain what coherence is.'[177] However, it is possible to distinguish coherence from some other evaluative criteria, with which it is frequently identified.

Coherence is not mere logical consistency. 'It is true that consistency is one requirement for coherence, that inconsistency is obviously a very serious sort of incoherence.'[178] But the absence of contradiction among beliefs is not a sufficient condition of their coherence, since logically consistent beliefs may show very little in the way of coherence. For example, the beliefs that Fulton produced the first workable steamboat, and that light from the sun takes some eight minutes to reach the earth, are logically consistent, but there is no appreciable coherence between them, in the sense of mutual epistemic support.

And coherence is not equivalent with explanation. Explanatory relations are one significant type of relation that may hold between beliefs in a coherent system, but there are others types of coherent relations as well, including evidentiary and inferential relations. For instance, a criminal suspect's blood-stained trousers may provide *evidence* that the suspect committed the crime, without constituting an *explanation* of the crime. Additionally, one may form a belief concerning the height of a tree, by *inferring* it from one's other beliefs about the length of its shadow and principles of geometry, without thereby *explaining* the height of the tree. However, there being a close connection between coherence and explanation, the existence of unexplained anomalies within a set of beliefs certainly 'detracts from the coherence of the system.'[179]

The second problem with coherence is that, even if we accept this minimal conception of coherence, we can not use it as a criterion to evaluate prospective theories unless we have some measure of the extent to which a theory coheres with one's other beliefs, since almost any belief will exhibit *some* coherence with a given body of beliefs. How coherent is a particular body of beliefs? Unfortunately, a non-arbitrary means of determining the amount of coherence in a system of beliefs does not seem to be at hand.

---

[177] Audi (2003) 198.
[178] BonJour (1985) 95.
[179] Ibid., 99.

And third, supposing, however, that we could measure the coherence of a set of beliefs, we have, nonetheless, no standard of how much coherence is rationally required for one to assent to a belief. Is it ever rational, for example, for one to accept a belief that reduces the overall coherence of his system of beliefs? Similarly, is it ever rational for a scientist to maintain belief in a theory that is currently incapable of explaining some observation or experimental result? Presumably the answer to these questions is yes. As Kuhn has observed, '[t]here are always some discrepancies'[180] between theory and fact. Likewise, Larry Laudan has 'conclud[ed] from the existing historical evidence that . . . [t]heories are generally not rejected simply because they have anomalies,'[181] and Paul Feyerabend writes that 'there is not a single interesting theory that agrees with all the known facts in its domain.'[182] Even Popper, the arch falsificationist, occasionally allows that 'a limited amount of dogmatism is necessary for progress: without a serious struggle for survival in which the old theories are tenaciously defended, none of the competing theories can show their mettle.'[183]

A belief may cohere strongly with a subset of a person's beliefs, while being only weakly supported by another subset. But a theory and an anomaly may both be so well established in the scientist's mind, that the scientist is willing to accept the incompatibility, with the faith that the problem will be 'set right sooner or later . . . We therefore have to ask what it is that makes an anomaly seem worth [accepting], and to that question there is probably no fully general answer.'[184]

How long may one defend dogmatically an embattled theory, before such tenacity becomes irrational? When is it reasonable to adopt a new theory that, despite significant defects, seems to offer the promise of solving some old problems? How much support does a belief require from other beliefs, in order to make it rational to accept it? It seems clear that Kuhn is correct, that there are no fully acceptable answers to such questions.

---

[180] Kuhn (1970) 81.
[181] Laudan (1992) 144.
[182] Feyerabend (2001) 21.
[183] Popper (1992) 98.
[184] Kuhn (1970) 82.

But this is just what we should expect. Beliefs are commitments to the truth of propositions, which go beyond the evidence supporting those propositions. Indeed, as Kuhn acknowledges, a revolution in thought is nothing more than a 'special sort of change involving a certain sort of reconstruction of group commitments.'[185] Such beliefs rest inevitably upon our faith and trust that they are correct. A scientist who chooses to embrace a new paradigm or theory, because he 'feel[s] that the new proposal is on the right track,' does so with the 'faith that the new paradigm will succeed with the many large problems that confront it, knowing only that the older paradigm has failed with a few. A decision of that kind can only be made on faith.'[186] Therefore, our willingness to adopt a new belief, or to hold onto an old one in the face of conflicting evidence, depends largely upon the character, temperament, and prejudices of the believer. And this is the case whether the believer is an individual philosopher or a community of scientists. Kuhn has noted, for example, that those who advance revolutionary ideas in science are 'usually . . . so young or so new to the crisis-ridden field that practice has committed them less deeply than most of their contemporaries to the world view and rules determined by the old paradigm.'[187] Likewise, Charles Darwin 'look[ed] with confidence to the future, to the young and rising naturalists,'[188] as those most likely to be persuaded by his arguments for natural selection.

We can conclude, then, that coherence is a value—like simplicity, consistency, and elegance—that *informs* an individual's decision about whether or not to accept a theory, without *determining* that decision. Each of us weighs the evidence for himself, and makes a determination about the adequacy of that evidence. 'There is no neutral algorithm for theory-choice, no systematic decision procedure which, properly applied, must lead each individual in [a] group to the same decision.'[189] The extent to which a theory coheres with other

[185] Ibid., 181.
[186] Ibid., 158.
[187] Ibid., 144.
[188] Darwin (1998) 389.
[189] Kuhn (1970) 200.

beliefs is merely one factor that a person takes into consideration in deciding whether to accept the theory.

We can instructively liken this epistemological view to the ethical theory of W. D. Ross.[190] For Ross, we have a number of presumptive, *prima facie* moral duties, of which he lists seven: fidelity, reparation, gratitude, justice, beneficence, self-improvement, and non-malificence. Accordingly, we have duties to repay our debts, to be faithful to our promises, to help other people, to improve ourselves, and so forth. Our *actual* duty in a particular situation is the *prima facie* duty that is relevant and controlling in the situation, and when conflict inevitably arises between our *prima* facie duties, the moral agent must simply choose, as it were, according to his best lights, which duty shall govern. No objective principle specifies what our moral duty is in any given situation. Ross seems to suppose that the *actual* duty will be intuitively obvious to a normal, educated, morally-sensitive individual.

Now, in epistemological matters, as we have seen, each of us acknowledges a number of epistemic criteria, or values, which typically include elegance, consistency, coherence, simplicity, and predictive power. Faced with the choice of whether or not to accept a hypothesis, we weigh the hypothesis in the scales of these values, as it were, in order to assess its acceptability. However, the individual application of these criteria may yield different assessments of acceptability. A proposed theory might be elegant, without having much predictive power, or coherent without being simple.

As with decisions about our ethical duties, there is, as we have said, no general rule that determines how to evaluate the evidence for and against a particular hypothesis. Epistemic values, like *prima facie* duties, may pull in different directions. Even what counts as relevant evidence is open to dispute. There are a variety of factors that we take into account in deciding whether to believe a hypothesis. But different people apply these factors differently, and even the same person will not apply them in the same way in every instance, nor

[190] Ross (1930).

always accord them the same weight. Following James, we may give the general name of 'temperament' to this individual way of '*see*[*ing*] things . . . in [one's] own peculiar way.'[191]

Of course, in recognizing that each of us has his own susceptibility to certain kinds of hypotheses, and in giving that susceptibility or credulity the name of temperament, we do not thereby explain our willingness to believe. But it is nonetheless important to appreciate the significance played by temperament in a person's decision to accept a given hypothesis, because failure to recognize the rôle of temperament in one's own belief formation often leads to the charge that someone else is being irrational in refusing to adopt that same belief. For example, Elliott Sober writes, in his introductory *Philosophy of Biology*, 'that people behave unscientifically when they refuse to consider relevant evidence.'[192] This claim seems to assume that we have some independent standard for determining what counts as relevant evidence in a particular controversy. But this assumption is not credible, in light of our failure to identify any such standard, and given the way that individuals, including scientists, actually form their beliefs.

For example, the fierce debate between Galileo and his contemporaries, including Tycho Brahe, concerning the truth of the Copernican doctrine, is probably best not regarded as a refusal by either side to consider relevant evidence. The evidence that one considers relevant to resolving a given dispute depends largely on the other beliefs that one holds. Thus, both the Copernicans and their opponents evaluated the sun-centered hypothesis by means of beliefs and values that they found credible. '[B]ased on the facts, the theories and the standards of the time,' Feyerabend writes, the idea championed by Galileo that the earth moved was 'absurd.'[193] But given a dissatisfaction and frustration with the Ptolemaic system, as a theory that could no longer stand under the burden of its observational inadequacies and ad hoc inventions,[194] it seemed reasonable to Copernicus 'to try whether, by assum-

[191] James (2000) 9, *Pragmatism*, 'The Present Dilemma in Philosophy,' emphasis in the original.
[192] Sober (1993) 28.
[193] Feyerabend (2001) 129.
[194] Kuhn (1985) 139.

ing some motion of the Earth, sounder explanations than [Ptolemy's system offered] for the revolution of the celestial spheres might be so discovered.'[195] In its favor, the Copernican system provided a simple and elegant explanation for why Mercury and Venus are always observed near the Sun, and why the 'normal [eastward] motion of all planets . . . is occasionally interrupted by brief intervals of westward or "retrograde" motion.'[196] Thus, each side found compelling evidence to support its position.

The same analysis would apply to other scientific revolutions,[197] such as the debate over Alfred Wegener's theory of continental drift in the first half of the 20th century, or the contention over the Darwinian mechanism for evolution in the last decades of the 19th century. Darwin himself writes that '[a]ny one whose disposition leads him to attach more weight to unexplained difficulties than to the explanation of a certain number of facts will certainly reject my theory [of natural selection].'[198] In evaluating a hypothesis, of course, one should consider the evidence that bears on that hypothesis, but differences in beliefs are frequently best explained simply by noting that people attach different weights to the evidence that they consider relevant. In the continuing debate between the creationists and the evolutionists, for example, each side is, more or less, cognizant of the 'evidence' cited by the other camp, but finds it insufficient to overturn its own explanatory theory. Each side obviously attaches much greater significance to the deficiencies in the other position, than do proponents of that position. Whereas creationists, for instance, are apt to speak of 'gaps' in the fossil record as evidence that falsifies evolution by natural selection, and the current lack of a credible theory of abiogenesis as a critical deficiency in the origin of life controversy, evolutionists are unlikely to be particularly troubled by these criticisms. As Kuhn writes, 'every problem that normal science sees as a puzzle can be seen, from another viewpoint, as

[195] Ibid., 142, quoting the Preface to Copernicus' *De Revolutionibus* (1543), trans. John F. Dobson and Selig Brodetsky.
[196] Kuhn (1985) 47.
[197] Cohen (1985) provides an excellent and detailed discussion of scientific revolutions.
[198] Darwin (1998) 389.

a counterinstance and thus as a source of crisis.'[199] Thus, the accusation that one has not considered relevant evidence, like the charge of irrationality, is typically nothing more than an expression of disagreement with one's interpretation of the evidence.

In this discussion of coherence, we have considered but a few of the problems that beset the use of coherence as an evaluative criteria. Coherence is difficult to define and to quantify, and as only one of our evaluative criteria, it is unclear how coherence should be related to these other criteria in assessing the acceptability of a hypothesis. It seems that we frequently balance these criteria against one another, such that the strong presence in a hypothesis of some epistemic virtues can compensate for a relative lack of others. But how this balancing is to be carried out, within the dictates of reason, we probably can not hope to resolve, without resort to arbitrary rules, which would, in any event, certainly fail to describe the way that people actually evaluate hypotheses. I have suggested that this evaluative process is similar to ethical deliberations, in which we attempt to discern our moral duty in a particular situation, given the facts of the situation as we know them, and the moral principles and *prima facie* duties that we recognize. There are no algorithms that lead to a 'correct' result. Peter Lipton summarizes the situation well in the preface to his book *Inference to the Best Explanation*. '[I]t is amazingly difficult,' he writes, 'to give a principled description of the way we weigh evidence. We may be very good at doing it, but we are miserable at describing how it is done, even in broad outline.'[200] Now let us turn to the subject of Lipton's book, the claim that 'explanation [i]s a key'[201] to how we evaluate hypotheses.

Explanation

Explanation seems to be important to rational inquiry about the world. Probably most philosophers and scientists would agree with the view of Philip Gasper that '[e]xplanation is an important goal of scientific inquiry.'[202] According to Carl Hempel and Paul Op-

[199] Kuhn (1970) 79.
[200] Lipton (1991) ix.
[201] Ibid.
[202] Gasper (1990) 289.

penheim, it 'is one of the foremost objectives of [any] . . . scientific research . . . to go beyond a mere description of its subject matter by providing an explanation of the phenomena it investigates.'[203] Smart argues that inference to the best explanation is 'the core notion of epistemology.'[204] For Richard Swinburne, '[t]he many, complex and coincidental cry out for explanation.'[205]

But unfortunately, as we have seen in our previous discussions of the other evaluative criteria, 'there exists considerable difference of opinion as to the function and the essential characteristics of scientific explanation.'[206] Thus, even if we agree with Sober that '[h]ypotheses are accepted, at least partly, in virtue of their ability to explain,'[207] we need some account of what it means for a hypothesis to explain, and how that explanatory power contributes to our acceptance of the hypothesis.

A number of different accounts of explanation have been offered. Some philosophers have suggested, for example, that an explanation is nothing more than 'find[ing] things well grounded in our understanding,'[208] such that the world does not surprise us. On this account, an explanation is a pragmatically useful instrument for getting round in the world, and avoiding unexpected shocks. Other philosophers, such as Philip Kitcher, have argued that explanation amounts to a unification of knowledge. This is a reductive view, which explains a phenomenon by deriving it from more comprehensive phenomena, thereby 'reduc[ing], in so far as possible, the number of types of facts we must accept as brute.'[209] Still others have asserted that '[s]cientific explanations explain by telling us of an entity what it is.'[210] In this group, we must include Aristotle, for whom an explanation of something states the causes that are at work in it. Since 'men do not think they know a

---

[203] Hempel and Oppenheim (1953) 319.
[204] Smart (1990) 1.
[205] Swinburne (1990) 189.
[206] Hempel and Oppenheim (1953) 319.
[207] Sober (1990a) 73.
[208] Clark (1990) 206.
[209] Kinoshita (1990) 298, quoting Kitcher, P., 1981, 'Explanatory Unification,' *Philosophy of Science*, 48, 507-531.
[210] Kinoshita (1990) 304.

thing till they have grasped the "why" of it (which is to grasp its primary cause),'[211] Aristotle provides the 'why' of a thing in term of his four causes: the material, formal, efficient, and final causes.

But two accounts of explanation have received the most attention, and it will render our discussion more manageable if we limit it to these two. The basic distinction between these two accounts of explanation concerns whether we should 'regard the notion of causation as essential to scientific explanation.'[212] As Nancy Cartwright observes, 'in explaining a phenomenon one can cite the causes of that phenomenon; or one can set the phenomenon in a general theoretical framework.'[213] The second of these accounts, with which we begin our discussion, seeks to replace 'causal facts with facts about . . . the laws and equations of high level scientific theories.'[214]

The covering-law, or deductive-nomological, model of explanation, which Hempel did much to promote, explains an event by showing that it is entailed by a combination of antecedent conditions and general laws. By way of illustration, Hempel asks us to consider a mercury thermometer plunged quickly into hot water. The column of mercury drops temporarily, but then rises quickly. 'How is this phenomenon to be explained?'[215] Hempel refrains from offering an explanation in terms of causes. Instead, he explains the event in terms of established regularities, or laws. Thus, the heat reaching at first only the glass tube, the glass expands, providing a larger space for the mercury, whose surface level drops; but when the heat reaches the mercury, by conduction through the glass, the mercury itself expands more than the glass, and its surface rises accordingly.

This explanation, according to Hempel, consists of two types of statements: antecedent conditions and general laws. And these 'two sets of statements, if adequately and completely formulated, explain the phenomenon under consideration: They entail the con-

[211] Aristotle, *Physics*, Bk. II, Ch. 3, 194b20, trans. by R. P. Hardie and R. K. Gaye.
[212] Smart (1990) 3.
[213] Cartwright (1995) 379.
[214] Ibid.
[215] Hempel and Oppenheim (1953) 320.

sequence that the mercury will first drop, then rise.'[216] Thus, the general laws, or established natural regularities, concerning 'the thermic expansion of mercury and of glass, and . . . the small thermic conductivity of glass,'[217] together with the fact that a glass tube partly filled with mercury was immersed into hot water, entail that the mercury will behave as observed. This subsuming an event under general laws is the essence of explanation.

A request for an explanation of an event is therefore answered with a description of the laws and conditions that entail the occurrence of that event. The fall of a stone released from one's hand near the surface of the earth is explained by deducing the fall from conditions that existed at the time of release, plus Newton's laws of motion and gravitation. It does not simplify Hempel's account unduly to say that the descent of the stone is thus explained by saying that this is what stones and other massy objects do in these circumstances.

There are at least three justifications for an account of explanation that eschews any appeal to causation, and relies instead on the application of general laws. First, 'worries about causation go back at least as far as Hume.'[218] According to Hume, experience does not furnish us with direct apprehension of any necessary causal relation between cause and effect. We are never able 'to discover any power or necessary connexion[,] any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other.'[219] Experience reveals that certain events, such as the release of a stone, are generally followed by certain other events, such as the descent of the stone towards the earth. If there exists a necessary connection between the events, that connection is not revealed to experience, which sees only the sequence of events.

Consequently, if we translate our ordinary talk of causation into the Humean account of causation, 'we arrive at the [Hempelian] account [of explanation]: an event of a

[216] Ibid.
[217] Ibid.
[218] Gasper (1990) 286.
[219] Hume (1975) §7, Part 1.

certain kind is explained by citing a general law (or laws) that relates events of that kind to events or conditions of some other kind.'[220] If a cause is understood as nothing more than a lawful generalization, then explaining why something occurs, giving a causal explanation of the event, amounts to providing the lawful generalization.

Second, this account of explanation is thought to be more precise than other such accounts, and this precision 'is essential if we are to have objective criteria for assessing when something has been explained.'[221] In our discussion of coherence, we noted some of the difficulties inherent in the use of that criterion to evaluate hypotheses. One of these was the problem of determining how much coherence is required to render a hypothesis rationally acceptable. We suggested that there is no clear, non-subjective resolution to the problem. Nevertheless, we certainly 'want to be able to tell accurately when . . . a new theory has sufficient explanatory resources to be accepted or preferred over its competitors.'[222] The Hempelian account of explanation promises just such objective precision in hypothesis evaluation. In comparing two hypotheses, the objectively preferable hypothesis is the one that is more successful at explaining phenomena, where success is measured in terms of the ability of the hypothesis to yield accurate predictions.

The third justification of this account of explanation is that it avoids appeal to unobservable theoretical entities. A causal explanation of an event must postulate the existence of unobservable entities or 'occult powers' that cause the event; but the covering-law model of explanation depending only on observed regularities in the world can presumably dispense with any unverifiable talk of unobservables. Newton famously captured the spirit of this suspicion regarding unobservable processes in his 'hypotheses non fingo.' The causal model being highly speculative, it must therefore be regarded as inferior to a model of explanation that is justified by direct observation.

[220] Gasper (1990) 286.
[221] Gasper (1995) 289.
[222] Ibid.

But despite these ostensible justifications, the model has been subjected to a number of withering criticisms. One class of criticisms 'show[s] that fitting the model is not necessary for something to be an explanation.'[223] Not all explanations are predictions, so it is 'a mistake to suppose that all scientific explanations . . . aspire to deductive-nomological status.'[224]

As Smart observes, it 'is well known [that] Hempel's model suffers from counter-examples.'[225] Explanations in evolutionary biology can provide such counter-examples. In explaining why a species of organism has a particular trait, biologists give a Darwinian explanation. They attempt to show how the possession of the trait in question contributes to the fitness of the organism in its environment. 'An organism's fitness is its ability to survive and reproduce, which is represented in terms of probabilities.'[226] Organisms that possess the trait enjoy an edge in survival and reproduction over their conspecifics that do not possess it. Thus, we explain the presence of the trait by referring to the advantage it confers upon its possessor. As Hempel recognizes, 'for the phenomenon of mimicry, the explanation is sometimes offered that it serves the purpose of protecting the animals endowed with it from detection by its pursuers.'[227] But, he admits, 'the fact that a given species of butterflies displays a particular kind of coloring cannot be inferred from . . . the statement that this type of coloring has the effect of protecting the butterflies from detection by pursuing birds.'[228] The increase in fitness that derives from the possession of a trait does not deductively entail that an organism will possess the trait. Thus, Darwinian explanations in terms of natural selection are not explanations at all under the deductive-nomological model of explanation. Consequently, we must deny either that natural selection can explain features of the biological world, or that the Hempelian model comprehends all legitimate explanations.

---

[223] Gasper (1990) 287.
[224] Lipton (1991) 51.
[225] Smart (1990) 13.
[226] Sober (1993) 36.
[227] Hempel and Oppenheim (1953) 329.
[228] Ibid.

A second class of counter-examples to the covering-law model includes those that purport to 'show that fitting the cover-law model is not sufficient for something to be an explanation.'[229] I mentioned one such counter-example previously, in distinguishing coherence from explanation. We can deduce the height of a tree from the length of its shadow, the position of the sun, and the laws of geometry, but 'we would not want to say that the length of the shadow *explains* the height of the [tree].'[230] The 'classic case,' writes van Fraassen, 'is the *barometer example*.'[231] From the falling of the barometer we infer the coming of the storm, though we do not explain the change in the weather. Similarly, Edwin Hubble deduced the expansion of the universe from observations of the red-shifted spectral lines of distant galaxies, together with the laws of optics, the Doppler effect, and so forth, but we would not suggest that Hubble had thereby explained why the universe expands.

A third class of objections challenges the explanatory power of laws. For Hempel, we provide an explanation of an event when we can deduce that event from conditions antecedent to the event plus general laws. But, as philosopher Stephen Clark asks, '[h]ow is the particular conjunction of A and B "explained" by saying that there is a discoverable set of A-things conjoined with B-things?'[232] Answering the question why this conjunction occurs by saying only that it always does 'is not an explanation, but a refusal to take puzzles seriously at all.'[233] Clark cites approvingly a passage from Wittgenstein's *Tractatus* that '[t]he whole modern world-view is based on the illusionary deception that the so-called laws of nature offer explanations of natural phenomena.'[234] And Philip Gasper suggests that the covering-law model may ultimately be tautological. If the occurrence of an event is explained as nothing more than a specific instance of a general pattern, then A follows B be-

---

[229] Gasper (1990) 287.
[230] Ibid., emphasis added; Smart (1990) 14, gives a similar example.
[231] van Fraassen (1995a) 319.
[232] Clark (1990) 196.
[233] Ibid.
[234] Ibid., 196-197, quoting L. Wittgenstein, 1961, *Tractatus Logico-Philosophicus*, trans. by D. F. Pears and B. F. McGuinness (London: Routledge & Kegan Paul), 6.371.

cause A's follow B's, which is simply a logical tautology. But, of course, 'tautologies are not explanatory.'[235]

There are other problems as well. There is the difficulty of defining what counts as a general law. Hempel acknowledged the 'difficulty . . . of stating rigorous criteria for the distinction between the permissible and the non-permissible' predicates in our general laws.[236] And Gasper worries 'whether it is possible to specify which general statements count as genuine laws . . . without reference to unobservables or to causal factors.'[237] Another problem with the covering-law model concerns the supposed sharp distinction between observable and non-observable entities. An event is explained if it could be deduced from general laws and antecedent conditions. The model requires that the deduced consequences be 'capable, at least in principle, of test by experiment or observation.'[238] But if, as seems likely, '[o]bservation statements [themselves] are always made in the language of some theory and will be as precise as the theoretical or conceptual framework that they utilize is precise,'[239] then the test of the theory is not a straightforward observation at all, as the model requires, but a whole subset of one's theoretical beliefs, which go far beyond the antecedent conditions and general laws.

'Problems of the kind just mentioned have persuaded many philosophers that the covering-law model of explanation is fundamentally flawed.'[240] The causal model of explanation, on the other hand, according to which 'we explain phenomena [either] by giving their causes or . . . by giving a mechanism linking cause and effect,'[241] purports to avoid the problems besetting the covering-law model. As Swinburne writes, in order 'to explain the occurrence of some event or process we cite the causal factor which, we assert, brought it

[235] Gasper (1995) 293.
[236] Hempel and Oppenheim (1953) 342.
[237] Gasper (1990) 287.
[238] Hempel and Oppenheim (1953) 321-22.
[239] Chalmers (1982) 29.
[240] Gasper (1995) 293.
[241] Lipton (1990) 247.

about or keeps it in being.'[242] This is the 'commonsense view of what an explanation is.'[243] For example, to 'explain why smoking causes cancer, . . . [we] give information about the causal mechanism' that links smoking and cancer.[244]

This approach seems to enjoy some advantages over the previous model. First, causal 'explanations do not have to meet the requirements of [the covering-law] model, because one need not give a law to give a cause, and one need not know a law to have good reason to believe that a cause is a cause.'[245] Second, since causality is uni-directional, proceeding from cause to effect but not vice-versa, we do not confront the problem exemplified in the example of the tree and its shadow. Again, the problem for Hempel's model is that one can deduce either the height of the tree from the length of its shadow, or the length of the shadow from the height of the tree, but it seems mistaken to suppose that the length of the shadow somehow 'explains' the height of the tree. But under the causal model, 'the reason [the tree] explains the [the shadow] but not conversely is that causes explain effects and not conversely.'[246]

Third, the causal model admits Darwinian explanations in accounting for biological adaptations. The covering-law model fails to explain such adaptations, because the usefulness of a trait does not entail that an organism will possess it, but we can explain such adaptations on the causal model. The possession of an adaptive trait causes an increase in the organism's fitness, its ability to survive and reproduce, and gives the organism an advantage over the other members of its species that do not possess the trait. Natural selection favoring organisms with the trait, those without it tend to be eliminated, leaving the favorable trait, so to speak, in possession of the field.

Fourth, the causal model of explanation may be more consonant with actual scientific practice than the covering law model. For Cartwright, 'given the way modern theories

[242] Swinburne (1990) 177.
[243] Gasper (1995) 289.
[244] Lipton (1991) 32.
[245] Ibid., 32-33.
[246] Ibid., 33.

of mathematical physics work, it makes sense only to believe their causal claims and not their explanatory laws.'[247] In her view, mathematical laws are treated by physicists as heuristic calculating devices, whereas causal explanations are regarded as either true or false descriptions of the natural world. If Cartwright is correct that scientists do not regard mathematical treatments of a phenomenon as providing explanations of that phenomenon, then the claim by Hempel to have uncovered the 'basic pattern of scientific explanation'[248] must be doubted.

Finally, the causal model does not depend upon any dubious distinction between observable and theoretical entities. Hempel adopts a model based on laws in order to eliminate reference to speculative unobservable entities, but observation is itself inherently theory-laden and interpretive. The causal account of explanation, however, acknowledges that a belief in causes does not differ in kind from the other beliefs that one holds. Such 'causal' beliefs are accepted according to the same evaluative criteria as our other beliefs.

Nevertheless, despite these advantages over the covering-law model, the causal model faces significant difficulties of its own. First, there is a question whether causal powers themselves are explanatory. Against the Hempel model, Stephen Clark argues that it does not explain an event to say that events of that kind always occur under similar circumstances, but he likewise 'doubt[s whether] the situation is improved by transferring our attention to causal powers.'[249] There seem to be two possible sources of his doubt. There is the question whether it is meaningful and explanatory to speak of causal *powers* at all. Chalmers observes that the 'majority of philosophers seem reluctant to accept an ontology which includes dispositions or powers as primitive.'[250] But, like Chalmers, 'I do not understand their reluctance.'[251] The other possible source of concern is that causal powers may not be *ultimately* explanatory, because the most basic causal powers can not themselves be

[247] Cartwright (1995) 379.
[248] Hempel and Oppenheim (1953) 319.
[249] Clark (1990) 197.
[250] Chalmers (1999) 219.
[251] Ibid.

explained in terms of anything else. Not all brute facts can be explained. The most fundamental properties and powers of matter may simply be given. But if these fundamental causal powers remain inexplicable, then it might seem that anything explained by means of them is not itself truly explained. However, I see no reason to disagree with Lipton's assessment that 'we can know a phenomenon's cause without knowing the cause of the cause.'[252] We uncover the explanations that are available to us.

'The second and perhaps the most obvious objection to the causal model of explanation is that there are non-causal explanations.'[253] For example, a 'mathematician may explain why Godel's Theorem is true, and a philosopher may explain why there can be no inductive justification of induction, [without] cit[ing] causes.'[254] Moreover, we may be interested in an explanation of what something is, rather than an explanation of what caused it. For example, early chemists, such as Lavoisier, sought to explain the nature of heat. 'Their question was not what caused heat, but rather what sort of entity it was; they were not sure if it was a substance, an effect, a process, or an activity.'[255] A similar question confronts the contemporary philosophers and scientists who attempt to explain the mind. Is the mind an epiphenomenon of the brain, a separate substance, or merely the functional activity of interacting neurons? Some philosophers, called psychological eliminativists, have even denied that there is a phenomenon to be explained.[256]

This suggests what we should perhaps have suspected all along. Explanation does not name a homogeneous group of activities; it is not a single thing. Sometimes we are indeed interested in what something is, rather than its cause. Other times we seek the cause of the phenomena. And still other times, as Hempel recognizes, the statement of a law satisfies our desire for explanation. '[S]cientists are often and perhaps primarily interested in ex-

[252] Lipton (1990) 247.
[253] Lipton (1991) 33.
[254] Ibid.
[255] Kinoshita (1990) 307.
[256] E.g., Paul Churchland has argued for the 'elimination' of the mind, as a remnant of 'folk psychology,' and Daniel Dennett sometimes writes as if he, too, is sympathetic to that position.

plaining regularities, rather than particular events,'[257] so subsuming the regularity under a general law may satisfy their desire for an explanation. Thus, in seeking an account of explanation, we must recognize 'the important but unsurprising point that different people are interested in explaining different phenomena,'[258] and that they have different requirements about what will count as an explanation.

Before concluding this discussion of explanation, something should be said about the realist argument that 'the best explanation of [a] theory's success is that the (observable and unobservable) entities, mechanisms and events which it postulates, actually exist (or closely resemble what actually exists).'[259] Charles Darwin himself defended his theory of natural selection by means of such an argument: 'I cannot believe,' he writes, 'that a false theory would explain, as it seems to me that the theory of natural selection does explain, the several large classes of facts above specified.'[260] According to this type of argument, called Inference to the Best Explanation, a theory's truth is taken to be the best explanation of its explanatory and predictive successes. Another possible explanation for a theory's 'predictive success is that the predictions that have been checked just happen to be some of the true consequences of a false theory . . . [but this] is just to say that the predictive success is a fluke.'[261] It would be 'miraculous [if] the theory should have these explanatory successes, yet not have something importantly true about it.'[262]

Inference to the best explanation is an example of what Peirce calls an abduction. An abduction is a preference for a hypothesis, prior to any testing, based on its ability to explain the facts. Thus, 'an explanatory hypothesis—which is just what abduction is— . . .

[257] Lipton (1990) 264.
[258] Ibid., 265.
[259] Gasper (1990) 294.
[260] Darwin (1998) 388.
[261] Lipton (1991) 159.
[262] Ibid., 184.

cannot be admitted, even as a hypothesis, unless it be supposed that it would account for the facts or some of them.'[263] Accordingly, abduction has this form:

> The surprising fact, C, is observed;
> But if A were true, C would be a matter of course,
> Hence, there is reason to suspect that A is true.[264]

In the debate between realists and anti-realists concerning the existence of non-observable theoretical entities, the realist hangs much weight upon an argument to the best explanation. He supposes that 'a realist construal of scientific theories is required to give an adequate explanation of the experimental and observational testing of hypotheses.'[265] For example, '[t]hose who take laws seriously tend to . . . assume that the fact that a law *explains* provides evidence that the law is true.'[266]

But we can 'infer the truth of an explanation only if there are no alternatives that account in an equally satisfactory way for the phenomena.'[267] Thus, the realist must address Duhem's provocative question: 'Shall we ever dare to assert that no other hypothesis is imaginable?'[268] The history of science is replete with instances of well-grounded theories that were upset when some scientist imagined a more credible alternative. For example, the existence of a subtle and undetected 'luminiferous ether' was once posited as the best explanation for the propagation of light waves through space, the ether providing a medium for their transmission, as the atmosphere does for compression waves. The German physicist Heinrich Hertz expressed his 'certain[ty] that all space known to us is not empty, but is filled with a substance, the ether, which can be thrown into vibration.'[269] And even several years after the failure of the famous Michelson-Morley experiment to detect any movement of the Earth relative to the ether, Hermann von Helmholtz could write, '[t]here can no longer be

---

[263] Peirce (1955) 151.
[264] Ibid.
[265] Gasper (1995) 296.
[266] Cartwright (1995) 380.
[267] Ibid.
[268] Duhem (1954) 190.
[269] Jaki (1966) 82, quoting H. Hertz, 1889, 'On the Relations between Light and Electricity,' in *The Miscellaneous Papers of Heinrich Hertz*, trans. D. E. Jones (London, 1896), 314.

any doubt that light waves consist of electrical vibrations in the all pervading ether.'[270] That explanation is not accepted anymore. Consequently, the claim that some theory is true, or that some theoretical entity is real, because that theory or entity is the best explanation of some phenomena, must be regarded rather as an article of faith, than a species of proof.

Lipton argues, moreover, 'that [this] truth argument has no force for the non-realist,'[271] since it begs the question against him. According to a non-realist, such as van Fraassen, 'acceptance of a theory may properly involve something less (or other) than belief that it is true.'[272] One may accept a theory because it is empirically adequate, without believing that it is literally true; 'theories need not be true to be good.'[273] The realist, however, tries 'to argue from a theory's past observational successes [i.e., its empirical adequacy] to its truth.'[274] But this assumption that empirical adequacy is 'truth-tropic'[275] is precisely what the anti-realist denies. Consequently, in order to avoid begging the question against the anti-realist, the realist needs to provide a justification of abduction; he needs to establish, and not just presuppose, that the fact that a theory would explain if true, is reason to think that it *is* true. But the justification of abduction is fraught with the same problems as the justification of induction. The prospect of solving these problems is dim indeed.

But although the realist argument 'is almost entirely without probative force,'[276] the intuition remains that there is something correct about it. Lipton captures that intuition: 'If I were a scientist, and my theory explained extensive and varied evidence, and there was no alternative explanation that was nearly as lovely, I would find it irresistible to infer that my theory was approximately true.'[277] 'Explanatory power,' writes van Fraassen, 'is something

---

[270] Jaki (1966) 83, quoting the Introduction by H. von Helmholtz to H. Hertz, *Principles of Mechanics*, trans. by D. E. Jones and J. T. Walley (London, 1899), xvi.
[271] Lipton (1991) 161.
[272] van Fraassen (1995b) 187.
[273] Ibid.
[274] Lipton (1991) 160-161.
[275] Ibid., 161.
[276] Ibid., 184.
[277] Ibid.

we value and desire.'[278] It is one of the factors, along with consistency, coherence, simplicity, elegance, and so forth, that we take into account in deciding whether to accept a theory. Van Fraassen usefully reminds us that such acceptance need not include the belief that the theory is true; we may accept a theory because, given our purposes, it is good enough.

Conclusion

It may be useful, in a few lines, to weave together a few of the individual threads that run through this long chapter. First of all, we must distinguish between a belief and its justification. The criteria we apply to determine whether a belief is acceptable are different from the test to determine whether that belief is true. I have argued, with respect to the test of the truthfulness of a hypothesis, that we have on hand no definitive test, nor the prospect of acquiring one. The skeptical challenge finally can not be met, because there is no non-circular means of validating and justifying our beliefs. And I have argued that the process we go through in assessing the acceptability of a belief does not follow any definite rules.

There are a number of factors or epistemic virtues that we balance in evaluating a proposed hypothesis. I looked at a few of these criteria, including consistency, simplicity, coherence, and explanation, and concluded that none of them is necessary or sufficient for acceptance of a belief. There are no definite rules that determine how to apply these criteria, nor how to weigh them when there is a conflict. As Darwin wrote in the introduction to his *Origin of Species*, in evaluating a hypothesis '[a] fair result can be obtained only by fully stating and balancing the facts and arguments on both sides of each question.'[279] This balancing, which is common to all disciplines, including science, theology, philosophy, and history, is unavoidably subjective and idiosyncratic. I find much truth in Feyerabend's insight that '[t]he comparison [of hypotheses] and the final decision [between them has] much in common with the comparison of life in different countries (weather, character of people, melodiousness of language, food . . . etc.) and the final decision to take a job and to start life

[278] van Fraassen (1995a) 326.
[279] Darwin (1998) 4.

in one of them.'[280] In making such a decision, one evaluates the alternatives in terms of factors that one finds important, and according to the weights that one places on those values. Both of these are unavoidably personal choices, because decision making, like hypothesis evaluation, involves personal commitment to the result.

In the next chapter, I shall attempt to state and balance, according to my best lights, some of the arguments for and against a program of reductive materialism. Consequently, I shall, as Peirce writes, 'suggest certain ideas and certain reasons for holding them true; but then, if you accept them, it must be because you like my reasons, and the responsibility lies with you.'[281]

[280] Feyerabend (2001) 259.
[281] Peirce (1955) 3.

CHAPTER 2

ONTOLOGICAL REDUCTIONISM

What strange and curious metaphysical systems have been imagined! And how much more wonderful if one of them should be true! What all of these systems share, besides their quaintness, is an attempt to formulate an answer to the question that Quine considers central to metaphysics: What is there?[1] The subject of the present work is an extended treatment of one such system, a picture of reality presented by contemporary science, which I shall denominate, for reasons that will become clear, reductive materialism.[2]

For the materialist, simply put, reality consists of matter in motion. Thus, as Marjorie Grene writes, 'whatever one truly asserts about the world expresses, in the last analysis, some change in the configurations of matter described by physics.'[3] Sometimes this view is called physicalism, rather than materialism, presumably because physicalism is thought to be more inclusive than materialism, and the person who describes himself as a physicalist wishes to include in his scientific worldview more than just matter in motion. Indeed, J. J. C. Smart, who regards 'physicalism [a]s a better word' for his view than materialism, includes within that view all 'the entities which need to be postulated by physics.'[4] Among the physical things that he regards as necessary for physics, 'even though they are not material things,'[5] are numbers, sets, assertions, Occam's razor, and Platonic Forms. Al-

[1] Quine (1953) 1, 'On What There Is.'
[2] It may be helpful to bear in mind that my purpose in this chapter is not to offer a critique of materialism, but rather to challenge claims that the materialist is necessarily committed to reductionism.
[3] Grene (1971) 16.
[4] Smart (1978) 382.
[5] Ibid.

though Smart 'confesses' that he is not '100 per cent happy' about the need to posit such abstract objects, he can not conceive how they are to be avoided.[6]

Since I am not concerned in this work with the ontological status of such abstract objects, I shall not follow Smart and others in using the more inclusive, and less well defined, term 'physicalism' to describe the worldview of contemporary science. According to the materialist worldview, as I shall understand it, 'there are no concrete existents, or substances, in the spacetime world other than material particles and their aggregates;'[7] real things are material. Reductive materialism, however, goes beyond the basic materialist claim that everything is made of matter.[8] It includes additional assertions, to which not all materialists subscribe. The reductionist adds the claim that the science of 'physics is causally [and explanatorily] complete (i.e., all fundamental causal forces are physical forces, and the laws of physics are never violated).'[9] For the reductive materialist, everything can be explained in terms of the interactions of material particles and their properties. Therefore, reductive materialists are united in their opposition to the doctrine of vitalism,[10] which, despite various vague formulations, implies that 'there is some mysterious something about living things exempt from physico-chemical laws,'[11] or that 'living things contain some principle of a quite peculiar nature and quite different from anything that is found in matter that is not living.'[12]

The reductive materialist view encompasses ontological, causal, and explanatory claims.[13] Let us take each of these in turn. First, an ontological inventory of real things would include the basic material particles, their properties, and the fundamental interactions between those particles. As physics professor Roger Faber puts it in his book on mechanis-

[6] Ibid., 384.
[7] Kim (1996) 211.
[8] Trout (1995) 387.
[9] Horgan (1993) 560.
[10] Rosenberg (1986) 73 regards this as a doctrine 'that no responsible biologist could seriously credit.'
[11] Grene (1971) 15.
[12] Singer (1959) 37.
[13] Owens (1989) 59.

tic reduction, the reductive materialist believes that 'the full truth about anything whatever resides in stories about the particles of physics';[14] and in the words of theoretical physicist Paul Davies, '[t]he physical world is populated by material objects that occupy locations in space and have qualities like extension, mass, electric charge and so on. These objects are not inert, but move about, change and evolve in accordance with [the] dynamical laws . . . of physics.'[15] Non-material entities, such as immortal souls, would therefore be excluded. Likewise, the traditional notion of the mind as a mental substance that thinks and wills and actuates the body, like the pilot of a vessel, has to be abandoned; materialism, writes Davies, 'denies the existence of mind altogether,' and reduces all 'mental states and operations . . . to physical states and operations.'[16]

For 'contemporary physicalists,' the familiar objects of experience, such as stars and stones and owls, are nothing more than 'aggregates' (or collections) of the 'basic particles recognized in physics.'[17] These 'ordinary things . . . acquire a shadowy ontological status because, according to the atomist program, parts are specifiably *more real* than the wholes they compose.'[18] But since it is inconvenient and impracticable for us to describe our world in term of interacting elementary particles, which 'alone are allowed to count as "real,"'[19] 'we adopt, when possible, the expedient of attaching names to collections of large numbers of atoms, treating these groups . . . as honorary individuals.'[20] Accordingly, our classification of the world into objects and kinds is arbitrary, and uninformative about reality.

Consider an illustration. Oceanographers are accustomed to partition the ocean in various ways, according to their interests. They may be interested in the difference between

[14] Faber (1986) 7.
[15] Davies (1983) 73.
[16] Ibid., 82.
[17] Kim (1996) 227. Kim appears to use 'aggregate' as a general synonym of 'collection,' and does not distinguish aggregates, or non-chemical mixtures of discrete objects, from compounds, or 'distinct substance[s] formed by chemical union of two or more ingredients in definite proportion by weight.' *Merriam-Webster's Collegiate Dictionary*, 2003, 11th ed. (Springfield, Mass.), 255.
[18] Faber (1986) 4, emphasis added.
[19] Grene (1971) 24.
[20] Faber (1986) 25.

marine habitats close to shore, and those in the open ocean, and therefore divide the ocean into zones to reflect this difference. The open ocean, which they call the pellagic zone, is further divided into the waters above the continental shelf, and the waters beyond it. The ocean can also be divided into strata according to depth, or according to the amount of light that can penetrate into each. But from the standpoint of reality, all of these divisions, and the many others that oceanographers employ, are all equally artificial, for if the water molecules are regarded as the only real things, then there is no *real* difference between any of these scientific divisions, because they are all alike in being composed of the same stuff.

Similarly, the reductive materialist insists that any division of the material world into objects and kinds, like the division of the ocean into levels of depth, is purely arbitrary. In reality, there is only a background sea of interacting material particles, and those collections of particles to which we choose to give names have no *ontological* status beyond the reality of the constituent particles.

Second, the reductive materialist claims that the causal powers of ordinary objects derive from the elementary particles and their interactions. As Terence Horgan writes, 'any broadly materialistic metaphysical position needs to claim that physics is causally complete.'[21] This means that 'there are no new causal powers that magically accrue to [objects, such as minds,] over and beyond the causal powers'[22] of the interacting constituent particles. Although we are accustomed to 'ascribe causal agency to macroscopic things . . . true causal agency resides in the particles alone.'[23] The alternative to this reductionist view of causality, about which we shall have much more to say, would allow that an object might possess 'genuinely novel'[24] *emergent* causal powers that can not be predicted from, or explained in terms of, a complete knowledge of the particles and their interactions. Consequently, the

[21] Horgan (1993) 573.
[22] Kim (1996) 232.
[23] Faber (1986) 5.
[24] Kim (1996) 227.

reductive materialist is concerned to show that there are, and can be, no such emergent properties or causal powers.

And third, the reductive materialist claims that 'everything is *explicable* in terms of physics.'[25] In principle, there are no mysterious relations between macroscopic objects and the parts that compose them. Once we have acquired a complete understanding of the properties of the parts, and their relations with one another, we will be able to explain the properties of the object, for the reductive materialist believes that all the properties of the object 'are reducible to, and hence ultimately turn out to be, physical properties'[26]

This claim that the properties of the object are reducible to the properties of the particles and their interactions raises the central contention amongst materialists. As we have said, all materialists are agreed, ontologically speaking, that objects are composed of material particles; there are no 'extra-physical objects [such as] entelechies, psychical phenomena and the like.'[27] But they are not similarly agreed about whether objects can be *reduced* to their constituents. Whereas the reductive materialist believes that '[n]ew properties cannot spontaneously emerge in complex systems without being explicable in terms of the structure and components of the system,'[28] the non-reductive materialist argues that there are, or at least might be, features or aspects of the world that can not be properly understood merely in terms of the constituent particles and the laws that govern them; these features or aspects of the world resist *reduction* to anything else.

Consequently, in order to assess the relative strengths of these positions, we must understand what it means to reduce something to something else. But acquiring this understanding is made difficult, as a number of writers have observed, by the confusing 'number of different reductionist theses'[29] that one finds in the literature on the subject. John Searle, for example, 'find[s] at least . . . five different kinds of reduction in the theoretical litera-

[25] Brooks (1994) 803, emphasis in the original.
[26] Kim (1996) 212
[27] Smart (1978) 383.
[28] Brooks (1994) 803.
[29] Howard (1979) 163.

ture.'[30] This difficulty is compounded because, in their discussions of reduction, 'many authors shift uncritically from one [kind of reduction] to another.'[31]

According to the most basic formulation, reduction is merely one type of relation that can exist between different kinds or realms of things. To say that one thing reduces to another is to say that a particular type of relation exists between them. Therefore, we must consider the two components of a reductive relation; namely, the type of relationship that reduction is, and the things between which this relationship holds.

The reductive relation is often expressed by saying that something 'is just' or 'is nothing more than' something else.[32] For instance, to say that light can be reduced to electromagnetic radiation is to say that light is just, or is nothing more than, electromagnetic radiation. Of the two things that figure in a reductive relation, then, we can say that one of them is real, and the other is simply another, less accurate, description of the real thing. The one that is reduced, therefore, is eliminated as a real thing in its own right; when we successfully reduce one thing to another, we in effect *reduce* the number of real things in the world. Reduction is 'a method of eliminating unnecessary entities from the world's content.'[33]

But what are the things that philosophers attempt to reduce and thereby to eliminate from the world's content? The answer is a great many things. But these fall conveniently into two broad categories. First, one may seek the *ontological* reduction of the objects of experience and their properties; the reductive materialist, as we have seen, wanting to reduce them to complex arrangements of material particles. Or second, in *scientific* reductionism, one may attempt to reduce the terms, laws, or theories of one science to the terms, laws, or theories of some other science. The ultimate goal of most scientific reductionists is to show that natural science is really nothing more than physics.

[30] Searle (1992) 113.
[31] Barbour (1997) 231.
[32] Trout (1995) 387.
[33] Mumford (1994) 420.

In addition to ontological and scientific reduction, many discussions of reduction distinguish a third form: methodological reduction. This refers to the scientific strategy of seeking the explanation of a thing in terms of its parts. According to this view, 'we can best understand a whole by *analyzing* it (conceptually, at least) into its component parts.'[34] For example, an ecologist may seek to understand an ecosystem by analyzing the behaviours and interactions of the various individual organisms that make up that ecosystem. Rosenberg describes methodological reduction as merely a 'commitment' to the scientific method of seeking 'lower level' explanations for 'the behavior of the stuff of nature.'[35] But as Barbour recognizes, '[o]ne could adopt reduction as a practical research strategy without claiming that all biological theories will be derived from chemical theories, or that nothing exists in the world except material particles.'[36] I shall have nothing more to say about methodological reductionism in what follows.

I am concerned here with the two main types of reduction: ontological and scientific. Ontological reduction, as we have said, can be divided into two varieties, as it concerns the reduction of objects or properties. Object ontological reduction seeks to reduce things, including the objects of experience, to their ultimate natures. In this way one explains 'some phenomenon . . . by showing that "in some sense" it really *is* something else.'[37] Common illustrations include the claims that water is really H_2O, lightning is just electrical discharge, and a gene is nothing more than a segment of DNA.[38] In each of these three pairs, only the second item is real; in being reduced, water, lightning, and genes are shown not to be real things in their own right. To use Faber's colorful words, in 'the great drama of the universe . . . [these things are not] featured on the marquee.'[39] Thus, by showing that 'objects of cer-

[34] Howard (1979) 163, emphasis in the original.
[35] Rosenberg (1986) 72.
[36] Barbour (1997) 231.
[37] Girill (1976a) 69.
[38] Nagel (1995) 159 gives these examples of 'successful reduction.'
[39] Faber (1986) 11.

tain types . . . [are] nothing but objects of other types,'[40] the ontological reductionist seeks to demonstrate 'the redundancy of any ontology other than that of basic physics.'[41]

Property ontological reduction, also called micro-reduction, 'is a form of ontological reduction, but it concerns properties.'[42] '[T]he principal aim of a micro-reduction is to *explain* the properties of wholes in terms of the properties of their parts, the fundamental laws governing their parts, and . . . the structure of the whole.'[43] Micro-reduction is thus a type of explanation, in which we seek an 'explanatory insight into *how* and *why* the observed regularities hold at the macro-level'[44] by appeal to the properties of the parts at the micro-level.

But as an examination of the literature on reduction readily reveals, there is no consensus about what it means to explain upper-level properties in terms of those at the lower level. As Robert Causey recognizes, 'the current analyses of explanation are far from being entirely satisfactory.'[45] In the previous chapter,[46] I discussed a number of different philosophical accounts of explanation, and concluded, unsurprisingly, that explanation has no univocal meaning. We simply seek different kinds of explanation, according to our interests and the phenomenon to be explained. Sometimes we want an explanation of what something is; other times we are interested in the cause of a phenomenon; and still other times we are satisfied that an event has been explained by the statement of a law. For example, a person regarding the northern lights (the aurora borealis) might seek an explanation of what the phenomenon is, or what causes it, or under what conditions it occurs, or even all three.

So, then, what kind of explanation is required in micro-reduction? What does it mean to explain the macro-properties of an object by claiming that they are nothing more than the micro-properties of the interacting component parts?

[40] Searle (1992) 113.
[41] Foss (1995) 412.
[42] Searle (1992) 113.
[43] Causey (1972) 415, emphasis added.
[44] Kim (1996) 216, emphasis added.
[45] Causey (1972) 418.
[46] §4

First, the causal model of explanation does not seem appropriate for reduction. In the course of defending the scientific reductionist claim that 'biology is nothing more than chemistry,'[47] Kenneth Schaffner refers to 'the causal mechanism responsible for [upper-level] properties,'[48] thereby suggesting a kind of causal explanation of those properties. But if biological properties are really *nothing more than* chemical properties, for which claim Schaffner believes there is 'persuasive'[49] scientific evidence, then it seems odd, at the very least, to suggest that the chemical properties *cause* those biological properties, because that would seem to imply, paradoxically, that chemical properties cause themselves!

It also seems that the Hempelian model of explanation is inappropriate for micro-reduction. According to that model, the macro-properties of an object are explained if they can be predicted or derived from a complete description of the micro-properties of the parts, together with their arrangement. For Hempel, 'the same formal analysis . . . applies to scientific prediction as well as to explanation.'[50] So, for example, if neuroscience ever develops to the point at which it is possible to deduce a person's psychological state from the state of his brain, that deduction would constitute an explanation, in the covering-law model, of his psychological state.

The problem with this 'view that . . . explanation consists of nothing but derivation'[51] is that it fails to recognize the 'clear conceptual distinction between identities and correlations.'[52] Identities 'require empirical justification, but do not require scientific explanation,'[53] while a correlation 'needs both justification and explanation.'[54] Consider, for example, that well-worn illustration of an identity: 'the morning star is the evening star.'[55] It may be necessary to justify empirically the claim that the morning and evening stars are in fact

[47] Schaffner (1969) 326-7.
[48] Ibid., 333.
[49] Ibid., 326.
[50] Hempel and Oppenheim (1953) 322.
[51] Causey (1972) 413.
[52] Ibid., 417.
[53] Ibid., 407.
[54] Ibid., 414.
[55] Schaffner seems to be fond of the illustration; e.g., Schaffner (1967) 143 and (1969) 329.

one and the same star—the planet Venus actually—but, as Causey points out, 'once this identity is accepted, we do not try to explain why it is true.'[56] An established identity between two things *is* the explanation of their relation. Correlations, however, are relations that require an explanation between the two things. To take one example, scientists have discovered a statistical correlation between exposure to asbestos fibers and the occurrence of mesothelioma, but 'in the practice of science we are not satisfied with an explanation that fails to establish the connection'[57] between the exposure and the development of cancer. We want to know *why* this correlation holds. Merely identifying a correlation 'fail[s] to fulfill the central goal of reduction, namely, that of enhancing our understanding of why emergent phenomena occur as they do.'[58]

Now let us return to the relation between a person's psychological state and the state of his brain. Is this an identity or a correlation? As Causey admits, '[w]hen dealing with a particular case . . . it may be quite difficult to decide whether it is an identity or a correlation.'[59] But here it does not seem very difficult to recognize that the relation between the mind and the brain stands in need of explanation, unlike, for instance, the relation between the morning and evening stars, which presents no mystery. It is still true to-day, as Thomas Nagel wrote thirty years ago, that 'we have at present no conception of what an explanation of the physical nature of a mental phenomenon would be.'[60] As Kim concedes, 'it is difficult to see how [the intrinsic properties of mental experience] could be reductively identified with anything else.'[61] Perhaps it is simply the case 'that any correlation we might find between a phenomenal quality of experience and a neural property must be taken as brute and fundamental, not amenable to further explanation.'[62]

---

[56] Causey (1972) 414.
[57] Faber (1986) 12.
[58] Kim (1996) 229.
[59] Causey (1972) 417.
[60] Nagel (1995) 159.
[61] Kim (1996) 236.
[62] Ibid., 236-7.

So it seems that neither the causal model nor the covering-law model of explanation is an appropriate description of explanation in the reductive relation. This leaves the third model of explanation, according to which we explain something by giving an account of what it is. At first blush, this would seem to be what we mean by a reductive relation. We explain what a thing is by showing that it is really something else; we 'identify' the reduced thing with the reducing thing; water, for example, is reduced or explained by identifying it with H_2O. However, this model is incomplete, because, as we have said, an identity must be justified in order to be explanatory; an unjustified identification is merely an unwarranted stipulation. We need grounds for thinking that one thing really is some other thing. This means that if there is any question about the relation between two things, if the 'correlation between [them] . . . is extremely mysterious,'[63] we do not dispel the mystery simply by stipulating that one of them is the other. A philosopher of the mind does not explain the qualitative aspects of experience by declaring that the mind is the brain. An identification explains only 'once this identity is accepted.' What we require then is an account of how we justify the identification of one thing with another.

Let us pursue the subject of explanation by way of a couple of venerable examples of 'successful' micro-reductions. '[S]olid objects [have the property of being] impenetrable by other objects, . . . [and] resistant to pressure.'[64] The individual component particles do not have this property. For example, ice has much greater resistance to pressure than liquid water, even though ice and liquid water are both composed of molecules of H_2O. But this property of impenetrability is explained by identifying it with the 'vibratory movements of molecules in [the] lattice structures'[65] of such solid objects; accordingly, impenetrability is nothing more than another name for the disposition of certain arrangements of molecules to resist pressure. This identification is justified because, and only to the extent that, we understand why certain arrangements of molecules show such resistance to pressure. Thus,

[63] Causey (1972) 414.
[64] Searle (1992) 114.
[65] Ibid.

the identification of the property of impenetrability with the micro-structure of solid objects is supplemented with an account of how that micro-structure is able to resist pressure; that is, an account of how the micro-structure is able to do the things that solid objects do. This gives us a successful micro-reduction of the macro-property of impenetrability.

Material objects have the property of *having heat*. This means, among other things, that the objects have the power to bring about 'the expansion of mercury columns in thermometers, sensations of heat, and combustion.'[66] Various accounts of heat have been given throughout history. First, Aristotle regarded heat as one of the four primary irreducible qualities,[67] which, along with cold, wet, and dry, combine to form the four fundamental material elements: earth (cold/dry), air (hot/wet), fire (hot/dry), and water (cold/wet).[68] Second, heat has also been treated as a substance. According to the caloric theory, which flourished from the eighteenth century to the middle years of the nineteenth, 'heat was a subtle, invisible, weightless fluid, passing freely between the particles of bodies.'[69] And third, Newton and Boyle accepted the modern view that heat is the motion of particles of matter. This kinetic theory posits that heat 'is nothing more than the mean kinetic energy of molecule movements;'[70] that is, 'being hot is nothing over and above having some amount of molecular K. E.'[71] Thus, we can effect a micro-reduction of the property of *having heat* by identifying heat with the motion of the molecules that constitute an object, *and* explaining how the motion of those molecules can cause an expansion of mercury in thermometer, sensations of heat, and so forth.

These examples illustrate the relation between reduction and explanation. From these brief discussions, we can see that there are ostensibly *three* elements involved in a micro-reduction: first, an arrangement of micro-particles that constitute the object; second, the

[66] Jackson (1980) 27.
[67] Aristotle, *On Generation and Corruption*, Bk. 2 Ch. 2, 329b17-330a29.
[68] Ibid., Bk. 2 Ch. 3.
[69] Dampier (1957) 94.
[70] Searle (1992) 113.
[71] Jackson (1980) 27.

macro-property of the object, such as heat or impenetrability; and third, the effects of the property on other things in the world; that is, what the property *does*. The effects of heat, as we have seen, include the combustion of inflammable objects and sensations of heat. There are two questions that must be answered with respect to micro-reduction: How are these three elements related in a micro-reduction, and what account of explanation is appropriate?

The micro-reductive answer is as follows. First of all, the macro-property is not a real thing in its own right, in addition to the arrangement of particles; heat is not a thing independent of the collection of particles that constitute the hot object. What then is the macro-property? It is nothing more than 'a disposition'[72] of the collection of particles to affect other things in certain ways. Heat is just a name for the power of an arrangement of moving particles to cause combustion and sensations of heat. This 'identity' between the property and the causal powers of the arrangement does not need an explanation; the property and the causal power are simply the same thing.

Second, a micro-reduction provides a *causal* explanation for the connection between the arrangement of particles and the effects that we associate with the property; and '"connection" here means "mechanism."'[73] 'Hence, the particles, not the . . . upper-level [properties,] perform the causal activity'[74] What needs explanation then is how heat, that is, molecular motion, can cause a column of mercury in a thermometer to rise, or cause objects to combust. A mere lawful correlation between molecular motion and combustion would be insufficient to effect a reduction of heat, because correlations themselves require explanation.

Finally, the micro-reduction justifies the identity between the macro-property and the arrangement of micro-parts by showing that the micro-parts can explain all the causal effects that we associate with the property. The reductionist can not 'simply stipulate that this cor-

[72] Causey (1972) 415.
[73] Faber (1986) 12.
[74] Ibid., 133.

relation is an identity'[75] without justification. But by providing a micro-explanation of all the effects attributed to the property, we can be confident that the property is nothing more than a disposition of the micro-parts. As Searle explains, 'once we know that [combustion and sensations of heat] are caused by [molecular motions], we then redefine the word [heat] in terms of [molecular motions].'[76]

Obviously, the challenge of micro-reduction is not simply to assert an identity between the macro-property and the micro-parts, but to justify that identity by explaining how the micro-parts cause the effects we associate with the macro-property. Micro-reduction is different from scientific reduction, in that it employs 'an explanatory strategy that is prima facie different,'[77] and 'should thus be kept distinct'[78] from scientific reduction, with which most of the literature on reductionism seems to be concerned.

Scientific reduction, also called inter-theoretic reduction, concerns itself with the lawful description of the world presented by natural science.[79] In an inter-theoretic reduction, one reduces some scientific theory to some other theory by showing that 'the laws comprising the reduced theory are deductively valid consequences of the laws comprising the reducing theory.'[80] For example, to reduce Mendelian genetics to molecular biology, which 'has widely been held' to be 'the test of reductionism,'[81] one would have to show that Mendel's laws of segregation and independent assortment, or at least corrected versions of them, can be derived from the laws of molecular biology. This derivation, if it could be accomplished,

---

[75] Causey (1972) 415.
[76] Searle (1992) 115.
[77] Girill (1976a) 70.
[78] Howard (1979) 163.
[79] Since Nagel's classic discussion of 'the reduction of theories' (1961, Ch. 11), the bulk of the literature on scientific reduction has addressed the reduction of scientific *theories*. Less attention has been paid to the reduction of scientific branches or disciplines, such as biology or genetics, for it seems to be assumed that these are only sets of theories, and that reducing each of the component theories results in the reduction of the branch or discipline. For a critique of this view, and a defense of the distinction between scientific theories and fields, see Darden and Maull (1977).
[80] Bickle (1992) 49.
[81] Rosenberg (1986) 90.

would show that, in a sense, 'the wider reducing theory "contains" the narrower reduced one as a special case.'[82]

It is a curiosity that although most of the literature on reductionism is devoted to scientific reduction,[83] genuine examples of successful inter-theoretic reduction seem to be quite rare in the history of science; Searle observes that 'the same half dozen examples are given over and over in the standard textbooks.'[84] The standard illustration is the reduction of 'classical thermodynamics [to] molecular statistical mechanics.'[85]

Now, having distinguished briefly the different types of reduction, I shall consider at some length the arguments that have been adduced for and against each type. In the following chapter, I shall treat scientific reduction; in the balance of this chapter, I shall take up the controversy between the reductive and the non-reductive materialists with respect to object and property ontological reductionism, and provide some assessment of the strength of each position.

If the ontological reductionist is correct in his thesis about the nature of reality, then it follows that a complete description of the world in terms of the interacting particles and their properties would leave out nothing real. A hypothetical Laplacian being, standing aloof from the world, with a complete knowledge of each particle and its properties, would know all that there is to know, on an ontological level, about the world, even though, remarkably, its account would include no mention of any of the things that fill our experience: physical objects, mental objects, even conscious beings, such as ourselves.[86] In the same

---

[82] Ibid., 91. For an up-to-date discussion of the relation between classical Mendelian genetics and molecular biology, see Sterelny and Griffiths (1999) chs. 6 and 7. Hull (1974) provides a useful introduction to the subject, and raises a number of criticisms of this reduction.
[83] Howard (1979) 163.
[84] Searle (1992) 113.
[85] Horgan (1993) 575.
[86] Of course, our hypothetical Laplacian being might 'notice' that there are stable clumps of interacting particles, but these groupings, which we call macro-objects, would not be regarded as having a separate ontological existence apart from the constituent particles. As Faber (1986) 5 explains, 'In our ordinary talk about the world, we habitually ascribe causal agency to macroscopic things which, according to [reductive materialism], are either conglomerates of atoms or whirling patterns whose atomic memberships continually change. We

way, a complete molecular description of the ocean would say nothing about the arbitrary, artificial divisions that oceanographers have imposed upon it. The objects of experience, including living creatures, being nothing more than collections of particles, according to the reductionist, a complete account of all the constituent particles would exhaust all that there is to say ontologically about the objects.

Amongst materialists, the central controversy in reduction is whether non-reductive materialism is 'a viable metaphysical position.'[87] Materialists are all agreed that everything real is composed of material particles, that there are no immaterial substances. But they disagree about whether this commitment to materialism entails a further commitment to the stronger thesis of reductive materialism. As one might expect, there are three types of non-reductive materialism, which correspond to the three forms of reduction that we have identified: object ontological reduction, property ontological reduction, and scientific reduction. It is not clear, however, that writers on reduction have always been scrupulous in distinguishing these.

The non-reductive materialist may therefore resist the claims of reductionism in one of three ways. First, while allowing that an object contains nothing more than material particles, the non-reductionist may nonetheless argue that the object in question is a real thing that 'in some sense "really and independently exists"'[88] and as a result must be included on the ontological inventory along with the particles of which it is composed. This assertion that the object is not just a collection of particles typically follows from the belief that the object has one or more properties that can not be micro-reduced to properties of the constituent particles. This is a second type of non-reductive materialism, according to which objects have properties or 'causal powers that magically *emerge* at a higher level and of which there is no accounting in terms of lower-level properties and their causal powers and

name these things as causal agents for convenience, to simplify the stories we tell, or out of ignorance; but true causal agency resides in the particles alone.'

[87] Ibid.

[88] Gellner (1956) 160.

nomic connections.'[89] The third type, which we shall discuss in the following chapter, 'hold[s] that a viable . . . materialistic position need not assert that the special sciences generally, and mentalistic psychology in particular, are reducible to physics.'[90]

Emergence

The viability of non-reductive materialism, as it concerns objects and their properties, thus depends upon the viability and coherence of the concept of emergence. The term has been rendered in various ways, some of them more helpful than others. Sometimes philosophers circularly define non-reductive emergence simply as the opposite of reduction. According to Kim, emergentism is the doctrine that '[e]mergent properties are irreducible to, and unpredictable from, the lower-level phenomena from which they emerge.'[91] The general idea seems to be that the interaction of material particles inexplicably brings into existence novel macro-properties that can not be predicted from even a complete knowledge of the particles and their properties. Let us try to be more specific.

I find three distinct types of emergence tangled in the literature, only one of which is truly incompatible with reduction. I shall consider it last. The first type of emergence holds that the macro-properties of the whole are in fact *unpredictable* from knowledge of the parts and their interactions; that is, 'an emergent property [is] a property at one level of organization that cannot be predicted from properties at lower levels of organization.'[92] In the past, a proponent of this emergent position might have suggested, by way of illustration, that 'the peculiar properties of water . . . could not be deduced from our understanding of the properties of hydrogen and oxygen.'[93] But as modern chemistry has now developed to the point that 'the properties of water can be explained by reference to the properties (including the

[89] Kim (1992) 18, emphasis added.
[90] Horgan (1993) 575.
[91] Ibid., 228.
[92] Howard (1979) 164.
[93] Mayr (1982) 63, attributing this position to T. H. Huxley in 1868.

interactive propensities) of neutrons, protons, and electrons,'[94] a different example is enjoying wide currency.

One is more likely to be told these days that a full understanding of the properties of the individual brain cells, including their interactions, does not permit us to predict or account for the properties of the mind. Although the individual neurons are incapable of thinking, a vast collection of them, interconnected by nerve fibers, has the emergent, unpredictable capacity of thought. Kim acknowledges the frustrations of many of those working in philosophy of mind, when he writes that the 'emergentists may have been right after all in despairing of giving an intelligible account of how consciousness could, or should, emerge from the molecular processes of the nervous system.'[95] But absent an explanation of how the brain could cause the features that we associate with the mind, such as the presence of consciousness and the 'phenomenal properties of experiences,'[96] the properties of the mind have not been micro-reduced.

In summary, the type-one emergentist argues as follows: if micro-reductionism is correct, then given an understanding of the component parts and their interactions, we should be able to predict the properties of the whole; but we can not make such a prediction in many cases, the most conspicuous illustration being our inability to predict the properties of the mind from the properties, structure, and functioning of the brain cells; therefore, micro-reductionism is false.

The argument has an initial plausibility, particularly 'if we assume [with Hempel] that . . . explanation and prediction are symmetrical,'[97] since micro-reduction requires an explanation of the relation between the property to be reduced and the properties of the constituent parts. However, I have already argued that explanation is not equivalent to prediction, and in micro-reduction particularly, the type of explanation required is an account of

---

[94] Faber (1986) 8.
[95] Kim (1996) 236.
[96] Ibid.
[97] Brittan (1970) 454.

how the structure and properties of the micro-parts can bring about the effects associated with the upper-level property. Providing such a micro-explanation need not amount to a prediction.

The more significant problem with this account of emergence is that our current inability to explain the relation between a macro-property and the micro-constituents does not establish that such an explanation is unavailable *in principle*. After all, the property of superconductivity was discovered almost half a century before scientists found a satisfactory quantum explanation of it,[98] thereby showing that superconductivity is nothing more than a disposition of the particles under certain conditions. For the reductionist, failure to explain or predict the behaviour of an object does not refute reductionism; it merely indicates that 'the analysis required is more complicated, and perhaps for that reason more difficult, than at first suspected.'[99] Accordingly, the reductionist is confident that the mind will eventually yield to reduction, as additional knowledge of the brain explains mental phenomena, in the same way that we have come to understand that heat and superconductivity are nothing more than dispositions of certain collections of particles.

The second type of emergence concerns the *organization* of particles in a collection. According to this variety of emergence, the arrangement or 'relational property [of the particles] . . . is held to be emergent because, however much we know about the other properties of the [constituents], it is not clear that we are able to predict [how they] are assembled.'[100] All objects are ultimately alike, in that they are all made of the same fundamental material particles, so there must be something besides the matter that distinguishes objects from each other, and explains their causal powers. This additional ingredient that distinguishes one material object from another is what the ancient Greeks called form: the arrangement or structural relation of the parts. The structure of the particles and their properties determines

[98] John Bardeen, Leon Cooper, and John Schrieffer were awarded the 1972 Nobel Prize in Physics for their work in 1957 providing a quantum explanation of superconductivity, which was discovered by Kamerlingh Onnes in 1911.
[99] Howard (1979) 169.
[100] Ibid., 164.

the causal powers of the object. For example, 'a *kind* of molecule (e.g., an H_2O molecule) [is] defined in terms of its parts (two H atoms and an O atom) plus a description of how these parts are structurally oriented with respect to one another;'[101] and it is the 'three-dimensional *structure* [of a protein that] causally determines its effects, and in particular all its biological functions.'[102]

Can this *arrangement* of the particles, their *structural* relations, be micro-reduced to properties or dispositions of the particles? As the non-reductionist points out, this structure is underdetermined by the parts themselves, since the same particles can be arranged in a myriad of structures; carbon atoms, for example, can 'crystalliz[e] into graphite or diamond, [and] the same proteins can be assembled into different structures.'[103] The properties of the particles are thus not a sufficient condition for explaining how those particles will be related; 'knowledge . . . of all the components of the cell would not . . . enable us to predict that these components come together in just the way they do to form a cell.'[104] A micro-reduction explains the properties of an object in terms of the parts and their arrangement, without explaining the arrangement itself. But without such a micro-explanation of the properties of the whole, including the arrangement or relational properties of the constituent particles, there can be no reduction.

This argument, and a slightly different version that will be discussed under scientific reductionism, have succeeded in persuading many philosophers that reductionism, in the words of biologist Ernst Mayr, 'is at best a vacuous, but more often a thoroughly misleading and futile, approach.'[105] However, this argument in favor of emergentism ultimately fails, I think, to undermine the reductionist attempt to explain the material world in terms of material particles and their relations.

[101] Causey (1972) 409.
[102] Rosenberg (1986) 73, emphasis added.
[103] Goodwin (1979) 110.
[104] Howard (1979) 164.
[105] Mayr (1982) 63.

So how can the reductionist explain the structure of an object? It is important to notice that the question whether the reductionist can explain the structure of an object is different from the question whether he can explain the properties of an object in terms of the parts and their arrangement. For example, the reductionist might very well be able to explain why one arrangement of carbon atoms gives rise to the properties of a diamond, and another arrangement produces the properties associated with graphite, without at the same time being able to explain, by considering only the properties of carbon atoms, why a particular collection of such atoms happen to be arranged in one structure or the other. The emergentist has correctly observed that among the many ways that a set of particles can be arranged, we may not be able to tell from the properties of those particles alone which of the *possible* arrangements will be the *actual* arrangement.

But this is not a failure of micro-reduction. It bears repeating that a micro-reduction explains the properties or causal powers of an object in terms of the dispositions of the parts plus their arrangement. A property of an object is just a name for the ability of an object to do something, or to affect other things in some way: a *solid* body has the ability to resist pressure, a *warm* body has the ability to raise thermometers. A micro-reduction then explains how the particles, arranged in some fashion, have this ability, and cause these effects. For example, the reductionist explains why an arrangement of carbon atoms in sheets has the properties of graphite, including its lubricating abilities. Thus, the arrangement of a collection of particles helps explain the properties of the whole; the structure is not itself explained in a micro-reduction. In technical language, the structure is an *explicans*, and not an *explicandum*.

Does this mean, then, as the non-reductionist insists, that the arrangement of the particles is an emergent, irreducible property of the object? Not at all. By an arrangement of particles, we mean two things: first, the relative positions of the particles with respect to one another, which we might call their pattern or configuration, and second, the causal interactions of the particles; that is, we want to know where the particles are located, and how they are bound together. Let us consider first the reductionist treatment of patterns or configurations. The pattern of the particles that make up an object is nothing more than a summary or 'arithmetical aggregation'[106] of the positions of the individual particles. Once we have accounted for the location of each particle, we have accounted for the pattern of the particles; and in order to account for the relative location of each particle, we must consider the interaction of that particle with other particles, some of which we do not classify as part of the object. For instance, the configuration of carbon atoms that constitute a diamond is partly determined by the pressure exerted on them by other particles. But any arbitrary grouping of particles into an object will necessarily leave out some of the influences on the selected particles. So the fact that the relative positions of the individual particles in an object are influenced by other particles outside the object is hardly reason to suppose that the summation of these individual positions is an emergent feature.

Now let us consider how the reductionist explains the causal interaction between the component particles of an object. In brief, as one might expect, such interactions are explained in terms of the chemical properties or dispositions of the interacting particles. The interactions between atoms, with a bit of popular over-simplification, 'can be explained . . . as a sharing or swapping of electrons (or both) between atoms, in an attempt to achieve the desirable state of having filled outermost [electron] shells.'[107]

---

[106] Rosenberg (1985) 72.

[107] Gribbin (1999) 69. Chemists distinguish a number of different chemical bonds. In an ionic chemical bond, such as that between the sodium and chlorine atoms in table salt, one atom gives up a negatively-charged electron and the other atom seizes it, thus leaving the 'donor' atom with a positive charge, and giving the 'recipient' atom a negative charge; and since oppositely charged ions (atoms with an electrical charge) attract each other, the two atoms form a stable ionic bond. In a covalent bond, atoms share pairs of electrons, which interact with the nuclei of both atoms to form a stable attraction. In a molecule of methane, for instance, one carbon atom is covalently bonded to four hydrogen atoms (CH4). Each of the hydrogen atoms, with one electron, shares one of the four electrons in the outer shell of the carbon atom, thereby giving each hydrogen atom a stable configuration of two electrons, while the carbon atom shares all four electrons provided by the hydrogen atoms, and achieves a full, stable complement of eight electrons in its outer shell. And a hydrogen bond exists between molecules, and though much weaker than either ionic or covalent bonds, this bond is especially important in organic molecules. For instance, the folded structure of a protein molecule is partially held in place by intermolecular hydrogen bonds between the amino acids.

In summary, the non-reductionist has claimed that the form or structure of an object is an emergent property, which can not be explained in terms of the properties of the particles that compose the object. But, as we have seen, the reductionist can explain both the configuration of the particles and how they are bound together, as the consequences of the properties of the particles, and their interactions with other particles. Therefore, the form of an object being neither mysterious nor inexplicable, it can not properly be regarded as an emergent, irreducible feature of an object.

The third type of emergence, a 'stronger,'[108] 'much more adventurous conception,'[109] I shall call ontological emergence. It has also been dubbed 'British emergentism,'[110] for the group of British philosophers who developed and promoted it from roughly the middle of the nineteenth century through the first quarter of the twentieth: Samuel Alexander, George Henry Lewes, Alexander Bain, A. O. Lovejoy, C. Lloyd Morgan, and C. D. Broad.[111]

Ontological emergentism can be understood as involving two theses, one concerning the nature of emergent properties, and the other concerning explanation. Kim describes the first of these as the doctrine that '[w]hen aggregates of material particles attain an appropriate level of structural complexity ("relatedness"), genuinely novel properties emerge to characterize these structural systems.'[112] This has been called a form of 'property dualism,'[113] because an emergent object possesses both physical properties—those that can be micro-reduced—and non-physical properties—those that can not be reduced to properties of the component parts and their relations. These non-physical properties are taken to be 'fundamental force-generating properties, over and above the force-generating properties'[114] possessed by the particles. An emergent causal power is like a genie rising out of a rubbed

[108] Howard (1979) 164.
[109] Searle (1992) 112.
[110] Horgan (1993) 557 attributes the name to Brian McLaughlin, 'The Rise and Fall of British Emergentism,' in *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, Beckerman, Ansgar, et al., eds. (Berlin: Walter de Gruyter, 1992).
[111] Kim (1996) 226 and Horgan (1993) 557 identify these figures as proponents of this type of emergentism.
[112] Kim (1996) 227.
[113] Ibid., 228.
[114] Horgan (1993) 557.

lamp,[115] in that it 'gets squirted out by the behavior of the [interacting particles], but once it has been squirted out, it then has a life of its own.'[116]

The second thesis of ontological emergentism is that these non-physical properties 'are metaphysically *sui generis*, unexplainable in more fundamental terms.'[117] For the reductionist, remember, we explain a property by showing how the component particles and their relations could cause the phenomena we associate with that property; to reduce lightning to an electrical discharge, for instance, one must show how an electrical discharge can do the things that lightning does: cause thunder, light up the sky, strike trees, and so forth. But the emergentist believes that certain properties of objects can not be explained in terms of particles and their relations. For example, '[i]f consciousness were [ontologically emergent], then consciousness could cause things that could not be explained by the causal behavior of the neurons.'[118] The emergentist 'may legitimately point out that, in a contingent universe, not every fact gets explained,'[119] and perhaps the relation between micro-particles and macro-properties is just such a 'brute, unexplainable matter of fact.'[120]

Moreover, as Kim recognizes, 'the emergentist would have [no] problems with [the existence of] . . . laws connecting emergent properties with lower-level properties,'[121] because the emergentist seeks an explanation of that connection, an explanation 'that renders the [upper-level] phenomena intelligible by explaining why they occur under just those conditions in which they do occur.'[122] So even if it is possible to establish a nomic correlation between upper and lower-level properties, this does not constitute a reduction of the upper-level properties unless the correlation itself is explained. As neuroscience progresses, it may

[115] Long after I had written this sentence, I discovered in McGinn (1995) 272 the following epigraph attributed to Julian Huxley: 'How it is that anything so remarkable as a state of consciousness comes about as a result of initiating nerve tissue, is just as unaccountable as the appearance of the Djin, where Aladdin rubbed his lamp in the story . . .'
[116] Searle (1992) 112.
[117] Horgan (1993) 560.
[118] Searle (1992) 112.
[119] Faber (1986) 229.
[120] Kim (1992) 22.
[121] Kim (1996) 228.
[122] Ibid.

become possible to predict certain mental experiences from the state of one's brain, just as we can deduce the occurrence of a dream state from the observation of rapid eye movement (REM); but the reductionist has not thereby shown that the mental experience is nothing more than a brain state unless he can also explain how the brain is able to produce such experiences. The availability of such an explanation is central to the denial of ontological emergence, because, as Jeffrey Foss notes, the mere ability to predict mental states from brain states 'is quite consistent not only with eliminativism and reductionism, but even with dualism, supervenience, and epiphenomenalism.'[123]

But even without an explanation of some upper-level property in terms of lower-level conditions, ontological emergence is, 'for many people, impossible to accept; it is too uncomfortable.'[124] Consequently, the position is often simply dismissed out of hand. For example, Searle doubts that there are any ontologically emergent properties, 'because the existence of any such features would seem to violate even the weakest principles of the transitivity of causation.'[125] Other writers seem to be unaware of the doctrine of ontological emergence. In his long chapter on biological reductionism,[126] Rosenberg does not discuss emergence at all, as far as I can tell. And although he does briefly mention 'holism' and 'vitalism' as alternatives to the reductionist view that biochemical structures are the 'explanatory foundation of all biological phenomena,' he disparages them, without argument, as positions 'that no responsible biologist could seriously credit.'[127] Whether or not he means by holism what I am calling emergentism, is a matter that is not easily decided, but it is just as well, for there are a number of writers who have made clear their objections to ontological emergence. I turn to those now.

One of the first difficulties that is apt to occur to someone is this: how can there be something in the object that was not first in the parts? Unless we are prepared to abandon

[123] Foss (1995) 405.
[124] Grene (1971) 21.
[125] Searle (1992) 112.
[126] Rosenberg (1986) 69-120.
[127] Ibid., 73.

the hoary *a priori* principle that something can not come from nothing, it would seem that there must be *some* source for the properties that emerge in the object. However, the objects of experience are made of nothing more than material particles, as all materialists are agreed, so if the properties of such objects do not reside in the particles, then there would seem to be nothing from which the properties can emerge. Thus, the emergentist position would appear to be inviable from conception.

The emergentist is certainly not proposing that emergent properties should come from nothing; rather, such properties of the object emerge, as it were, from the interaction of the particles, from their union. Recall the two theses of ontological emergentism: novel causal powers appear at certain levels of organization, and they can not be explained in terms of the properties of the particles and their interaction. Thus, the emergentist conceives of an emergent property as a property that an object has simply by virtue of its nature, just as a material particle has a suite of properties that belong to it by virtue of its character as a particle. When a particle of matter comes into being, coalescing out of energy, it possesses a variety of inherent properties and causal powers, including the ability to attract other particles, and be attracted by them. But whence do these fundamental properties derive? We might say that they do not derive from anywhere; they are simply the natural characteristics of what it is to be a particle: something can not be a fundamental particle without them. The emergentist wants to offer a similar description of ontologically emergent properties.

The emergentist allows that many objects—perhaps most objects—are just collections of particles, whose macro-properties are nothing more than the properties of the particles; but other objects, he asserts, have properties that belong only to the objects. These emergent properties do not exist in the world prior to the coming into being, the gathering together, of the object from the concrescence of material particles. It is one of the inexplicable principles of the universe that material particles have un-derived properties, and according to the emergentists, it is another such principle that certain objects, certain collections of particles, have unique properties of their own, which, we might say, accompany the object into the world.

The universe contains mysteries that we can not currently explain. Some we will no doubt dispel in time, but other mysteries, ineluctably, must remain, because 'when theorists succeed in [explaining some mystery in terms of] deeper principles[,] the aura will simply float over to those principles.'[128] The reductionist would merely restrict the unexplainable to the level of the particles, explaining everything else by way of those. But the emergentist, on the other hand, apprehends no reason why there should not also be inexplicable properties at the level of objects; perhaps reality is just more complicated than the reductionist would have us believe. There may be a hierarchy of principles that determine the properties of things, from the individual particles to collections of them.

A second objection derives from the failure of the emergentist to provide any *a priori* argument to show that the world is, in fact, hierarchically structured in this way. Of course, this neglect on the part of the emergentist is readily explained: there would seem to be no such argument available, because, as a number of writers have noted, whether it is possible to reduce some object, property, or theory, 'is finally an *empirical* question.'[129] We shall simply have to see whether, in the long run, the proposed reduction can be effected. The reduction of mental properties to neural properties, for instance, 'will require a great deal more knowledge of neurology and psychology than is presently available.'[130]

We may conclude, when 'the whole show is over,'[131] that certain properties do emerge at higher levels, when objects come into being; but this can not be determined beforehand. Indeed, it would seem that it can not be determined at any point. The emergentist can never prove the existence of an emergent property, because the only evidence he can adduce to support his position is that we have failed so far to reduce it; but as we have seen,

---

[128] Faber (1986) 229.
[129] Fodor (1995) 433; others making this point include Bickle (1995) 37, Oppenheim and Putnam (1995) 411, and Place (1995) 108.
[130] Godbey (1978) 434.
[131] Fodor (1995) 433.

our inability to reduce an object to its component parts, given the knowledge we have at a particular time, does not show that the object can not be reduced *in principle*, with greater knowledge. There would seem to be 'no *logical* obstacle to eventual reduction.'[132] Thus, we may conclude only that, in our present state of knowledge, we have failed to explain how the object could derive its properties from its material parts; we have not thereby proven that the reduction can never be completed.

But the mere possibility of ontological emergence does have significance for the debate about reduction, because if it is possible for objects to possess emergent properties, then materialism is theoretically separable from reductionism. The usual understanding of materialism seems to deny this theoretical possibility. According to this understanding, if we accept the materialist ontology, which admits that only the most fundamental particles of matter and their aggregates are real, then it is supposed that we must allow the reductionist thesis to follow as a matter of course; only our ignorance of the basic level of matter stands in the way of a complete reduction of all the objects of experience.[133]

In support of ontological emergence, the non-reductionist will point to our repeated failures to explain some phenomenon as evidence that it can not be explained. For example, Colin McGinn writes that the mind-body problem having 'stubbornly resisted our best efforts' to solve it, perhaps 'the time has come to admit candidly that we cannot resolve the mystery.'[134] Although McGinn's claim is epistemological, rather than metaphysical,[135] a similar argument lies at the heart of the emergentist's belief. On the other hand, the reductionist who wants to assert the inseparability of materialism and reductionism must identify some difficulty with the emergentist position, in order to show that the world does not admit the possibility of emergent properties.

---

[132] Rosenberg (1986) 72, emphasis added.
[133] This seems to be the view of Rosenberg (1985) ch. 4.
[134] McGinn (1995) 272.
[135] Ibid., 286.

That brings us to the third argument against emergentism. Ontological emergentism being more mysterious than ontological reductionism, emergentism is, therefore, less adequate as a theory of reality. The principle upon which the argument rests seems plausible: to paraphrase Ockham, one should not multiply mysteries beyond necessity. Because it is one of the requirements of an acceptable metaphysical theory that it should explain the world, a theory that explains more of the world, all other things being equal, is preferable to one that explains less. The appeal to mystery is merely a way of caulking the interstices of ignorance in one's theory. Of course, the reductionist does not deny that his own theory contains an ineliminable element of mystery, the ultimate material particles having properties that can not be explained; but since the emergentist introduces additional mystery, by also ascribing inexplicable properties to objects, the emergentist presumably provides a less satisfying account of reality. In reductionism, 'the mysteries are confined to the ultimate level of reduction, where they remain ultimately mysterious,'[136] while emergentism admits both lower-level mysteries, and higher-level ones to boot. For example, Smart accepts the reductionist account of the mind, because the alternative 'involves a large number of irreducible psycho-physical laws . . . of the queer sort, that just have to be taken on trust.'[137]

But this argument provides little reason to think that emergentism is false. Having already rejected in the first chapter the principle that simplicity is a criterion of truth, I shall not repeat those arguments here, except to say that we may desire to remove traces of the inexplicable from our account of the world, without being able to remove them. The desire for explanation does not guarantee that it is available. If the reductionist is unable to explain some object in terms of its constituent particles, then, unless the doctrine of emergence involves some demonstrable incoherence, it must be considered a possibility that the properties are simply irreducible, and thus belong to the object as emergent properties. It would be

[136] Faber (1986) 30.
[137] Smart (1995) 130.

impermissible, then, to conclude that the only inexplicable properties belong to the particles, simply because we should like more tidiness in our theories.

A fourth criticism of ontological emergence, closely akin to the preceding argument, accuses the emergentist of committing the informal fallacy of *argumentum ad ignorantiam*, the argument from ignorance. This is the fallacy of thinking that something is true merely because no one has proven that it is false, or that something is false just because no one has proven that it is true. Obviously, our present ignorance concerning the cause of something does not guarantee that a cause does not exist. Therefore, it is illegitimate for the emergentist to conclude, simply because the reductionist has failed to explain some property of an object, such as the mental properties of the brain, that the property in question can not be explained.

Two responses are sufficient to dispose of this objection. First, while it is true, strictly speaking, that we can not deductively conclude from the fact that an explanation is not known that one does not exist, we are often inclined, nevertheless, as a general inductive principle, to proportion our belief that something will not be found, to the amount of effort that has been expended searching for it. 'Pure logic is not the only rule for our judgments.'[138] Repeated failures to explain some phenomenon by means of a given theory increase our confidence that an explanation can not be found with that theory, and encourage the search for another theory with greater explanatory resources. If reductionism continually proves unable, for example, to account for the properties of the mind, then, at some point, it would not seem to be egregiously irrational to consider the possibility that perhaps reductionism is false. Indeed, 'we may find it childish and unreasonable . . . to maintain [a theory] obstinately at any cost.'[139] Our patience for a theory to perform its explanatory promises is not everlasting.

[138] Duhem (1954) 217.
[139] Ibid.

Second, the emergentist would indeed be guilty of reasoning fallaciously if he argued as follows: the reductionist has not provided an explanation for the properties of the mind; therefore, reductionism is incapable of ever providing the requisite explanation. This conclusion obviously does not follow from the premise, since it is possible that a reductionist explanation will be found at some point. However, this is not the argument made by the emergentist.

The emergentist and the reductionist are both attempting to formulate an acceptable metaphysical hypothesis. The emergentist, aware of the reductionist's present inability to account for some anomalous features of the world, such as the mental attributes of the brain, ventures his own hypothesis that such features are merely contingent brute realities. Certain arrangements of particles just give rise to novel causal powers, and there is an end of explanation. The motivation of the emergentist to propose his own metaphysical theory need not depend on any fallacious reasoning concerning the eventual outcome of the reductionist hypothesis; rather, the emergentist having lost confidence in the explanatory power of reductionism, determines to cast his lot with another theory. One need not deductively conclude that the other side will necessarily fail, before one advances an alternative proposal.

However, these two responses are really beside the point, because the charge that the emergentist is motivated by flawed reasoning to propose his metaphysical theory, tells us nothing about the truth of that theory. This is sometimes called the genetic fallacy, which is 'the mistake of confusing the origin of a claim with its evidential warrant and undermining the claim by calling attention to its origin.'[140] Even if the emergentist has insufficient evidence to conclude deductively that reductionism is false, the emergentist may nevertheless be correct. Therefore, we have failed to show that emergence must be false.

A fifth criticism argues that because ontological emergence violates well-established principles of causality, the reduction of all objects of experience and their properties must in

[140] Moreland (1987) 211.

principle be possible, even if our current knowledge does not permit us to perform some reductions in practice. This is the type of argument to which I take Searle to be referring when he worries that ontological emergence may 'violate . . . the transitivity of causation.'[141] The argument has been developed and defended by Jaegwon Kim,[142] whose account informs this discussion.

We have previously identified a property of an object as the ability of that object to do something; for instance, a colored object has the ability to absorb certain wavelengths of light, while a round object, such as a billiard ball, can 'roll in a straight line on a flat surface when struck.'[143] Therefore, 'having a mental property must endow the thing that has it with powers to affect courses of events in its neighborhood.'[144] But if these mental properties are taken to be emergent, and hence irreducible to the properties of the component parts, then 'this must mean that they have causal powers that are different from those of [the parts].'[145] On the emergentist view, then, the mind can do things, and affect other things, in ways that are unexplainable in terms of the interactions of the brain cells.

As we have also seen, the emergentist allows that there may be nomic relations between the lower-level particles and the upper-level properties, such that it may be possible to predict the properties of the object given the state of the micro-parts, even if the relation between those two levels is inexplicable. Indeed, as Kim recognizes, the emergentist is 'committed to there being such laws: *When appropriate "basal conditions" are present, emergent properties must of necessity emerge*.'[146] This is because it is the structural organization of the particles that gives rise to the emergent properties of the object. After all, the genie can not appear out of a lamp that does not exist.

---

[141] Searle (1992) 112.
[142] Kim (1996) 229-233.
[143] Mumford (1994) 427.
[144] Kim (1996) 230.
[145] Ibid.
[146] Ibid., 228, emphasis in the original.

'How does [the upper-level object] manifest its causal powers?'[147] With respect to mental properties, we might ask how one mental state causes another mental state. Since each mental state is dependent upon a physical state, which gives rise to the properties of that mental state, the only way to change the mental state is to change the physical base upon which it depends. As Kim puts it, '[i]f [a physical state] is a realizer of [a particular mental state], then when [that physical state] is present, [the mental state] must be there, too,'[148] and can not be there without some physical realizer or other, which is capable of giving rise to the emergent mental properties. Therefore, Kim writes, '[i]t is plausible to think that the only way for validating [the mind's] causal role is for the following to be the case: [one mental state] causes [another mental state] by causing its realizer, [i.e., a particular physical state], to be instantiated.'[149] But in order for the mind to cause a particular physical state to be instantiated, the mind would have to be able to affect its own component parts, which gave rise to the emergent mind in the first place. This ability of something to affect its own parts is called downward causation, and the emergentist would seem to be necessarily committed to its possibility.

However, the concept of downward causation is widely frowned upon by philosophers. For example, Kim cautions that if we allow the 'causation of physical processes by nonphysical properties, . . . [then] the causal closure of physical is breached';[150] we 'retrogress to the Cartesian picture that does not allow, even in principle, a complete and comprehensive physical theory of the physical world.'[151] E. A. Gellner wonders whether we do not all feel intuitively that 'the insubstantial cannot constrain the substantial?'[152] Charles Taylor worries that if 'the neurophysiological account cannot encompass' the mind, we shall be 'forced back on to dualism, [although] this scarcely seems plausible to the contem-

[147] Ibid., 230.
[148] Ibid., 231.
[149] Ibid., italics removed from the original.
[150] Ibid., 232.
[151] Ibid., 233.
[152] Gellner (1956) 168.

porary mind, even if [from a universal perspective] it may be as tenable as any other hypothesis.'[153] And J. J. C. Smart simply can not bring himself to 'believe' that there could be 'ultimate laws of nature' that give rise to ontologically emergent properties, for such laws have 'a queer "smell" to them.'[154]

I detect three primary concerns in these objections. First, downward causation seems to violate all that we know about the operation of causation in the world; experience shows that causation is transitive, always running from the cause to the effect, and never in the other direction. Second, ontological emergence tears an explanatory hole in our physical picture of reality, because it leaves us incapable, even in principle, of accounting for the relation between the properties of an object and those of its component parts. And third, ontological emergence just seems strange, because it allows that an object could have causal sway over its own components, when it is those very components that give rise to the causal power.

Those familiar with the literature in philosophy of mind will no doubt recognize these criticisms as similar to those made against substance dualism. I do not know why, but the attacks on dualism are almost always conducted by means of a barrage of rhetorical questions, against which the dualist is presumably thought to be defenseless. The same questions crop up again and again in practically every critical discussion of dualism. For instance, I casually selected from my shelf an old philosophy text,[155] and in the section devoted to mind and matter, I find, sure enough, these criticisms of mind-body dualism: if the mind and body are truly of different substances, '[h]ow does it happen that bodily injury can produce amnesia or other mental disturbances?' and 'how are we to explain the apparent ability of our minds to exert control over the activities of our bodies?' The writer concludes that 'these and other questions which present themselves show clearly enough that [sub-

[153] Taylor (1971) 58.
[154] Smart 1995) 118-9.
[155] Clifford Barrett, 1935, *Philosophy: An Introductory Study of Fundamental Problems and Attitudes* (New York: MacMillan Company).

stance dualism] is not to be accepted without further explanation.'[156] In other words, dualism does not conform to our accepted view of causation, and it defies explanation; and without that explanation, well, the whole theory is just too strange to be believed.

Leaving the substance dualist to fend for himself, we may ask how the emergentist can respond to these criticisms. The first point to note is that these are not really criticisms of emergentism at all, but merely restatements of the emergentist position, expressed with a sneer. The emergentist regarding emergent properties as *novel*, *inexplicable* causal powers that arise at higher levels of organization, he is hardly abashed by the reductionist's complaints that emergent properties are causal anomalies, which can not be explained.

Of course the relation between the properties of the object and those of the parts is puzzling and inexplicable; that is precisely what it means for properties to be emergent. As biologist Ernst Mayr writes, '[e]mergence is a descriptive notion which . . . seems to resist analysis.'[157] The real question that must be addressed is whether purported instances of emergence might constitute *exceptions* to the 'usual' behaviour of the world; and this question is not answered by assuming, from the beginning, that the 'usual' causal order does not admit of exceptions, since whether or not this is true is what we want to know.[158] Nor, obviously, is the matter settled by Smart's nose, which detects a 'queer smell' in the doctrine, since others, including Mayr, regard it as one of the 'most interesting characteristics' of emergent objects that they 'can affect properties of components at lower levels.'[159]

Supervenience

Our sixth and final criticism of ontological emergentism will occupy most of the remainder of this chapter. This criticism will require the introduction of a new concept, that

[156] Ibid., 88-89.
[157] Mayr (1982) 63.
[158] Incidentally, a similar response may be given to Hume's argument against miracles (Hume (1975) ch. 10), which 'appears to be clearly question-begging. To say that uniform experience is against miracles is to implicitly assume that miracles have never occurred. It seems almost embarrassing to refute so sophisticated an objection by such a simple consideration, but nevertheless, this answer seems . . . entirely correct.' Craig (1994) 150-51.
[159] Mayr (1982) 64.

of supervenience, but briefly the criticism seeks to disclose a hidden incoherence in the doctrine of emergentism, by showing that emergentists can not consistently claim that emergent properties *depend* upon the constituent particles, but are *irreducible* to them. It is clear that the emergentist *is* committed to both claims; but whether he is rationally entitled to do so is, of course, something that we must determine.

As we have seen, emergentists acknowledge the existence of lawful 'correlations between emergent properties and their basal conditions.'[160] They accept that if certain arrangements of particles are present, then the object composed of those particles will necessarily manifest certain emergent properties, though 'there is no explanation for why [these] emergent properties come into being, or why they generate the specific non-physical forces that they do.'[161] As I mentioned earlier, emergentists used to illustrate the doctrine of emergence by means of water, whose 'aquosity,'[162] they asseverated, could not be explained in terms the properties of hydrogen and oxygen, and their interactions. Although the example turned out to be an unfortunate one for the emergentists, since the desired explanation can now be provided, the illustration demonstrates nonetheless that the emergentists recognized a lawful relation between emergent properties and the properties of the interacting component parts. When hydrogen and oxygen are bound together in an H_2O configuration, the properties of water will emerge, with nomic regularity. Thus, the emergentist admits the existence of laws that permit the derivation or prediction of emergent properties from facts about the micro-state; what he denies is the existence of any explanations of those laws, which explanations are necessary for reduction.

Some philosophers, however, seem to believe that this 'apparently harmless admission . . . entails that there is a systematic reductive relationship between . . . different levels of organization.'[163] According to this view, if the emergentist admits that the emergent

---

[160] Kim (1996) 228.
[161] Horgan (1993) 557.
[162] Mayr (1982) 63, citing T. H. Huxley.
[163] Rosenberg (1986) 116.

properties depend upon the micro-structure, such that the emergent properties will necessarily arise whenever the micro-structure is present, then he can not deny the thesis of reductionism; he must concede that the emergent properties are really nothing more than dispositions of that micro-structure. Accordingly, the challenge to ontological emergentism can be stated as the claim that a property of an object can not be both ontologically emergent and nomically dependent on the micro-particles. Philosophers characterize this nomic dependence of the object on its parts as a *supervenience* relation, so the question we must seek to answer is this: Can the ontological emergentists hold, 'consistently with their other principles, that emergent properties are *supervenient* (in the contemporary philosophical sense) on lower-level properties?'[164] I agree with Horgan that '[t]he answer to this question, as far as I can see, is affirmative.'[165]

In responding to this criticism of ontological emergentism, I shall offer both a brief answer and a much longer answer that depends upon a close analysis of the concept of supervenience and its relation to reduction. First, the short answer. There is simply no obvious contradiction that I can discover between the thesis that emergent properties are lawfully correlated with lower-level properties and the two theses that I have ascribed to ontological emergentism. These two theses assert that an emergent object has '(i) . . . fundamental causal properties . . . over and above physical causal forces; and (ii) the connections between lower-order and higher-order properties . . . could be metaphysically fundamental, hence unexplainable.'[166] But those theses are not inconsistent with the position that there are laws connecting the upper and lower-levels properties, such that when a particular lower-level structure is present, novel, inexplicable upper-level causal powers will necessarily arise.

Moreover, it is far from clear, at any event, how ontological emergentism is supposed to differ from the view, which all reductive materialists accept, that the fundamental

[164] Horgan (1993) 559.
[165] Ibid.
[166] Ibid.

particles of matter themselves have basic, novel, inexplicable causal properties. The emergentist and the reductive materialist both believe that there are things that arise, or come into existence, with causal powers that can be lawfully predicted but not explained. Their disagreement concerns only the identity of these things. Since the reductionist admits that the causal properties of, say, quarks and leptons are inexplicable, brute facts, he would seem to have no grounds for dismissing as incoherent the emergentist's ascription of such properties to objects.

It is interesting to note that Kim, who discusses ontological emergentism at some length,[167] does not attempt to show that the position is incoherent; nor does he argue that our ability to predict or derive macro-properties from information about a micro-state entails the reduction of the micro-properties to the micro-state. For Kim, the real problem with ontological emergence, as we have seen, is the problem of downward causation; 'if we reject reductionism, we are not able to see how [emergent] causation should be possible.'[168] With respect to the reduction of the mind, Kim frames the debate as a 'profound dilemma':[169] either we embrace mind-body reductionism, without understanding the connection between them, or we accept that the mind has emergent properties, without understanding how the mind could do anything. 'If our considerations of emergentism and downward causation are generally correct, then if reductionism goes, so goes the intelligibility of mental causation.'[170] But, significantly, Kim never suggests that the theses of ontological emergentism are incoherent.

Rosenberg, however, does seem to believe that reductionism follows from the dependence of the object upon its parts, for he writes that anyone who accepts the 'finitude of nature' and the 'assumption of determinism' is committed to reductionism.[171] For Rosenberg, 'the only obstacle to [reduction] is the relationship between the actual complex-

---

[167] Kim (1996) 226-233.
[168] Ibid., 237.
[169] Ibid.
[170] Ibid.
[171] Rosenberg (1986) 116.

ity of things . . . and our powers to express and manipulate symbols representing their properties.'[172] Reduction is merely a practical, technical problem; there is no question of emergent properties that can not be explained in terms of lower-level properties. In fact, as I mentioned previously, his discussion of reductionism contains no mention of emergence. But Rosenberg is mistaken to think that the derivation of upper-level properties from lower-level properties amounts to a micro-reduction of the upper-level properties, because a mere derivation or prediction, which is roughly what Rosenberg means by determinism, leaves out an explanation of the correlation between the two levels. To render this into philosophical language, reduction does not follow from supervenience; there is no inconsistency in believing that an object supervenes on its material components, even though it is not reducible to them.

Now we turn from the shorter answer to the longer one. What do philosophers mean by supervenience? At bottom, supervenience, like reduction, is merely a type of relation. In a reductive relation, as we know, one thing is shown to be nothing but something else, but in a supervenience relation, one thing is shown to be merely dependent upon, or determined by, something else. 'Supervenience can be understood as a relation of *dependence* of a "supervenient" class upon a "subvenient" class.'[173] For example, 'wholes do not exist in their own right but depend upon the existence of their parts;'[174] hence, the whole supervenes upon the parts.

Supervenience means literally 'to come above': one thing—the 'upper' one—supervenes on, or comes above, some 'lower' thing or things. The term originally seems to have been used in ethical theory, to describe the relation that exists between moral evaluations, such as goodness, and the objects to which those evaluations might be properly ascribed, such as a person or an action; for example, goodness supervenes on the particular set of properties or behaviours that constitute a good action. The lower, subvenient thing gen-

[172] Ibid.
[173] Mumford (1994) 423.
[174] Faber (1986) 13.

erally consists of a class, such as the class of characteristics that 'make' something good or beautiful, the facts upon which a true theory depends, or 'the neural events' upon which Searle's 'present back pain' supervenes.[175] Consider, for example, the beauty that supervenes on a work of art. A painting is beautiful, to the extent that it is, because of its various physical characteristics: the blots of paint on the canvas and their relative locations. The beauty of any particular work 'depends' on these characteristics, such that any other painting that is physically indistinguishable from it would be similarly beautiful. 'Any two works of art that are physically indiscernible,' Kim writes, 'must of necessity be aesthetically indiscernible.'[176]

Philosophers define supervenience in two different ways, which are represented in a passage from G. E. Moore, who provided an early expression of supervenience, though he did not use the term. Referring to the supervenience of ethical concepts, Moore writes: '[I]f a given thing possesses any kind of intrinsic value in a certain degree, then not only must that same thing possess it, under all circumstances, in the same degree, but also anything *exactly like it*, must, under all circumstances, possess it in exactly the same degree.'[177] From this statement, we can identify the two ways of formulating the thesis of supervenience.

First, supervenience can be accounted a form of determinism. To say that some property supervenes upon some collection of lower-level things, means simply that when the lower-level things are present, the upper-level property will *necessarily* also be present; or, to paraphrase Moore, the object *must possess* a supervenient property when its subvenient base is present. Sometimes this determinist sense of supervenience is expressed by saying that the subvenient properties 'fix' the supervenient properties. As Kim observes, '[w]hat higher-level properties a given entity has are totally fixed by the lower-level properties and relations

[175] Searle (1992) 126.
[176] Kim (1996) 222.
[177] Moore (1951) 261, 'The Conception of Intrinsic Value,' emphasis in the original, cited in Horgan (1993) 555.

characterizing its parts.'[178] Or as David Owens puts it, with a theological twist, '[i]f the non-physical facts [of a world] supervene on the physical facts, then all God has to do in order to fix the non-physical facts about a world is to fix the physical facts about that world. Once the physical features of a world are determined, its non-physical features are determined also.'[179]

Second, a property that is supervenient upon something will necessarily also be supervenient upon a physically identical thing; two things that are exactly alike in all their lower-level physical properties will be exactly alike in their upper-level supervenient properties. 'Supervenience assures us that there can be no difference in the non-physical state of two worlds without there being some difference in their physical state.'[180] Rosenberg describes it this way: 'If a set of properties *A* supervenes on another more basic set of properties *B*, then no two objects that share identical properties from the more basic set *B* can differ in the properties they share with set *A*.'[181] Objects can not differ in their bottom subvenient properties without differing in their upper-level supervenient properties. For example, if the mind supervenes on the brain, then two brains that are identical in all their physical properties will be indistinguishable in their psychological properties; the 'sameness of neurophysiology,' writes Searle, 'guarantees sameness of mentality.'[182]

While philosophers are agreed about the meaning of supervenience, there is no similar agreement about the importance of supervenience in the debate between reductionists and their opponents. This disagreement arises, I think, because our understanding of supervenience does not specify how the supervenient properties depend on the subvenient properties, or just how the lower-level properties fix the upper-level properties. In the philosophy of mind, for instance, epiphenomenalists and identity theorists both acknowledge some kind of supervenient relation between the mind and the brain, but they have a fundamental dis-

[178] Kim (1996) 222.
[179] Owens (1989) 60.
[180] Ibid.
[181] Rosenberg (1986) 113.
[182] Searle (1992) 125.

agreement about precisely what kind of dependence this is. The epiphenomenalist explains the dependence of the mind on the brain by saying that the brain causes the mind; the qualitative aspects of the mind 'cause nothing physical [themselves] but are caused by something physical.'[183] The identity theorist, on the other hand, explains the mind's dependence on the brain by means of a reduction; the properties of the mind are fixed by the brain because the mind is nothing more than the brain.

Thus, showing that one thing is supervenient or dependent on another does not explain the kind of dependence that holds between them. For this reason, compared to reduction, Brooks remarks, supervenience is a 'feeble, milk and water pabulum.'[184] Reduction is a *type* of supervenient relation. I shall mention three such types of supervenience.

First, the identity relation, which says that two things that are identical in every respect are really the same thing under two different names, is a species of the supervenience relation. We might call this 'reflexive supervenience.'[185] For example, a fiddle and a violin are the same thing, because, having all the same properties, they satisfy Leibniz's law of indiscernibility, and each could be said to be supervenient on the other, given our two definitions of supervenience. First, when the physical properties of the violin are present, the musical properties of the fiddle are necessarily also present, and *vice versa*; the physical properties of each instrument can be said to fix or determine the upper-level properties of the other. Second, some other object having all the physical characteristics of either of them, will also have the supervenient properties of the other. It is arbitrary which of the instruments is taken to provide the subvenient base that determines the supervenient properties of the other.

I am aware that some philosophers have rejected 'the absurd position that something can supervene on itself.'[186] They argue that supervenience is an 'asymmetrical relation

[183] Jackson (1982).
[184] Brooks (1994) 803.
[185] Scott Kleiner suggested this name to me.
[186] Mumford (1994) 435.

[such] that the subvenient and supervenient can be distinguished.'[187] I find two reasons why these philosophers require that supervenience be an asymmetrical relation. First, it seems odd, to say the least, to suggest that a thing can fix or determine itself. Determinism proceeds from one thing to another; for instance, the fitness of an organism is determined by its 'anatomical, physiological, and behavioral properties . . . and their interaction with physical properties of the environment,'[188] but the contrary does not seem to hold. The fitness of the organism does not fix its molecular properties nor the properties of the environment. And second, supervenience must be asymmetrical in order to account for the asymmetry of reduction. Reductions proceed in one direction; one thing is reduced to something else, but not *vice versa*. For example, 'a cloud is a mass of tiny particles and nothing else.'[189] By reducing the cloud, we eliminate it as a real thing; it is simply another way of referring to a mass of droplets suspended in the atmosphere. Reduction is an asymmetrical relation, and since reduction can be understood as a supervenient relation plus an explanation of that relation, the supervenient relation must be asymmetrical in order to provide reduction with its asymmetry.

However, it seems to me that supervenience can describe a reflexive relation. I acknowledge, of course, that reductions are asymmetrical; a cloud reduces to a mass of droplets, but the droplets do not reduce to a cloud. Reduction must be an asymmetrical relation because the reduced item is eliminated as a real thing; the mass of water droplets that constitute a cloud are 'specifiably more real'[190] than the cloud. Nevertheless, the two accepted definitions of supervenience do not preclude the supervenience of a thing upon itself. All the definitions require is that one set of properties, called subvenient properties, fix or determine a second set of properties, called supervenient properties, and that any object having the same subvenient properties as another object will also have the same supervenient prop-

[187] Ibid.
[188] Rosenberg (1986) 164.
[189] Place (1995) 110.
[190] Faber (1986) 4.

erties. But these definitions do not specify which set of properties is to be regarded as subvenient and which as supervenient.

Now, in most supervenient relationships, it is clear which properties are subvenient, because it is obvious in which direction the ‘property fixing’ runs. Consider a couple of examples. In the supervenient relation between the physical properties of a statue and its aesthetic properties, there is no question which are the subvenient properties, and which the supervenient properties. The physical characteristics obviously fix the aesthetic properties, because any other statue having identical physical properties will necessarily have the same aesthetic properties; but the aesthetic properties do not likewise determine the physical characteristics, because equally beautiful works of art may have very different physical characteristics.[191] As a second example, consider Rosenberg’s claim that an organism’s fitness, its ability to survive and reproduce, supervenes on properties of its anatomy and physiology, and properties of the environment in which the organism lives. This is clearly a type of asymmetrical supervenient relation. The phenotypic characteristics of the organism and the properties of its environment determine the organism’s fitness, since any other organism with an identical environment and phenotype would have the same fitness. However, the reverse is not true; fitness does not fix the organism’s traits and environment, since two organisms could be equally fit but inhabit different environments, or have different phenotypic traits.

These examples are characteristic of most supervenient relations, in that the ‘property fixing’ can run in only one direction. But examples of reflexive supervenience, in which the determination runs both ways, are not hard to find. Let us consider a couple of these. ‘[T]he property “is copper,”’ Petrie notes, ‘will supervene on the property “has atomic number 29” and vice versa.’[192] Here the ‘property fixing’ runs in both directions. We can regard the property ‘is copper’ as subvenient, since that property fixes the supervenient property

[191] Scott Kleiner helpfully reminded me of this fact in his comments on an earlier draft of this chapter.
[192] Petrie (1987) 127.

'has atomic number 29'; or we may treat the property 'has atomic number 29' as the subvenient property, inasmuch as there can not be two objects alike in having atomic number 29, but differing in respect of their being copper.

Mumford discusses another example, the supervenient relation between the 'disposition . . . of a billiard ball to roll in a straight line on a flat surface when struck' from any direction, [193] and the '"property-complex" of being an equally balanced rigid sphere.'[194] Which of these is the subvenient property that fixes and determines the other: the disposition to roll in a straight line when struck from any direction, or being an equally balanced rigid sphere? The answer, Mumford concludes, is that each 'has as good a claim' to be the subvenient property as the other,[195] because the definitions of supervenience 'fail[ ] to distinguish the subvenient and the supervenient.'[196] As he observes, 'there is indeed no way two particulars could be alike in [being balanced rigid spheres] and yet differ in respect of [their disposition to roll in a straight line when struck from any direction],' nor 'could [two particulars] be alike in possessi[ng the disposition to roll in a straight line when struck from any direction] and differ in [being balanced rigid spheres].'[197] I conclude, then, that reflexive supervenience, in which two objects or sets of properties supervene on each other, is a legitimate type of supervenience.

A second type of supervenience is '*causal* supervenience.'[198] This is the supervenient relationship that exists between two items or events, one of which is the cause of the other. Although the definitions of supervenience 'do[ ] not . . . contain the word "cause,"'[199] causal relations typically satisfy the requirements of supervenient relations. In a causal relation, the effect supervenes on the total causal circumstances that brought it about, because when-

---

[193] Mumford (1994) 427.
[194] Ibid., 428. Mumford distinguishes a billiard ball from a wheel, which also has a disposition to roll in a straight line. Unlike a wheel, a billiard ball 'can roll instantly in any direction, according to where it is struck, whereas a wheel is limited to certain directions.' Ibid.
[195] Ibid., 429.
[196] Ibid., 430.
[197] Ibid., 429.
[198] Searle (1992) 125.
[199] Owens (1989) 61.

ever those causal circumstances are present, the effect must necessarily occur, and any other identical causal circumstances must be attended by the same effect. For example, '[w]e can sensibly say,' writes Mumford, 'that the flame supervene[s] upon striking the match.'[200] Although the causal event he calls 'striking the match' would presumably include other causal factors, such as the presence of oxygen, the presence of those causal factors would necessitate the occurrence of the event he calls the flame; the same causal factors guarantee the same effect, which is the essence of supervenience.

The third type of supervenience is the reductive relation itself. The reduced object supervenes on the constituent particles, because the existence of the particles is sufficient to ensure the occurrence of the higher-order properties of the object, and two indistinguishable collections of material particles must have the same supervenient properties. A cloud is reducible to a mass of droplets suspended in the atmosphere; thus, a cloud is supervenient upon the mass of droplets, because the droplets fix or determine the properties of the cloud.

A supervenient relation alone is not sufficient to constitute a reduction. As Owens puts it, 'Supervenience does not entail Reductionism.'[201] Supervenience tells us only that given one set of properties, another set of properties is guaranteed to occur; reduction, however, adds the requirement that this guarantee be explained. 'Microreductions provide us with an explanatory insight,' Kim writes, 'into how and why the observed regularities hold at the macrolevel.'[202] A reduction of a cloud to an aggregation of ice and water particles, for example, must explain, among other things, how and why a mass of water particles has a fleecy, fluffy, white appearance; unlike reduction, supervenience merely requires that similarity at the bottom level ensures similarity at the top level. Thus, supervenience is compatible with both reductionism and emergentism. An emergent property is a novel causal power that supervenes on an arrangement of particles, but can not be explained in terms of those particles and their interactions. '[W]e should be sensitive to the possibility,' Horgan

[200] Mumford (1994) 426.
[201] Owens (1989) 63.
[202] Kim (1996) 216.

reminds us, 'that . . . it will not be possible to give an account of putative higher-order properties under which their ontological supervenience on the physical could be successfully explained.'[203] In other words, it may not be possible to explain all supervenient relations. 'Supervenience says nothing about the explanatory power of physical theory,'[204] but reductionism requires that an explanation be given of the relation between the reduced and the reducing properties. Indeed, as Brooks notes succinctly, '[r]eductionism is supervenience plus explicability.'[205]

Therefore, supervenience comprises various types of relations, and the claim that one thing is supervenient upon another tells us nothing about the specific type of dependence: reflexive (or identity), causal, or reductive. Each of these types of supervenience differs subtly from the others, and should be distinguished from them. Strictly speaking, reduction is neither a causal relation, nor an identity; and likewise, an identity is not a causal relation. Let us take each of these in turn.

First, reduction is not a causal relation. In a causal relation, the effect is different from the cause; they are, as Hume observes, 'two [separate] objects.'[206] In a reductive relation, on the other hand, a property of an object 'is nothing more than' a property of an arrangement of component parts; accordingly, the property of an object is not separate from the causal power of the interacting particles, as an effect is separate from a cause. For example, the flame is caused by all the circumstances that constitute the event we call striking the match, but the flame is not thereby reduced to the striking match. These are separate events, one of which, we conclude, brings about the other. But in a micro-reduction, the upper-level property is not *caused* by lower-level occurrences: it *is* the lower-level occurrences. Heat is not caused by the molecular motion of atoms; rather, it just is the molecular motion of atoms.

[203] Horgan (1993) 580.
[204] Owens (1989) 61.
[205] Brooks (1994) 804.
[206] Hume (1985) 122.

However, a reduction is sometimes mistakenly treated as a causal relation, because the lower-level events do cause the *effects* that we associate with the upper-level property. This is perhaps a fine distinction, so I want to be clear about this. A property, Kim reminds us, is the 'power' of 'the thing that has it . . . to affect courses of events in its neighborhood.'[207] For instance, a hot object has the power to cause combustion and sensations of heat, but combustion and sensations of heat are not, properly speaking, properties of the object; they are effects of the property. The lower-level events do not cause the property, they cause the effects of the property. Molecular motion does not cause heat, since heat just is molecular motion; instead, molecular motion causes the effects that we associate with heat, namely, combustion, sensations of heat, and an increase in the height of mercury in a thermometer. I conclude, then, that reduction is not a causal relation, though we do expect a reduction to explain how the reducing events can bring about the effects of the upper-level property.

Second, a reduction is not an identity, even though reductions and identities both are types of supervenient relations. The difference, briefly stated, is this: an identity is a symmetrical relation, while a reduction is an asymmetrical relation. 'If reduction is to mean anything at all,' observes Mumford, 'it must be an *asymmetrical* relation.'[208] With respect to an identity relation, we can say, for instance, either that a fiddle is a violin, or that a violin is a fiddle; both statements are true. But in a reduction the subject and predicate can not be exchanged in this way. For example, if a cell can be reduced to a complex arrangement of molecules, then it is true that a cell just is a collection of molecules, but it is not true that an arrangement of molecules is nothing but a cell, because the cell is not ontologically equivalent to the collection of molecules. A 'reduction makes a claim about *what there really is and what there really is not*.'[209] A reduction privileges ontologically the reducing objects. An iden-

[207] Kim (1996) 230.
[208] Mumford (1994) 422, emphasis in the original.
[209] Ibid., emphasis in the original.

tity, however, does not privilege ontologically one of the identical objects; the fiddle and the violin are, so to speak, on an ontological par with each other.

'The "is" [of reduction],' explains Owens, 'denotes realization or constitution rather than identity.'[210] Thus, to say that a cell 'is' a collection of molecules means that the cell is constituted or composed of a collection of molecules, and not that the cell and the collection of molecules are the same thing, as we intend when we say that the morning star is the evening star, or that Erich Weiss is Harry Houdini. In a sense, then, we might say that reduction is a kind of one-way, top-to-bottom identity. Appropriately, Searle calls reduction 'a peculiar form of the identity relation.'[211]

Incidentally, it is interesting to note that the so-called 'identity theory' in the philosophy of mind, according to which the mind is nothing more than brain processes, is in fact a durable misnomer, because proponents of this view assuredly do not wish to assert, what would be true of a genuine identity theory, that the brain is the mind! 'One supposes that psychology may be reducible to physics, but not that physics may be reducible to psychology!'[212] The identity theory, therefore, is more accurately described as a reductive theory of the mind, and when an identity theorist, such as U. T. Place, 'suggest[s] that we can *identify* consciousness with a given pattern of brain activity,'[213] he is presumably using Searle's 'peculiar form' of identification. Place himself makes this clear when he equates the thesis that '[c]onsciousness is a process in the brain' with the claim that '[l]ightning is a motion of electric charges,'[214] a standard illustration of a reduction.

And third, a causal relation is not an identity. The relation between the evening star and the morning star is different from the relation between two events, one of which brings about the other, as the flash of lightning causes the clap of thunder, 'since bringing about

[210] Owens (1989) 61.
[211] Searle (1992) 113.
[212] Oppenheim and Putnam (1995) 408.
[213] Place (1995) 106, emphasis added.
[214] Ibid., 108.

and causing are presumably asymmetric, while [identities] express symmetric relations.'[215] Therefore, because a causal relation involves two distinct events with an asymmetric relation between them, and an identity relation concerns a single object under two names, which necessarily bear a symmetric relation to each other, causal relations and identities must be regarded as different types of supervenient relations.

We can see, therefore, that supervenience comprehends at least three different types of relations. A supervenient relation is one in which *some* determinative relation exists between two things, though it does not specify or explain the nature of that determinative relation. A supervenient relation, Petrie writes, is 'nomologically necessary,' but '[i]t does not entail any more specific connections between the predicates referring to subvenient and the predicates referring to supervenient properties.'[216] Supervenience is, therefore, consistent with non-reductive relations, such as ontological emergence, which Horgan calls 'determinist emergence;'[217] additionally, G. E. Moore argues that non-physical moral properties supervene on physical properties, though the relation between the two levels remains mysterious. Horgan rightly points out that the 'supervenience of higher-order properties and facts on physical properties and facts cannot be enough to confer materialistic respectability.'[218] In order to attain 'materialistic respectability,' the materialist must show that an upper-level object or property can in fact be *reduced* to lower-level physical properties, and not just that the upper-level property is supervenient upon, or determined by, lower-level physical properties. Emergent causal powers that 'get squirted out by the . . . brain,' and mysterious non-physical moral properties that attach to physical events, are both admitted to be supervenient upon lower-level physical properties, but both are anathema to reductive materialism.

[215] Fodor (1995) 430.
[216] Petrie (1987) 126, footnote 10.
[217] Horgan (1993) 559.
[218] Ibid., 565.

In my earlier discussion of the importance of explanation in reduction, I mentioned Brooks' equation, '[r]eductionism is supervenience plus explicability.'[219] I want to suggest now that, given our understanding of supervenience, this claim is incomplete, and for that reason, not quite correct. Brooks is right, of course, that a reduction is a type of explanation, in that it explains how an upper-level property does the things that it does, by showing that it is really the lower-level properties, with which the upper-level property is 'identified,' that does them, such as bringing about combustion in the case of heat. But reduction is not just any old type of supervenience plus explicability. For example, causal supervenience plus explicability is not reduction; we can explain, for instance, how striking a match causally brings about a flame, which flame supervenes on the match-striking event, but we do not thereby show that the flame is nothing more than the striking match. If we wanted to correct Brooks' formula, we should have to say that reduction is *reductive* supervenience plus explicability; but, of course, this equation suffers from an uninformative circularity. We could try to avoid that result by defining reduction as asymmetric identity supervenience plus explicability, but whether this enhances our understanding of reduction I confess some misgivings.

This concludes my discussion of the sixth criticism of ontological emergence, that supervenience entails reductionism. I have been unable to discover any incoherence in the doctrine of emergence, so whether or not ontological emergence is actually a feature of the world seem to be irresoluble by *a priori* reflection. We must consult the facts of the world to decide the matter. Nevertheless, some philosophers are confident that metaphysical reduction is achievable in principle, and that our inability to reduce all the objects of experience to their component parts is merely an epistemological failure: we simply lack the knowledge requisite to explain the objects in terms of material particles. But inasmuch as these reductionists have failed to undermine the possibility of ontological emergence, their assertions are simply not justified. We do not know whether all objects and properties can be reduced

[219] Brooks (1994) 804.

to material particles, and presumptuous claims that reduction *must* be possible merely misrepresent the state of our knowledge. One may believe in materialism without being a reductionist, and one may accept that objects and their properties supervene on material particles without being a reductionist. Ontological emergence is consistent with both of these positions. For all we know, it may well be written into the metaphysical structure of the world that some objects have inexplicable, irreducible, emergent features, and in consequence the theory of universal reductionism must come to ruin.

'[A]rguments [in favor of ontological reductionism] seem hard to find,'[220] but the few that are available can be roughly divided into three categories. First, there are versions of what Jeffrey Foss calls the 'bandwagon argument.'[221] The history of science being replete with examples of seemingly irreducible objects and properties that were eventually reduced, it is just a matter of time, it is alleged, before every intractable object and property yields at length to reduction. One after another, 'the heavens, fire, acid, light, disease, [&c.] have been explained scientifically, and so, it is concluded, the mind will be next.'[222]

Second, there is what I call the assembly argument. According to Oppenheim and Putnam, who propounded this argument in their article 'Unity of Science as a Working Hypothesis,' if an object evolves or is synthesized from an assemblage of parts, then it is nothing more than those parts, nothing else having got into the mix, as it were; and the various sciences establishing that the objects of experience come to be from prior parts, we must conclude that these objects are nothing more than their parts, and therefore that reductionism is true.

And third, there is a causal argument, such as that 'construct[ed]'[223] by Faber, which contends that the only way to explain the causal interaction between objects is to suppose that every upper-level object is in fact nothing more than the parts that compose it. Our or-

---

[220] Garfinkel (1995) 445.
[221] Foss (1995) 407.
[222] Ibid., 408.
[223] Faber (1986) 16.

dinary description of the world refers to various levels of organization: 'Social systems are composed of individual organisms, organisms are systems of organs, organs are made up of cells, cells of molecules, molecules of nuclei and electrons, and so forth.'[224] Faber argues that interaction between objects on different levels is incoherent; a raindrop can not fall on a corporation, 'a hen . . . cannot peck a flock.'[225] Therefore, if an upper-level object is taken to be an irreducible whole, and not simply a 'collective noun,'[226] then the causal relation between that upper-level object and some lower-level object must remain mysterious, because they can not causally affect each other. Non-reductive theories, therefore, terminate in mystery, which can be dissipated only by reducing the upper-level objects to a collection of parts with which the lower-level objects can interact.

Obviously, these arguments could be developed at greater length, and with more sophistication, but enough has been said, I hope, to show that they leave intact the possibility of ontologically emergent properties. With respect to the bandwagon argument, the emergentist may allow that science has accomplished much in the way of reduction, but question whether we can read the future of science in a catalogue of its past successes. Whether or not the mind will continue to resist reduction is not resolved by pointing out that, after all, we did eventually figure out how to explain water. The assembly argument by Oppenheim and Putnam is entirely compatible with ontological emergence, for the fact that an object is composed of matter is insufficient to establish that it does not have emergent properties that supervene on the material parts. Since materialism does not entail reductionism, the reductive materialist must do more than argue that everything, from atoms to artefacts, is made of fundamental particles of matter. Finally, Faber's reminder that irreducible objects are irreducibly mysterious is not so much a criticism of emergence as a succinct description of it. In the end, Faber himself feels compelled to 'advocate[ ] a Cartesian dualism of mind and mat-

[224] Ibid., 13.
[225] Ibid., 17.
[226] Ibid., 20.

ter'[227] to explain the properties of the mind, though he acknowledges that we may 'simply [have to] admit the mystery'[228] of their interaction.

From my examination of the arguments for reductionism and against emergentism, I conclude that ontological emergentism remains a viable position. Whether or not there are in fact emergent features of the world that resist reductionist explanations is ultimately a question that must be answered afresh with each proposed case of emergence.

In the next chapter, I shall turn from the reduction of objects and properties to an examination of issues relating to the reduction of scientific theories.

[227] Ibid., 184.
[228] Ibid., 229.

CHAPTER 3

SCIENTIFIC REDUCTIONISM

The previous chapter addressed ontological issues. Specifically, I sought to determine whether, and to what extent, the familiar objects of experience and their properties can be explained by reducing them to the properties of their interacting component parts, upon which the upper-level properties supervene; or whether these upper-level properties might, in some sense, be regarded as emergent. Consequently, this investigation required an understanding of explanation, reduction, emergence, and supervenience, which I tried to provide.

In this chapter, instead of objects and their properties, I shall be concerned with the reduction of scientific claims, such as scientific laws and theories. Generally speaking, this intertheoretic or scientific reduction attempts 'the reduction of one theory or set of laws to another in the sciences.'[1] In the literature on scientific reduction, the lower-level reducing science is almost always either chemistry or physics, while the upper-level reduced science, which philosophers call a 'special science,' is usually either biology or psychology, and sometimes economics. Although the scientific reductionists and their anti-reductionist opponents typically agree that the objects studied by the special sciences, such as organisms or minds, are made of matter, and 'obey physical laws,' they disagree about the extent to which 'those laws state the sufficient as well as necessary conditions essential to the description and explanation of biological [and psychological] phenomena.'[2] In other words, scientific reductionists and anti-reductionists disagree about whether the laws and theories of

---

[1] Taylor (1971) 45.
[2] Grene (1971) 16.

physics and chemistry can truly explain the laws and theories characteristic of the special sciences.

In order to explain the source and scope of this disagreement, I shall divide this chapter into two parts. The first part will provide an analysis of scientific reduction, including what Ernest Nagel calls the 'formal conditions'[3] for reducing one law or theory to another, and the second part will examine, and ultimately reject, two influential arguments against scientific reductionism.[4]

Analysis of Scientific Reduction

To reduce one thing to another, as we have said, whether it be an object, property, theory, or law, means showing that the thing to be reduced is nothing more than something else, the reducing thing. Sometimes scientific reduction is likened to a kind of translation, 'which . . . enable[s] the physicist to restate in his own terms what the [special scientist] [is]

[3] Nagel (1961) 345.

[4] A brief note about the scope of this chapter. The great bulk of the literature on intertheoretic reduction seems to take for granted the existence of natural laws and of scientific theories composed of such laws: '[I]n the contemporary debate on reductionism,' '[a] theory . . . is taken to be a set of statements, which represent its "laws." It's convenient to think of a theory as constituted by its "basic laws" (or "axioms") plus all statements logically (and mathematically) derivable from them.' Kim (1996) 212. Accordingly, '[i]t is the job of physics to discover exceptionless laws,' writes Brooks (1994) 805, and the question posed in this 'contemporary debate' is whether, and to what extent, the laws of the special sciences can be reduced to those discovered by the physical sciences. It is this debate that I shall be joining in this chapter, though I am aware that some philosophers, with good reason, have questioned the basic assumptions that underlie the current discussion of intertheoretic reduction.

I mention here only a few of the problems that have been raised against the accepted view. First, scientific theories appear to be more than just sets of laws. A physical theory includes instruments, concepts, practices, and theoretical entities, which are as much a part of physical theories as are laws. '[T]he content of science is found not just in its laws but equally in its practices.' Cartwright (1989) 1. Second, theories are perhaps best construed rather as models than as sets of laws. 'It is with . . . models that we are mostly concerned in science, that is, with real or imagined things and processes which are similar to other things and processes in various ways, and whose function is to further our understanding.' Harré (1992) 174. 'The question for a model is how well it "fits" various real-world systems one is trying to represent.' Giere (1999) 93. And one may be 'inclined to think that even where the scientific models fit, they do not fit very exactly.' Cartwright (1999) 48. Third, one may 'question[ ] the usefulness of the concept [of natural laws] for understanding the practice of contemporary science as a human activity.' Giere (1999) 85. It is thought to be a mistake to treat natural laws after the Humean fashion as descriptions of universal regularities, because 'there are no such regularities to begin with,' Cartwright (1999) 71, except 'in the artificial environments of our laboratories.' Cartwright (1999) 37. '[S]trictly speaking, most purported laws of nature seem clearly to be false.' Giere (1999) 90. Instead of regularities, science seeks the 'dispositions, tendencies, powers, and capacities . . . of material systems'; '[w]e experiment on systems to find out how they are disposed to behave.' Chalmers (1999) 218. 'It is capacities that are basic [in science], and laws of nature obtain—to the extent that they do obtain—on account of the capacities [of material systems].' Cartwright (1999) 49.

saying when he offer[s] an . . . explanation for [some phenomena].'[5] Other philosophers speak of theory reduction as theory replacement,[6] because '[a]fter a . . . reduction ha[s] occurred, one theory ha[s] been "eliminated," at least in the sense that it ha[s] been explained'[7] by a more fundamental, lower-level theory; 'one realm of discourse has been shown to be eliminable in favor of another,'[8] because the reducing theory 'enables us to "accomplish the same purposes" as the [reduced theory].'[9]

But in order to capture the asymmetry required by reduction between the reduced and reducing theories, the scientific reductionist must go beyond the claim that the laws and theories of the special sciences can also be expressed in terms of laws and theories in physics or chemistry; the reductionist must add that the lower-level descriptions and explanations provided by chemistry or physics are, in some sense, preferable to those provided by the special sciences. The reductionist privileges the account provided by the lower-level science because of his ontological commitment to reductive materialism, which posits, as we saw in the previous chapter, that the science of 'physics is causally [and explanatorily] complete (i.e., all fundamental causal forces are physical forces, and the laws of physics are never violated).'[10] Thus, the scientific reductionist believes that physics provides a complete description of reality, which leaves out nothing of ontological significance; while the special sciences, he avers, provide a derivative, 'completely redundant'[11] description of the natural world.

Let us be try to be more explicit about the motivations for belief in scientific reductionism. I find in the literature two basic arguments for the claim that reductive materialism entails scientific reductionism. The first argument tries to prove scientific reductionism by means of the doctrine of materialism, whose basic thesis is that reality is composed of mat-

[5] Owens (1989) 59.
[6] E.g., Trout (1995) 390.
[7] Darden and Maull (1977) 60.
[8] Garfinkel (1995) 443.
[9] Ibid., citing Quine (1960) 260.
[10] Horgan (1993) 560.
[11] Foss (1995) 412.

ter; that is, 'the [ontological] domain is fundamentally comprised of the fundamental stuffs of physics, namely, electrons, neutrons, photons, space-time, and so on.'[12] Physics is the science of these things. Therefore, physics is the science of reality. As Alexander Rosenberg puts it, 'the reductionists' conviction that theories are reducible ultimately boils down to their metaphysical belief that things are reducible.'[13] The objects studied by the special sciences are nothing more than arrangements of material particles,[14] whose behaviour is the subject of physics; therefore, by describing and explaining the behaviour of the fundamental particles, physics accounts for the objects composed of, and reducible to, those particles.

A second argument depends upon the claim that physical events and entities are governed by physical laws. All the events or entities described in the special sciences are physical events or entities, governed by the laws of physics; therefore, since the laws of physics govern the events or entities of the special sciences, the behaviour of those events or entities can be explained in terms of the laws of physics. Marjorie Grene presents the argument in these words: 'For since everything is made of matter, . . . the fundamental laws of physics, plus initial conditions and boundary conditions, ought to give us the laws of all systems.'[15] Charles Taylor states the argument in this way: because all objects must 'obey the laws of physics and chemistry . . . [i]t follows that some form of reductivism must hold, that is, that higher level explanations, like the psychological, the sociological, etc., must be ultimately explicable on a more basic level, in terms of physics and chemistry.'[16] If we deny scientific reductionism, the reductionists will insist, then we must admit that there are upper-level ex-

---

[12] Ibid., 420.
[13] Rosenberg (1986) 90.
[14] Some philosophers would qualify this statement to include among the fundamental physical entities such things as electromagnetic fields. See Mary Hesse, 1961, *Forces and Fields* (London: Nelson).
[15] Grene (1971) 24.
[16] Taylor (1971) 57.

ceptions to the operation of the laws of physics and chemistry. But for many people, Taylor suggests, this 'doctrine of exceptionalism seems hard to justify intellectually.'[17]

It might seem that the anti-reductionist can avoid this second argument quite easily. He can admit that although the objects studied by the special sciences do 'obey' and are 'governed by' the laws of physics and chemistry, there are nonetheless higher-level laws in the special sciences in addition to those identified by physics and chemistry, and these additional laws provide explanations that supplement those of physics. For example, '[a]ll living systems do indeed obey the laws of physics,' explains Grene, 'but without countervening [contravening] the laws of physics they may well obey other laws as well.'[18] However, if one is committed to the reductive materialist view that the objects studied by the special sciences are in fact nothing more than collections of particles, then a physical explanation of each particle in the collection provides a complete description of the object. If the whole is nothing more than its parts, and if we can account for each of those parts in terms of physics, then we can account for the whole in terms of physics. There is nothing left over for the special sciences to explain.

Given the apparently close relation between ontological and scientific reductionism, it behooves us to spend a little time clearly distinguishing them. Materialism, which is simply the claim that whatever exists is made of matter, is opposed to the existence of immaterial substances, such as minds, souls, and vital forces, though some 'physicalists' have apparently made their peace with at least some abstract objects. Summarizing the results of the previous chapter, we can describe ontological reductionism in terms of four theses. First, a material object is nothing more than its components and their interactions. Second, an object's macro-properties are explicable in terms of the properties of the interacting components; the macro-properties are 'contained in' the micro-properties, so to speak. Third, properties do not emerge ontologically at upper-levels of organization from the interaction

[17] Ibid., 58.
[18] Grene (1971) 20-1.

of parts at lower levels. And fourth, upper-level 'entities,' such as those studied by the special sciences, are not candidates for ontological recognition, because upper-level entities are merely nominal clusters or groupings of lower-level entities.

Obviously, the primary difference between ontological and scientific reduction lies in the nature of the 'things' that each seeks to reduce: objects and properties versus theories and laws. In a scientific reduction, the relation between the reduced and reducing theories is customarily described in two ways. First, as in the micro-reduction of properties, the reduction of one theory to another 'would in effect show [the reduced theory] to be implicitly *contained in* [the base theory] as a subtheory and not [as] an independent theory.'[19] For example, scientists are satisfied that thermodynamics is contained in, or 'incorporat[ed] . . . within statistical mechanics and the kinetic theory of gases.'[20]

The other way of describing the relation between the two theories in a scientific reduction is to say that the laws of the reduced theory can be *logically derived* or *deduced* from the laws of the reducing theory. If the 'laws of the reduced theory can (more or less) be deduced from the laws of the reducing theory,' we can treat the reduced theory as 'nothing but a special case of the reducing theory.'[21] A classic example of this is the derivation of the gas laws, which relate the pressure, temperature, and volume of gases, from the kinetic-molecular theory of gases. These gas laws, such as Boyle's law, which states that the pressure and volume of a gas are inversely proportional to each other, 'can be understood and verified without introducing any reference to some postulated microscopic structure of [gases] and without assuming that [they] can be reduced to [the laws of] some other theory such as mechanics.'[22] But the gas laws can be derived mathematically from the kinetic theory of gases, which assumes that gases are composed of swarms of tiny, colliding, perfectly elastic molecules, so the kinetic theory 'fully suffice[s] for the description and explanation of

[19] Kim (1996) 213, emphasis added.
[20] Nagel (1961) 342.
[21] Searle (1992) 114.
[22] Nagel (1961) 343.

the [behaviour of gases].'[23] Thus, the reduced upper-level theory is shown to be unnecessary, redundant, superfluous, because the characteristics of gases represented by the gas laws can also be described and explained by the kinetic-molecular theory of gases, which theory is preferable to the gas laws, since it refers to more fundamental entities.

Therefore, if the laws of the special sciences can be logically derived from the laws of chemistry and physics, then the special sciences are shown to be nothing more than chemistry and physics. The scientific reductionist, who regards 'biological systems [as] nothing other than chemical ones,'[24] is 'eager to see as much nonmolecular biology as possible absorbed by the theoretical structure erected in molecular biology.'[25] And by reducing the laws of the special sciences to the lower-level laws of physics and chemistry, that will 'leave[ ] at most [only] the most fundamental laws of [the reducing theories] in need of possible further explanation.'[26] For the reductionist, the special sciences, which 'treat composite entities as units[,] will always be needed in practice,'[27] because it is impracticable to express all of the interesting generalizations about the world in terms of the behaviours of micro-particles, but if the laws of the special sciences can be shown to be special cases of the laws of chemistry and physics, then the special sciences are most assuredly not 'necessary in principle.'[28]

In conclusion, ontological reduction is the attempt to reduce material objects to collections of particles; on this view, upper-level objects are nothing more that nominal groupings or clusters of particles, to which we assign names, for the sake of utility, convenience, and our personal interests. The causal powers of these objects to affect other things in the world can be explained in terms of the properties of the interacting constituent particles. Scientific reduction is the attempt to reduce a scientific theory or law to the some other scientific theory or law, which usually means the reduction of some law in the special sciences

---

[23] Kim (1996) 215.
[24] Rosenberg (1986) 69.
[25] Ibid., 72.
[26] Girill (1974) 224.
[27] Faber (1986) 8.
[28] Ibid.

to a law in either physics or chemistry, by showing that the reduced law can be derived from the reducing law. The terms and concepts that appear in the laws of the special sciences are not included in an inventory of the world's ontologically real things. They are nothing more than groupings of objects that share similar features. Once we have admitted the truth of ontological reductionism, then scientific reductionism seems to follow as a matter of course, because if the objects of experience are nothing more than collections of interacting particles, whose individual behaviours can be explained by the laws of physics and chemistry, then there is nothing left over, so to speak, for the special sciences to explain, because the best explanations cite substantively real entities and processes. The upper-level sciences are nothing more than special cases of the lower-level sciences.

Formal Conditions of Reduction

How, then, does one reduce a theory? What are the formal 'conditions for reduction?'[29] As everyone seems to agree, the classic statement of scientific reduction, from which all discussions of reduction begin, is contained in chapter eleven of Ernest Nagel's *The Structure of Science*. Although it is common to observe that 'the set of conditions deemed adequate for reduction . . . remains a controversial issue,'[30] Nagel's model of inter-theoretic reduction provides considerable and estimable clarification of the issues. For Nagel, scientific reduction, as we have seen, is a type of deduction. In Nagel's words, 'a reduction is effected when the experimental laws of the [reduced] science . . . are shown to be the logical consequences of the theoretical assumptions . . . of the [reducing] science.'[31] Therefore, the reduction of a scientific theory, which Nagel understands as 'a set of experimental laws,'[32] involves the two fundamental requirements of any deduction. Nagel identifies these as the 'condition of connectability' and the 'condition of derivability.'[33]

[29] Nagel (1961) 338.
[30] Bickle (1995) 36.
[31] Nagel (1961) 352.
[32] Ibid., 338.
[33] Ibid., 354.

First, the condition of connectability requires that 'all of the theoretical terms of the reduced theory must be definable by means of the theoretical terms of the reducing theory,'[34] because, as Rosenberg explains, 'one set of statements cannot be derived from another unless the terms in which they are expressed have the same meaning. In logic, the failure to meet this requirement is known as the fallacy of equivocation.'[35] According to the second requirement, the condition of derivability, 'all the laws [to be reduced] . . . must be logically derivable'[36] from laws in the reducing science.[37] Therefore, combining these two conditions, we can say that a reduction of one theory to another occurs only when 'the basic terms (and entities) of one theory are related to the basic terms (and entities) of the other . . . and the axioms and laws of the reduced theory are derivable from [the axioms and laws of] the reducing theory.'[38]

This concept of reduction as deduction is easy enough to understand. If it can be shown that a statement follows deductively from, or is entailed by, a set of other statements, as the conclusion of a syllogism follows deductively from the major and minor premises, then the reduced statement is shown to be nothing more than the set of reducing statements, in the same way that the conclusion is just an explicit expression of what is contained implicitly in the premises. Thus, the reduced statement expresses nothing that is not already contained in the reducing statements. To recur to our previous example, if the gas laws are reducible to the molecular-kinetic theory of gases, then given the various laws that constitute that theory, one can deduce the several gas laws, because they are nothing more than explicit statements of consequences implicit in the kinetic theory of gases.

---

[34] Commoner (1979) 163.

[35] Rosenberg (1986) 91. Nagel (1961) 352-53 explains that the condition of connectability 'is based on the familiar logical canon that . . . no term can appear in the conclusion of a formal demonstration unless the term also appears in the premises.'

[36] Nagel (1961) 354.

[37] In addition, in order to reduce laws of the special sciences to those of more fundamental sciences, it will frequently be necessary to make artificial assumptions about the special sciences. For example, '[i]f we assume the force of small bodies near the earth to be zero, close approximations to Galileo's principles follow deductively from Newton's. If we assume the speed of light to be infinite, the mathematical formulae of Newton's laws may be inferred from Einstein's special theory of relativity.' Rosenberg (1985) 91.

[38] Schaffner (1967) 138.

This reduction illustrates one of the central issues in discussions of inter-theoretic reduction: how to satisfy the condition of connectability when terms that appear in the laws to be reduced do not similarly appear in the reducing laws. For example, Boyle's law relates the temperature and volume of a gas, but 'temperature' is not part of the theoretical vocabulary of the kinetic-molecular theory of gases. Therefore, 'the reducing theory must be supplemented by statements which take us from terms in the laws and postulates of the reducing theory to the terms essential but peculiar to the reduced theory.'[39] In order to deduce Boyle's law from the molecular-kinetic theory of gases, we must supplement that theory of gases with 'the additional postulate that the temperature of a gas is proportional to the mean kinetic energy of its molecules.'[40] Such postulates, which connect terms in the reduced theory with those in the reducing theory, are typically called 'bridge laws' or 'bridge principles,' because they bridge the terminological gap between the reduced and reducing theories.

However, the status of these bridge laws, or 'connecting principles,'[41] has been the source of much controversy in discussions of scientific reductionism. Although most writers on the subject are agreed that principles 'correlating the two vocabularies'[42] are necessary,[43] because 'meaning must remain invariant in the course of particular reductions,'[44] there is no consensus about how to answer 'the well-known problem of specifying the logical status of bridge laws.'[45] As Sarkar puts it, '[p]robably no aspect of reduction has been as controversial as the nature of the connections between the reduced and reducing theories.'[46]

The first 'question [that must be asked] is whether the[se] connections are conventional or factual.'[47] I take it that this question should be answered in favor of their factual

[39] Ager, Aronson, and Weingard (1974) 119.
[40] Nagel (1961) 355.
[41] Ager, Aronson, and Weingard (1974) 129.
[42] Kim (1996) 213.
[43] Ager, Aronson, and Weingard (1974) argue that bridge laws are not necessary for reduction, but only because they regard scientific reduction as a type of incorporation, and not as deduction.
[44] Brittan (1970) 451.
[45] Bickle (1992) 54.
[46] Sarkar (1998) 32.
[47] Ibid., 33.

status. As Kenneth Schaffner recognizes, bridge principles must 'have empirical support,'[48] because, as others have pointed out, '[t]o the extent that they are conceived of as conventional truths, . . . they introduce arbitrariness'[49] into scientific reduction. The reduction of one law to another is noteworthy and informative only if the terms in the reduced law are justifiably identified with terms in the reducing theory. For example, that the gas laws can be derived from the molecular theory of gases is a significant finding only if temperature really is mean kinetic energy of the molecules. If bridge principles are treated instead as mere conventions, or stipulations designed to make possible the derivation of one law from another, then 'there seems no good [scientific] motivation for going to such trouble.'[50] 'Hoked-up' connections between terms, as Rosenberg calls them, tell us nothing about the real relation between scientific laws and theories. Therefore, the reductionist's assertion 'that all scientific laws, including those of biology, can be translated into and are derivable from basic, universal laws of matter in motion'[51] is rendered vacuous unless those translations depend upon scientifically justified connections between theoretical vocabularies.

In the previous chapter, I argued that the micro-reduction of an object's upper-level property to the micro-structure of the object requires that the reductionist be able to explain how the micro-structure can cause the effects we associate with that property. If heat is identified with the kinetic energy of moving molecules, for instance, then the motion of the molecules must explain the characteristic effects of hot objects on other things in the world. This same requirement must be satisfied with respect to the bridge laws we employ in scientific reductions. Although it may be the case that the gas laws 'could be deduced from the [molecular theory of gases] *if* the temperature were in some way related to the mean kinetic energy of the molecular motions,'[52] this identification would still not be justified unless, as we said, the molecular motions could explain the effects we associate with temperature. In

[48] Schaffner (1967) 140.
[49] Ager, Aronson, and Weingard (1974) 121.
[50] Rosenberg (1986) 95.
[51] Grene (1971) 16.
[52] Nagel (1961) 344, emphasis in the original.

order to effect a reduction of a scientific law, the meaning of the terms that appear in that law must remain invariant through their identification with terms in the reducing theory. But if, for instance, molecular motions can not causally account for the behaviours of hot objects, then what we mean by temperature in the gas laws has not carried over to the identification of temperature with molecular motions; the meaning has not remained invariant, because the reduction has not preserved the meaning of temperature. Therefore, the ability of a bridge law to facilitate the derivation of one law from another is not in itself reason to treat the bridge law, and the reduction that depends on it, as successful and scientifically respectable.

A second source of controversy concerns whether bridge laws should be regarded as definitions. At first glance, it would seem obvious that the translation of the theoretical terms required by reduction is an unproblematic case of definition: terms in the reduced law or theory are simply defined by means of terms supplied by the vocabulary of the reducing theory. A problem arises, however, because we often treat definitions as 'analytic relations between the meaning of expressions, [and] not . . . empirically motivated and warranted correlations between the properties signified by them.'[53] Thus, when we say that some statement is true by definition, we typically mean that it is a necessary statement whose truth does not depend upon verification or justification. A child may sensibly question, for example, whether unicorns really exist, but it would be non-sensical to ask whether a unicorn is truly a one-horned horse, as if some evidence might come to light that would overthrow the definition. But bridge principles are not analytic truths, since, as Rosenberg observes, 'they must pass tests of coherence and compatibility with the remainder of science that are established independently of any claim that one particular law follows from another.'[54] Unlike definitions, bridge laws are, at least in principle, revisable in the light of

[53] Kim (1996) 213.
[54] Rosenberg (1986) 92.

new information. Therefore, we have good reason for denying that bridge principles are definitions, in the strict sense of necessary, analytic statements.

A third question is whether bridge laws are identities: 'Must bridge laws be *biconditionals* in form—that is, "iff" (↔) statements—rather than one-way "if-then" (→) conditionals?'[55] Most philosophers seem to assume that these laws are identities. For example, Schaffner regards reduction as requiring 'a one-to-one correspondence representing synthetic identity'[56] between terms in the reducing and reduced theories; this reduction identity, he explains, 'is in many ways analogous to the well-known synthetic identity: morning star = evening star.'[57] John Bickle describes bridge laws as 'contingent identities,' in which an upper-level event is 'contingently identical' to some lower-level state.[58] J. D. Trout writes that 'bridge laws are typically expressed as identity statements, such as "temperature = mean kinetic energy."'[59] And Jaegwon Kim, summarizing the literature, concludes, '[c]urrent discussions of reductionism . . . standardly assume that these [bridge] laws . . . must be biconditionals in form, giving for each property to be reduced a nomologically coextensive property in the base theory.'[60]

But why are these bridge laws typically treated as identities? What accounts for the reductionists' 'apparent infatuation with the biconditional?'[61] The answer is that scientific reductionists are concerned about more than the derivation of one law from another, which is the sole aim of Nagel reduction. As Sarkar notes, reductionists also have 'ontological concerns,'[62] which require that the bridge laws be identities. 'Such identities,' Kim explains, 'are [arguably] essential to the ontological simplification that we seek in theory reduction, for they enable us to dispense with [upper-level terms] as something in addition to [the terms

---

[55] Kim (1996) 214.
[56] Schaffner (1967) 144.
[57] Schaffner (1969) 329.
[58] Bickle (1992) 50.
[59] Trout (1995) 387.
[60] Kim (1996) 214.
[61] Sarkar (1998) 36.
[62] Ibid.

that appear in the lower-level theory].'[63] By reducing an object, property, theory, or law, we show that it is redundant, that it is nothing more than what it has been reduced to. So requiring identities in the bridge laws is simply a means of 'eliminati[ng] and replace[ing] the entities or predicates of the reduced theory by those of the reducing theory.'[64]

But there are at least two problems with this requirement that bridge principles be identities. First, in many cases, we may be able to translate an upper-level term into the vocabulary of the bottom-level theory, without being able to achieve the converse translation. For example, we can say of a gene[65] that it is a sequence of DNA, but it is 'notoriously difficult'[66] to specify the properties that a sequence of DNA must have to permit us say that any sequence having those properties is a gene. Moreover, science may not regard each of these translations as equally significant and explanatory. For instance, the translation of talk about genes into the language of molecular genetics may be far more important to scientists seeking the biochemical basis of disease, than the converse translations, which proceed from DNA molecules to genes.[67] As Sarkar concludes, 'the pursuit of explanation in molecular genetics does not normally require these converses.'[68]

The second problem is one that we encountered in the previous chapter: identifying the terms in the reduced theory with others in the reducing theory does not produce true identities at all. For example, the identification of a gene with a sequence of DNA is not equivalent to the identity between, say, an author and his pseudonym. Charles Dickens and Boz are identical; neither of them is ontologically more real than the other. But the motivation for regarding bridge laws as identities derives from an ontological commitment to the

---

[63] Kim (1996) 215.
[64] Sarkar (1998) 36.
[65] 'The most common sense of *gene* in molecular biology is "a *reading sequence*"—that is, a sequence of nucleotides that is transcribed into a piece of messenger RNA that is either translated into protein or used directly in the metabolism of the cell.' Sterelny and Griffiths (1999) 78.
[66] Sarkar (1998) 35
[67] Of course, we may seek to translate DNA sequences into higher units, such as genes or even phenotypic characters, as, for example, when we want to explain how DNA sequences under certain specifications can be expressed as Mendelian characters over successive generations.
[68] Sarkar (1998) 35.

greater reality, as it were, of the terms in the reducing theory. The DNA molecule is thought to be 'more real,'[69] ontologically speaking, than the gene, which, in the language of molecular genetics, is nothing more than a sequence of DNA. The gene and DNA do not occupy the same ontological plane, as do fiddles and violins, or Dickens and Boz. Bridge laws are not identities; they are, as no one should be surprised to learn, reductions, or, as I characterized them in the previous chapter, one-way identities. It is this reductive relation, and not a strict identity relation, that exists between the terms that appear in bridge laws. The upper-level term is reduced to, or identified with, terms from the reducing theory. Thus, the reduction of one law to another requires the reduction of the terms that appear in that law; and these term-reductions, which we call bridge laws, are nothing more than instances of ontological reduction. Consequently, much of the controversy surrounding these bridge laws would be dispelled if they were properly recognized as reductions, and not mistakenly regarded as definitions or identities.

I mention one example of this widespread confusion. 'The philosopher,' writes Rosenberg, 'treats reduction as a relation between theories and not between things . . . because [for one reason] a precise sense can be given to intertheoretical reduction more easily than to claims that one sort of thing is "nothing but" some more basic kind of thing.'[70] This claim may be criticized on several grounds, but the important point for our purposes is that Rosenberg apparently fails to realize that intertheoretic reductions, which he distinguishes from ontological reductions, can not succeed without ontological reductions. To use our familiar example, the intertheoretic reduction of the gas laws to the kinetic theory of gases depends upon the ontological reduction of temperature to the mean kinetic energy of the molecules. Temperature being a property of a gas, the bridge law identifying temperature with the motion of molecules amounts to an ontological property reduction. Consequently,

---

[69] Faber (1986) 4. According to the reductive materialist, 'ordinary things, human beings among them, acquire a shadowy ontological status because . . . parts are specifiably more real than the wholes they compose.' Ibid.
[70] Rosenberg (1986) 90.

Nagel's 'precise' model of intertheoretic reductions actually requires the sort of 'nothing but' relations between terms that Rosenberg mistakenly believes the philosopher can do without.

I conclude, then, that a proper understanding of bridge laws as ontological reductions will eliminate some of the controversies that have beset them. However, two other issues relating to Nagel's model of intertheoretic reduction remain to be considered: first, the need to 'correct' an upper-level theory before reducing it, and second, the nature of explanation proper to scientific reduction.

Nagel's account of reduction, many have argued, is incomplete, because it fails to recognize that many theories must be corrected before they can be reduced. If the theory to be reduced contains statements that are incorrect or otherwise deficient in the light of the reducing theory, then it would be perverse to attempt to reduce such statements. 'After all, we [do] not want to require the reducing theory to entail propositions already known to be false.'[71] A standard example of this need to correct a theory before it can be reduced is provided by Mendel's laws of genetics. Although 'Mendel was under the impression that the determiners [of heredity] were statistically independent of one another, . . . the[se] determiners (called genes) are for the most part "linked" with other genes.'[72] '[S]ince crossover, recombination, and linkage falsify Mendel's laws,'[73] the laws must be corrected to take into account what we know about the actual process of heredity, before they can be derived from the theory of molecular genetics.

However, the correction of Mendelian genetics necessary for reduction raises two distinct problems. First, if Mendel's theory is changed to accommodate new scientific findings, then what is the relation between Mendel's original theory and the corrected version, which we want to reduce? The answer, most philosophers seem to agree, is that the two theories 'must be related by "strong analogy," that is, they must be sufficiently similar that

[71] Ibid., 111.
[72] Schaffner (1967) 142.
[73] Rosenberg (1986) 111.

the second is just a revision of the first.'[74] 'In an important sense,' Schaffner writes, 'there [must be such] a strong family resemblance between [the two theories,] that one is justified in claiming that [they are] the same theory.'[75] But it is obviously quite difficult to give a very precise account of this 'strong analogy' requirement. How can we determine whether the second theory is only a revision of the first, such that they are really the same theory, or whether the second theory is in fact a new theory that simply incorporates what was correct about the first theory? This is another version of that old intractable philosophical puzzle: how much can a thing be changed before we feel justified in regarding it as a new thing? How many pieces of a ship can be replaced, for example, before it is not the same ship any-more? Or, in this case, how much change can a theory undergo, and still be recognized as the same theory?

But we can set this problem aside. What is important in science and for our understanding of the world is that inadequate theories should be corrected: deficiencies supplied, and falsities removed. Whether the improved theory is regarded as a revision of the old or as a new theory in its own right is unimportant if our primary concern is the reduction of upper-level laws that correctly describe the world. However, in changing Mendelian genetics to render it more susceptible to reduction, we confront our second problem: have we really *reduced* Mendelian genetics, or just *replaced* it with a better theory? 'What seems to be the case,' concludes Gordon Brittan, 'at least in the history of physics, is that theories have replaced, rather than [deductively] reduced, one another.'[76] For example, as Jeffrey Foss points out, because Newtonian mechanics and relativistic mechanics 'do not agree precisely' in their 'empirical content,' 'derivation of Newton's theory from relativity theory is impossible, strictly speaking, even for a restricted domain.'[77] Moreover, the concepts shared by the two theories, such as mass, time, and length, 'are defined by each in a way incompatible

[74] Kleiner (1975) 532.
[75] Schaffner (1969) 332.
[76] Brittan (1970) 453.
[77] Foss (1995) 410.

with the definitions in the other theory.'[78] 'But if, to repeat Nagel's point, the concept[s have] been redefined, then it is not [Newton's theory] that has been [deductively] reduced but something resembling it in certain ways.'[79]

It seems that we have something of a paradox. In order to reduce many theories and laws, we must first correct them, by constructing analogous theories, but if we correct them, then we have not really reduced these theories at all; we have merely replaced them with other theories that we can reduce, theories whose laws can be deduced from the laws of the reducing theories. Thus, many important theories, such as Newtonian mechanics or Mendelian genetics, turn out to be irreducible, even though they have been superseded by later theories. We can resolve the paradox in two ways. We can claim, as Schaffer does in the statement above, that the analogous theories we have constructed are in fact the same as the original theories, or deny that the irreducibility of the original, uncorrected theories has any significance for the progress of science.

The second of these seems preferable to me. In my view, the first option merely amounts to putting an old name on a new theory: Mendel's name is appropriated for our latest understanding of heredity, Darwin's name is attached to our current theory of evolution. And much confusion results. Disputes needlessly arise regarding the degree of similarity between the new theory and the original theory. Consider, for example, the vast amount of intellectual effort that has been expended trying to determine whether Darwin's original theory of evolution is consistent with Gould and Eldridge's theory of punctuated equilibrium, or whether an acceptance of punctuated equilibrium necessitates a revision or correction of Darwin's thought. Such debates, Rosenberg writes, 'have done modern biology a profound disservice,' because 'whether and to what extent Darwin [himself] was a "gradualist" . . . is not a biological question, but a biographical one.'[80]

---

[78] Ibid., 411.
[79] Brittan (1970) 453.
[80] Rosenberg (1985) 158.

It seems to me that the philosophical concern with strong analogies in theory reduction is misguided and unhelpful. Nagel was right, I think, to exclude theory correction from his account of theory reduction. We reduce those theories whose laws can be derived from the laws of some other theory, and if the theory to be reduced is incomplete or incorrect, such that its laws can not be derived from the laws of another theory, then we simply eliminate it as a candidate for reduction. For example, Mendel's laws have much in common with our current understanding of heredity—one dictionary of biology calls them 'the foundation of genetics'[81]—but they are inaccurate nonetheless. Consequently, they are not strictly reducible to molecular genetics. If we can formulate more accurate higher-level laws of heredity, though some philosophers doubt that biology actually contains any laws,[82] then these laws will be proper targets of reduction. However, the relation that exists between these laws and any previous laws of heredity, such as those proposed by Mendel, is ultimately irrelevant to the scientific understanding of the world. Whether or not there is a strong analogy between two theories, one of which improves upon the other, is important only for the historian. The scientific reductionist is uninterested in any similarity between the theory to be reduced and some previous theory, however great that similarity.

I turn now to the issue of explanation in scientific reduction. As we have said, reduction is a form of explanation. The reducing theory explains the laws of the reduced theory. '[I]n genetics there are two clearly identifiable theories, bodies of law, expression of regularity—one molecular and one nonmolecular, Mendelian.'[83] The reduction of nonmolecular genetics to molecular genetics provides an explanation of 'those [upper-level] regularities . . . at the level of biochemistry.'[84] The reduction of the gas laws to the molecular theory of gases explains, for example, why the pressure and volume of a gas are inversely proportional to each other, as represented in Boyle's law. The question I want briefly to consider

---

[81] *A Dictionary of Biology*, 1996, (Oxford: Oxford University Press), 314.
[82] Kleiner (1975) 524 cites J. J. C. Smart as a proponent of the view 'that biology proper contains no laws.' Brooks (1994) 805 agrees: '[B]iology has rather few laws and they have exceptions.'
[83] Rosenberg (1986) 90.
[84] Ibid.

here is whether explanation in scientific reduction differs from explanation in ontological reduction.

The ontological reduction of an upper-level property involves two steps. First, the upper-level property is *identified* with some characteristic of the collection of constituent parts. And second, this identification is *justified* by showing that the constituent parts are capable of producing the effects that we associate with the upper-level property. For example, we explain the whiteness of a cloud by identifying the color with the ability of the water droplets that constitute the cloud to appear white, and by showing that water droplets suspended in the atmosphere are capable of reflecting all wavelengths of light, thereby giving rise to the appearance of whiteness. Or we explain the heat of a body by identifying its temperature with the mean kinetic energy of the molecules that make up the body, and by showing that the vigorous movements of the molecules can produce the combustion and the effects on thermometers that we associate with hot objects.

But does explanation in scientific reduction involve both of these steps? Writers on scientific reduction certainly do not describe the process in such terms. But this is hardly surprising, since, it has been observed, 'the concept of explanation [involved in reduction] has not been elaborated at any length.'[85] Following Nagel, we are typically told that 'to explain a law or theory is to derive it . . . from other laws and theories.'[86] Explanations and reductions are thus logical equivalents. Therefore, reductions explain because they are explanations! But this will hardly be satisfying to one who seeks an understanding of how the reduction of an upper-level law explains that law.

Let us try to be clearer about the relation between scientific reduction and explanation. One way to explain something is to tell what it is. We explain a cloud by saying that it is a mass of suspended water droplets; we explain sunspots by identifying them as solar magnetic storms. These are examples of ontological reductions: identifying one thing with

[85] Brittain (1970) 454.
[86] Ibid., 447.

something else. But in order to justify these identifications, or one-way identities, if you will, we must show that the reducing thing is capable of producing the effects that we associate with the reduced thing. We must show, for instance, that solar magnetic storms can account for the darkness of sunspots.

In scientific reduction, we explain an upper-level law by showing that it can be derived from lower-level laws. This is another way of saying that the upper-level law is just an explicit statement of a generalization that is contained implicitly in the lower-level laws, as a conclusion is contained implicitly in the premises of a deductive argument. Once we are in possession of these premises or lower-level laws, we can derive or unpack the conclusion or upper-level laws. In deriving the gas laws from the molecular theory of gases, for example, we show that they are nothing more than implications of the molecular theory of gases. Thus, in summarizing scientific reductionism as 'the view that the laws of higher-level theories . . . can be explained in terms of, and thus replaced by, those of some lower-level theory,'[87] we mean that these higher-level laws are explained in the sense that they follow from, or are entailed by, the laws of the lower-level theory.

But this seems different from the explanation proper to ontological reductions. First of all, in an ontological reduction, it makes no sense to speak of logically deriving one object or property from another: the sunspot does not follow logically from the solar magnetic storm, the mean kinetic energy of moving molecules does not logically entail temperature. The concepts of deduction, entailment, and derivation express a logical relation between statements and sentences, not objects and properties. And second, explanations in ontological reduction involve two steps—identification plus justification—whereas Nagel reductions seem to require only one. The derivation of upper-level laws from lower-level laws both identifies the upper-level laws with those at the lower-level, and shows that this identification is justified.

[87] Trout (1995) 387.

Let me illustrate this distinction. In order to reduce a property, such as the viscosity of a liquid, we identify the property with some characteristic of the particles that constitute the liquid, in this case, the strength of the bonds between the molecules; and we justify that identification by showing how the strength of the molecular bonds explains the effects we associate with viscosity; that is, bond strength explains viscosity because, to put it simply, the stronger the bonds between the molecules, the more the fluid acts like a solid and thus resists flowing. Without this justification, the identification is nothing more than a stipulation, and thus we have no reason to believe that viscosity is really nothing more than molecular bond strength.

But now consider the well-worn reduction of the gas laws to the molecular theory of gases. According to Nagel's account of intertheoretic reduction, reducing a law means deducing it from other theoretical laws. If we can deduce Boyle's law from the molecular theory of gases, together with necessary bridge principles to match up the terms, then we have successfully reduced Boyle's law. Deducing an upper-level law from lower-level laws provides justification for the claim that the upper-level law is contained in the lower-level theory. Therefore, once we have shown that Boyle's law can be deduced from the molecular theory of gases, then Boyle's law is justifiably identified with the reducing theory, as a special case of that theory. No justificatory work beyond the deduction remains to be done. It would seem, then, that in a scientific reduction we explain an upper-level law simply by deducing it from a lower-level law.

To summarize, in an ontological reduction, we identify the upper-level object or property with lower-level phenomena, and we justify that identification by showing that the lower-level phenomena can cause the characteristics that we associate with the object or property to be reduced. In a scientific reduction, we logically deduce the law to be reduced from the reducing laws, and this deduction both identifies the law to be reduced with the reducing laws, and justifies that identification. In an ontological reduction, explanation consists of showing a causal relation between the lower-level phenomena and the effects of the upper-level property; but scientific reduction seems to equate explanation simply with deduction.

I want to suggest that this distinction between explanation in ontological reduction and that in scientific reduction is more apparent than real. The distinction vanishes upon a closer consideration of scientific reduction. What needs to be added to our account is the justification of the bridge laws that are essential to most intertheoretic reductions. As I have argued, bridge laws are ontological reductions. The identification of temperature with mean kinetic energy of moving molecules, necessary for the scientific reduction of some of the gas laws, is an ontological property reduction. This reduction of temperature to mean kinetic energy is not justified merely because it makes possible the derivation of the gas law from the molecular theory of gases. Many a meaningless conclusion could be logically deduced from a given set of premises if one were permitted to stipulate, without justification, that some term in the conclusion was to be identified with some term in the premises. In order to have confidence in a conclusion that is drawn from the premises, we need reasons to believe that the terms that appear in the conclusion mean the same as the terms in the premises, that temperature, for instance, really is nothing more than the mean kinetic energy of molecular motions.

Once we include bridge principles in our account of scientific reduction, we find that that explanation in scientific reduction involves the same two steps as explanation in ontological reduction. First, in order to explain a law, we must identify that law, and the terms unique to it, with laws and terms in the reducing theory. For example, Charles' law, which states that the volume of an enclosed gas is directly proportional to the temperature of that gas, is explained by identifying the law as a special case of the laws of the molecular theory of gases, and by identifying temperature with the mean kinetic energy of the moving molecules.

Second, this identification of laws and terms must be justified, and not merely stipulated. In the case of the laws, this identification is justified, as we have seen, by means of deduction; by deducing the law to be explained from lower-level laws, we justify the claim that the upper-level law is nothing more than a special case of the lower-level laws. The justification of the bridge laws, on the other hand, proceeds not by deduction, but by showing how the lower-level phenomena, with which the upper-level terms are identified, can cause the effects we associate with the upper-level terms. Deduction does not apply to ontological reductions, which are concerned with properties and objects, not statements. For example, the justification of the bridge law identifying temperature with the kinetic energy of moving molecules depends on showing that molecular motions can causally explain the effects of temperature on other things, such as thermometers.

To summarize, explanation in scientific reduction is not equivalent to deduction, because scientific reduction includes bridge laws, which are neither explained nor justified by deduction. We explain the gas laws by reducing them, that is, by showing that they are nothing more than special cases of the molecular theory of gases; and this reduction requires two steps: identifying the upper-level laws and terms with the laws and terms of the reducing theory, and justifying that identification by means of deductive explanation for the laws and causal explanation for the bridge principles.

This concludes my discussion of explanation in scientific reduction, and with it my analysis of scientific reduction. In part two of this chapter, I shall discuss at length two separate arguments against scientific reductionism, made by anti-reductionists who defend the irreducible autonomy of upper-level scientific laws and theories.

Arguments Against Scientific Reductionism

Scientific reduction is a type of reduction concerning scientific theories and laws; scientific reductionism, however, is the claim that *all* upper-level scientific laws are reducible to laws at the lowest level; 'all the special sciences reduce to physics.'[88] Accordingly, for the scientific reductionist, there is '*one* lowest-level set of laws into which all others are translat-

[88] Fodor (1995) 429.

able and from which they can be derived,'[89] and 'any laws other than the laws of physics' are redundant.[90] The anti-reductionist, while accepting the possibility of scientific reduction, denies the thesis of scientific reductionism. He accepts that some laws, such as the gas laws, may be derivable from the laws of a lower-level theory, but rejects the claim that all higher-level laws are reducible to the laws of physics.

A. The Irreducibility of Natural Kinds

The anti-reductionists offer a variety of arguments against scientific reductionism, but two of them being especially prominent, well-developed, and widely discussed in the philosophical literature, I propose to confine my discussion to these. I begin with an argument made famous by Jerry Fodor in his 'landmark'[91] article 'Special Sciences.' It is this argument, Bickle writes, that 'provides perhaps the main reason why psychophysical reductionism is out of fashion in the philosophy of mind.'[92]

Although he rejects scientific reductionism, Fodor makes certain concessions to the scientific reductionist. He gives 'his opponent [Nagel's] "bridge law" account of intertheoretic reduction,'[93] and '[t]aking a stand deep in territory usually thought to be held by reductionism,'[94] accepts 'that all the events that the sciences talk about are physical events,'[95] a position he calls token physicalism. This is the thesis, as Faber summarizes it, that '[e]very specific instance of a real thing or process is thoroughly physical; that is, nothing happens in any process that is not compatible with basic physical law or not in principle predictable by means of these laws from antecedent physical conditions.'[96] The burden of Fodor's argu-

[89] Grene (1971) 16.
[90] Foss (1995) 412.
[91] Bickle (1992) 48.
[92] Ibid., 47.
[93] Ibid., 49.
[94] Faber (1986) 123.
[95] Fodor (1995) 431.
[96] Faber (1986) 124.

ment, then, is to show that 'token physicalism is weaker than [and thus does not entail] reductionism.'[97]

The heart of Fodor's argument against scientific reductionism is the claim that proper bridge laws can not be established between the laws of the reduced theory and the laws of the reducing theory. Since 'bridge laws are what provide the crucial reductive linkages,' Kim writes, 'reasons for thinking that such . . . bridge laws cannot be had would yield a powerful argument against the possibility of reducti[on].'[98] Fodor tries to provide such reasons. His general argument against scientific reductionism can be briefly stated. If the laws and theories of the special sciences, such as biology and psychology, are to be reduced to physical laws and theories, then the terms that figure in those special science laws and theories must be identified with terms in physics, via bridge laws; but the required bridge laws can not be formulated; therefore, scientific reductionism is false.

The first premise of this argument is merely a restatement of the familiar requirement of Nagel reduction 'that the reduction of . . . the special sciences proceeds via bridge laws which connect their predicates with those of [the] reducing theories.'[99] In order to derive a law of the special sciences from physical laws, the terms unique to the special-science law must be translated into the language of the reducing theory. For example, the derivation of Charles's law from the molecular theory of gases requires translating temperature, which term does not appear in the molecular theory of gases, into the mean kinetic energy of the moving molecules; and the reduction of nonmolecular genetics to molecular genetics involves translating genes into the language of biochemistry.

It is the second premise denying the existence of appropriate bridge laws that requires defense. Fodor tries to provide that defense by means of two arguments, which I shall call the multiple realizability argument and the natural kinds arguments. We begin with the multiple realizability argument.

---

[97] Fodor (1995) 431.
[98] Kim (1996) 217.
[99] Fodor (1995) 430.

There are two ways of creating the bridge laws necessary for the reduction of one theory or law to another. First, we can find a term in the vocabulary of the lower-level science that corresponds coextensively with the term to be translated in the upper-level science. But this does not seem to be a promising means of creating bridge laws, because of the hierarchical structure of science. As Faber observes, 'we distinguish [upper and lower] levels [in science] according to the part-whole relation; lower-level laws qualify as such because they refer to the parts of upper-level things.'[100] This follows from the thesis of token physicalism: upper-level things are composed of lower-level parts. Therefore, since the objects referred to in upper-level laws and theories are composites of objects referred to in lower-level laws and theories, lower-level terms can not be coextensive with upper-level terms. For example, in the hierarchy of genetics, genes occupy a higher level than the nucleotides that compose them, and nucleotides likewise are found on a higher level than their molecular components. Consequently, there are no coextensive objects in molecular genetics that correspond to genes.

The second way of establishing bridge laws, which seems more hopeful, is to identify equivalents between upper-level terms and *groups* of entities at the lower level, which can stand for the upper-level terms. Accordingly, we want to find a *set* of lower-level objects that we can substitute for any instance of the upper-level term in a law or theory. This is possible for some upper-level terms. For instance, we can substitute H_2O for the upper-level term water, and each of the five nucleotides, which strung together make up DNA and RNA, can be replaced by a precise chemical description of their molecular components. These substitutions are possible only because each instance of the upper-level term has an identical physical base. Each molecule of pure water is alike in its chemical composition, and every thymine base in every DNA molecule is chemically indistinguishable from every other thymine base.

---

[100] Faber (1986) 125.

But it is not possible to provide such substitutions for many upper-level terms, because instances of those terms are not identical at the physical level. The instances of many upper-level terms may be realized in many different physical substrates. Philosophers describe these instances as being multiply realized. Examples are abundant. Kim mentions 'the widely accepted thesis that any mental [state] can have diverse physical realizations in a wide variety of biological organisms.'[101] Faber doubts whether there is any 'physical, chemical, or structural property possessed by all meteorites and only them.'[102] And Fodor observes that monetary exchanges, which figure in economic laws, differ in their 'physical description,' since '[s]ome monetary exchanges involve strings of wampum. Some involve dollar bills. And some involve signing one's name.'[103]

These upper-level scientific terms 'group[ ] together certain individuals or processes'[104] according to what is common to them at one level; but the individuals or processes may 'share little or nothing in common'[105] at the physical level. 'The point is,' Fodor concludes, 'that monetary exchanges have interesting things in common . . . [b]ut what is interesting about monetary exchanges is surely not their commonalities [at the physical level].'[106] There is no physical equivalent that can stand for every instance of a multiply realizable upper-level term. For example, because, as many philosophers argue, 'any mental state, say pain, is capable of "realization" . . . in widely diverse neural-biological structures in humans, felines, reptiles, mollusks, and perhaps other organisms further removed from us,'[107] there is no single neural-biological structure that can be substituted for every instance of that mental state. And if no lower-level equivalent can be found for the upper-level term, then the laws and theories that contain that term can not be reduced to the lower-level theory.

[101] Kim (1996) 218.
[102] Faber (1986) 126.
[103] Fodor (1995) 433.
[104] Longuet-Higgins (1970) 238.
[105] Bickle (1992) 48.
[106] Fodor (1995) 433.
[107] Kim (1992) 1.

Therefore, Fodor concludes, scientific reductionism, which states that all upper-level laws can be reduced to physical laws, must be false.

The 'obvious' reductionist response to this argument involves what Kim calls 'the disjunctive move.'[108] Since, according to Fodor's thesis of token physicalism, each particular instance of an upper-level term can be reduced to the physical level, then we can get a translation of the upper-level term by creating a disjunctive set of all the physical bases necessary to realize all the members of the upper-level term. The disjunctive bridge law connects an upper-level term with a disjunction of all the physical realizations of instances of that term. For example, if there is a distinctive neural state that instantiates pain in each type of organism, then the upper-level term pain can be identified with the disjunction of the various neural states that realize pain; thus, on this account, 'pain' is nothing more than, and can be translated as, *either the particular human neural state, or the particular reptile neural state, or the particular feline neural state, or* . . . Wherever the upper-level term pain appears in any law or theory, we can substitute this admittedly cumbrous disjunctive as an exact equivalent. This translation will overcome the barrier to reduction presented by multiple realizability.

At this point, the question we confront is the 'propriety'[109] of using such disjunctives as predicates in bridge laws. As we might expect, a number of anti-reductionist objections have been leveled against this 'disjunctive move.' E. A. Gellner complains that '[s]uch translations [are] clumsy, nebulous, long and vague, where the original [term] . . . was clear, brief and intelligible.'[110] Kim mentions the criticism that the members of these disjunctive sets 'show [little or] no resemblance,' although 'resemblance or similarity is the very core' of scientific terms.[111] And John Godbey points out that these disjunctions are often 'open-ended'; for example, 'a mental state . . . can only be identified with a disjunction of physical

[108] Ibid., 8.
[109] Kim (1996) 219.
[110] Gellner (1956) 162.
[111] Kim (1996) 219.

states, [but] . . . since we do not know what all these physical states are, the disjunction would be essentially open ended.'[112]

These criticisms, by themselves, are not particularly worrisome for the reductionist. Even if these disjunctive translations *are* long, nebulous, vague, open-ended sets of dissimilar members, the anti-reductionist needs to provide some reason to think that such sets can not appear in a scientific law. Fodor aims to provide this reason by means of what I call his natural kinds argument.

Briefly stated, Fodor objects to the use of disjunctive sets in bridge laws because, he claims, disjunctive sets are not proper scientific terms. A proper scientific term, according to Fodor, is a natural kind, which he defines in terms of a scientific law: 'the kind predicates of a science are the ones whose terms are the bound variables in its proper laws.'[113] As Kim explains, 'a given predicate *P* is a "kind predicate" of a science just in case the science contains a law with *P* as its antecedent or consequent.'[114] Therefore, unless a disjunctive set is either the antecedent or consequent of a proper scientific law, then it is not a natural kind, and thus not a proper scientific term that can appear in a scientific law. However, 'the disjunctive component of [a bridge law] is overwhelmingly likely not to be a kind predicate from any physical science.'[115] For instance, the vast disjunction that we would need to compose to translate a particular mental state does not appear in the laws or theories of any science. Consequently, that vast disjunction is not a kind predicate, and can not substitute for a kind predicate in a scientific law.[116]

Fodor's natural kinds argument, distilled to its essence, really amounts to nothing more than the claim that some groupings or classes of objects and properties are scientifi-

[112] Godbey (1978) 433.
[113] Fodor (1995) 432.
[114] Kim (1992) 9.
[115] Bickle (1995) 29.
[116] Fodor presumably does not wish to restrict kind predicates to those that actually appear in current scientific laws, for that would deny the possibility of our discovering new scientific laws that contain new predicates. Consequently, it is unclear on what grounds he can discount the possibility that scientists may in fact succeed in finding a physiological basis for mental states, such as pain response.

cally legitimate, and some are not. The legitimate groupings, which Fodor calls natural kinds, are the ones about which some science has actually seen fit to formulate laws. The disrespectable groupings, such as the cobbled together disjunctive sets, do not appear as predicates in the laws of any science. Therefore, in order to reduce a special science law, we must identify the 'natural kind predicates in [the reducing science] . . . which correspond to each natural kind predicate in [the reduced law].'[117]

In summary, Fodor argues that because upper-level terms are multiply realizable, they can not be translated into the vocabulary of a lower-level science without resort to disjunctive sets, and because disjunctive sets are not natural kinds, they can not be used to translate upper-level terms either. Consequently, the upper-level terms being untranslatable, the laws and theories in which those terms appear are irreducible to the laws and theories of a lower-level science.

The appeal of this argument frankly escapes me. Although the argument has 'succeeded in persuading a large majority of philosophers of mind to reject reductionism and type physicalism,'[118] I am mystified by its success. And even though it 'has been used by many nonreductivists as the main weapon against the possibility of psychophysical reduction,'[119] I am convinced that it is fundamentally flawed, as I shall try to show.

A number of objections to Fodor's argument have been published. I shall mention three of these, before turning to the criticism that seems to me most decisive against Fodor's position. First, Bickle argues that multiple realizability can not be an obstacle to reduction since the 'textbook example' of reduction, 'the intertheoretic reduction of classical thermodynamics to microphysics,' involves the translation of temperature, which is multiply realizable.[120] '[W]hen you think in terms of the velocity and momentum of each individual molecule,' he writes, 'you will see that there are an *indefinite* number of ways for a given ag-

[117] Fodor (1995) 431.
[118] Kim (1992) 3.
[119] Kim (1996) 234.
[120] Bickle (1995) 29.

gregate of molecules to realize any given temperature, e.g., 300°C.'[121] Or as Godbey puts it, '[o]bjects made out of different chemicals, organic and inorganic materials, things on earth and things on Mars—all of these things can be hot.'[122] 'Multiple realizability with a vengeance!'[123] But because no one doubts that temperature and thermodynamics have been successfully reduced, 'we should be[ ] suspicious of [Fodor's] argument.'[124]

Second, some philosophers argue that upper-level terms are not natural kinds, so there is nothing more to reduce than the individuals that are grouped under the upper-level term; and since Fodor acknowledges that individuals are nothing more than their parts—his token physicalism—there is no obstacle to reducing upper-level terms. Rosenberg argues, for example, 'that species are not natural kinds';[125] if he is correct, then there is no need to translate the term species into a natural kind at the lower-level, in order to reduce the biological laws and theories that refer to species. Similarly, Kim calls 'into doubt the unity and systematicity of [mental properties] as scientific kinds.'[126] Since a particular mental state, such as a pain, is nothing more than a neurological event, then the upper-level mental term pain is as 'heterogeneous'[127] as the neurological states that realize pains. Kim's point seems to be this: if each of the instances of an upper-level term is reducible to a lower-level event, then the upper-level term is really nothing more than a name for a collection of lower-level events. Thus, the upper-level term is nothing more than a disjunctive set. Given Fodor's token physicalism, this conclusion seems inescapable, and fatal to his position.

Third, Kim points out that Fodor has failed to provide any argument for the most crucial premise in his argument. The Nagel model of scientific reduction, as we have seen, requires that upper-level terms be translated into coextensive lower-level predicates. But why must these lower-level predicates be natural kinds, as Fodor insists? '[W]hy should

[121] Bickle (1992) 53.
[122] Godbey (1978) 434.
[123] Bickle (1992) 53.
[124] Godbey (1978) 434.
[125] Rosenberg (1985) 165.
[126] Kim (1996) 235.
[127] Ibid.

"bridge laws,"' Kim asks, 'connect kinds to kinds?'[128] All we have is Fodor's confident claim that '[i]f reductionism is true, then *every* [upper-level] kind is, or is coextensive with, a physical kind.'[129] He provides no argument for this assertion. 'What we need,' Kim concludes, 'is an *independent* reason for the claim that . . . [a] heterogeneous disjunction is unsuited for laws.'[130]

I find these arguments compelling against Fodor's position. But, to my mind, the strongest criticism of that position is the charge that Fodor has, so to speak, stacked the deck against the scientific reductionist. As Faber puts it, Fodor's insistence that upper-level terms be translated into coextensive natural kinds at the lower-level is 'an impossible demand.'[131] Lower-level terms can not be coextensive with upper-level terms, because upper-level objects are composed of lower-level parts, and a part of a whole can not be coextensive with the whole.[132] Thus, given Fodor's requirement of coextensive natural kinds, 'no lower-level [term] could possibly be coextensive with an upper-level [term], and Fodor's thesis would be trivially true.'[133] All that Fodor has shown, it seems to me, is that if we accept token physicalism, according to which each individual thing is a physical thing, made up of physical parts, then we will not be able to find a lower-level object that is coextensive with the upper-level thing. But, of course, no reductionist ever maintained that we could. No one should find it remarkable that the terms in molecular genetics are not coextensive with the terms in nonmolecular Mendelian genetics, since the objects described by Mendelian genetics are made up of the objects described by molecular genetics.

The most important part of Fodor's argument is the claim that upper-level terms are natural kinds. It is notoriously difficult to define satisfactorily a natural kind, but the general

---

[128] Kim (1992) 10.
[129] Fodor (1995) 432.
[130] Kim (1992) 10.
[131] Faber (1986) 125.
[132] An apparent exception occurs in the strange case of actual infinites, since, according to infinite set theory, the set of all natural numbers (i.e., the positive integers) has the same number of members as the set of all positive even numbers, which is a subset of the set of all natural numbers. J. P. Moreland, 1987, *Scaling the Secular City* (Grand Rapids, Michigan: Baker Books), 21.
[133] Faber (1986) 125.

idea seems to be that some divisions of the world are more natural, that is to say, more correct, than others. For example, biologist John Maynard Smith, defending the position that species are natural kinds, writes that 'the classification of animals into separate species often corresponds to real differences, and is not an artificial scheme imposed for convenience'; in other words (also his) 'the classification of animals into higher categories is not . . . wholly an arbitrary matter,'[134] as it would be if we classified animals according to their color. Yellow organisms, he suggests, do not form a natural kind.[135]

I read Fodor as attaching an ontological meaning to the concept of natural kinds, for he writes 'that there are special sciences not because of the nature of our epistemic relation to the world, but because of the way the world is put together.'[136] We divide the world into groups of similar objects, which we call kinds or types. Some of these kinds are 'important' enough that the sciences make 'generalizations' about them.[137] These important kinds, which Fodor calls 'kind predicates[,] . . . are those that figure in the laws of [a] science.'[138] Indeed, Fodor defines a natural kind as a predicate that appears in a 'proper'[139] scientific law.

Now the question we must ask in scientific reductionism is whether the terms used in science refer to ontologically significant classes of objects, or whether these terms are merely names for collections of lower-level entities that we find it interesting and convenient to lump together. Fodor accepts the former view. For him, Faber writes, 'gene pairs, states of cognition, and monetary systems are just as irreducibly real as electrons, protons, and neutrons.'[140] But why does Fodor think that upper-level terms are ontologically real things?

Let us consider first the option, which Fodor dismisses, that scientific terms, such as genes or monetary exchanges, are not natural kinds, but only names for classes of objects,

---

[134] Maynard Smith (1975) 39.
[135] Ibid., 40.
[136] Fodor (1995) 439.
[137] Ibid.
[138] Kim (1992) 10.
[139] Fodor (1995) 432.
[140] Faber (1986) 125.

similar in some respect, that we find it useful and convenient to group together. Such pragmatic groupings, depending upon our interests, have no ontological significance, so there is presumably no barrier to reduction. We can translate the class as a disjunctive set of lower-level objects. For instance, the upper-level term 'yellow thing' can be translated as an open-ended disjunctive set of objects that share the upper-level property of appearing yellow (e.g., *either the sun, or a lemon, or a lion, or a corn kernel, or* . . .); each instance of the upper-level term—each yellow thing—corresponds to one of the disjuncts in the disjunctive set. Consequently, it would be baseless to argue that the term 'yellow thing' is irreducible because there are no lower-level terms that are coextensive with it. Rather, the term is reducible precisely because it is an artificial construct; it has no reality beyond that possessed by its members.

Fodor believes that the natural kind predicates used in science are ontologically significant, and thus irreducible. They are not merely artificial groupings that can be translated as disjunctive sets of lower-level objects. But what is it that renders these terms real or ontologically significant, such that they are more than an artificial grouping of objects? As we saw in the previous chapter, a real thing is able to affect, and be affected by, other things in the world. 'Anything real must be part of the causal structure of the world.'[141] Upper-level objects merely express the causal powers of their constituent parts, unless they have their own ontologically emergent causal powers, which go beyond those possessed by their components, and which can not be explained in terms of those components.

But Fodor does not argue that the terms used in the special sciences refer to objects that have ontologically emergent causal powers; he does not claim that upper-level objects can affect other things in the world in ways that can not be explained in terms of the causal powers of the constituent parts. In fact, consistent with his token physicalism, he 'is willing to believe' that each instance of the upper-level term 'has a true description in the vocabu-

[141] Kim (1996) 230.

lary of physics and in virtue of which it falls under the laws of physics.'[142] Consequently, the upper-level term is not an ontologically real thing, with its own unique suite of causal powers. It is nothing more than a set of lower-level objects that share some similarity that we consider 'interesting'[143] or 'important.'[144] But this interesting, important upper-level similarity does not confer ontological respectability upon the collection of objects sharing that similarity. And a term does not become real simply because we call it a natural kind.

To summarize, each scientific term, unless some novel property emerges from the conceptual joining, is simply a collection of objects or events to which we have given a name; there is no reason to think that it is more than an artificial collection. The term is nothing more than a disjunctive set; so the translation of the term by means of a disjunctive set is not an *ad hoc* maneuver to effect reduction, but merely what is required by the artificial nature of the upper-level term. To avoid the reduction of upper-level terms, Fodor must show that something new has ontologically emerged that can not be explained in terms of the interacting constituent parts. He must show, for example, that genes or monetary exchanges have the causal ability to do things that can not be explained in terms of the interactions of their components, and this he has neither done nor attempted to do. The joining together of a collection of objects under a name does not call forth a new, emergent, irreducible object.

Fodor's position is based on the intuition that there really is something significant about the terms used in the special sciences. Understandably, Fodor takes the special sciences 'seriously,'[145] because the special sciences take seriously the existence of upper-level objects. Like many of us, Fodor seems to regard the existence of the everyday objects of experience as a starting point of philosophical reflection, and not as a hypothesis that must be proven. If we assume that these upper-level objects are ontologically real things in their own

[142] Fodor (1995) 433.
[143] Ibid.
[144] Ibid., 439.
[145] Ibid., 433.

right, which must be included in an inventory of the world's furniture, then it follows, of course, that they are not reducible to their components, since reducible objects are eliminated as real things in their own right. But this assumption begs the question against the reductionist, who denies that upper-level objects have any existence separate from their constituent parts. Consequently, Fodor has not so much mounted an argument against scientific reductionism, as announced a starting point that is inconsistent with that position.

In conclusion, I am skeptical of Fodor's anti-reductionism for a variety of reasons. Besides the three objections that I cited from the published literature, I discussed two additional criticisms. First, Fodor's requirement that upper-level scientific predicates be translated into lower-level scientific predicates is unreasonable, and tells us nothing about the relation between the laws and theories of the sciences. Since upper-level objects are composed of lower-level parts, it is impossible for the lower-level term to be coextensive with the upper-level term, as Fodor requires. And second, upper-level objects are nothing more than their components unless the upper-level objects have inexplicable, ontologically emergent causal powers. Consequently, whether or not an upper-level object can be reduced to its constituent parts is an ontological question, which we answer by trying to explain the causal powers of the object in terms of the properties of its interacting components. Ontological reductions have nothing to do with finding lower-level kind predicates that are coextensive with upper-level kind predicates. And because Fodor has given us no reason to think that the upper-level terms used in the special sciences have emergent properties, there is no reason to think that upper-level terms are irreducible to collections of lower-level objects.

B. The Irreducibility of Form

The second argument against scientific reductionism is a version of an argument we encountered in the previous chapter. As we saw, the anti-reductionist argues that the properties of upper-level objects can not be explained in terms of the properties of the components because it is the form or structure of the object that determines its properties, and that form is different from the parts, and is not reducible to them. For example, physicist Paul

Davies argues that the properties of a living cell are irreducible to the properties of the constituent atoms, because it is the 'complexity and organization'[146] of a cell that explains life. Since '[a]n atom of carbon, hydrogen, oxygen, or phosphorus inside a living cell is no different from a similar atom outside . . . life cannot be reduced to a property of an organism's constituent parts.'[147] It is the form of the cell that accounts for its properties, and this arrangement of the atoms is separate from the atoms themselves.

The argument against scientific reductionism claims that the form or structure of an object, which accounts for its behavioral properties, is not reducible to physics and chemistry. This is different from the ontological version, according to which the form or structure of an object is not reducible to the constituent particles. These arguments are rarely distinguished properly. In what follows, we shall be concerned exclusively with the argument against scientific reductionism.

The general thrust of the argument can be briefly stated. The laws of physics describe and explain the behavior of particles, but the behavior of the upper-level objects studied by the special sciences are determined by the *organization* of the particles that make up these objects; how the particles are put together, consistent with the laws of physics, is what accounts for the properties of upper-level objects. The laws of physics, however, do not explain the *form* of the object, so the behavior of upper-level objects can not be explained by the laws of physics.

The pivotal point of this argument, of course, is the claim that the laws of physics are inadequate to explain the organization or form of the particles that constitute an object. It will be easier to explain this claim, and to render it more cogent, if we examine a few specific examples of things that the anti-reductionists aver can not be explained by physics.

[146] Davies (1983) 59.
[147] Ibid.

First, Hilary Putnam argues 'that basic physics cannot be employed even to explain such physical truisms as that a square peg will not go into a round hole.'[148] A successful explanation of this fact must appeal to the concept of shape, both of the peg and of the hole. However, shape is not a concept that appears in physics, and the shape of the peg is not determined by the laws of physics, since those laws are consistent with other shapes. Thus, the laws of physics do not explain the inability of a round peg to fit in a square hole.

Putnam's argument, as I have presented it, depends crucially upon two claims. First, shape is not part of the vocabulary of physics, and second, the shape of an object is not determined by the laws of physics. We may be able to explain the behavior of material particles in terms of physics, but such an account would say nothing about the form of a collection of particles, about how those particles are arranged. And since it is the form or shape of a peg that determines whether it will fit into a particular hole, the laws of physics, which describe the fundamental properties of matter, will furnish no explanation of why the square peg will not fit in the round hole. Accordingly, the reductionist view of science, which regards all explanation as physical explanation, is an inadequate view of science.

Second, '[c]onsider the famous Möbius band.'[149] This is a strip of material that has been twisted once and the ends joined together to form a loop. It has strange, unexpected properties. By tracing a continuous line around the loop, one discovers that it has only one side and one edge! But these properties can not be explained by the laws of physics, for two reasons. First of all, the laws of physics underdetermine the shape of the Möbius band, since the collection of particles that make up the band may be given innumerable shapes, all of which are consistent with the laws of physics. And second, a complete physical description of these particles will include no mention of one-sidedness or one-edgedness. It is the form or shape, which belongs to the collection as a whole, that gives rise to these properties.

---

[148] Foss (1995) 418.
[149] Davies (1983) 95.

Therefore, the laws of physics are insufficient to explain the upper-level properties of a Möbius band, and scientific reductionism is false.

Third, '[i]t is instructive to consider machines.'[150] A machine functions because of its structure, the way its parts are put together. But this upper-level structure can not be derived from the lower-level laws of physics and chemistry, because 'the laws governing the particulars in themselves would never account for the organizing principles of a higher entity which they form.'[151] Moreover, the laws of physics are inadequate to explain the operation of the machine,because a complete physical description of each particle would not indicate whether the complex of particles fulfill a function, which is the essence of a machine. 'Hence a complete physical and chemical topography of an object would not tell us whether it is a machine, and if so, how it works, and for what purpose.'[152] Consequently, the laws of physics do not explain the functioning of a machine.

All machines, as physical objects, 'obey the normal laws of physics.'[153] Sometimes this is expressed by saying that machines are governed by the laws of physics, but as Charles Taylor reminds us, 'to say that a range of things is governed by a certain set of laws, . . . is not to say that these laws can account for all the behaviour of things of this range.'[154] 'Being convinced that life does not defy the basic laws of physics is not, of course, the same as saying that the laws of physics explain life.'[155] For example, the game of chess is governed by a definite set of rules, each legal move obeying those rules. But the outcome of the game is not *determined* by the rules of chess—otherwise, every game would terminate in the same result, since every game is played under the same rules. It is the higher-order principles of

[150] Brooks (1994) 807.
[151] Polanyi (1983) 34.
[152] Ibid., 39.
[153] Pattee (1970) 124.
[154] Taylor (1971) 61.
[155] Davies (1983) 65.

strategy that determine the outcome, which operate within the boundaries or constraints provided by the rules of the game.[156]

A game of chess, like the functioning of a machine, is 'under dual control.'[157] In the case of machinery, the laws of physics provide the lower-level constraints, but the principles of engineering enable the machine to harness those laws to fulfill its function. Though 'based on the laws of physics[, a machine is not] explicable by the laws of physics.'[158] An explanation of a machine requires both physics and engineering. For example, the operation of a motor is governed by the second law of thermodynamics, which prevents the motor from being perfectly efficient, but the law does not determine how the motor will operate within the physical boundaries left open to it. The engineer who designs a machine may not violate the laws of physics, of course, but within the constraints provided by those laws he is free to pursue a variety of design strategies to achieve the goal of a working machine. The laws of physics alone do not even explain why a well-designed machine works, and a poorly-designed 'broken' machine does not, since both are governed by the same laws of physics. This sort of explanation will involve the upper-level 'operational principles'[159] of mechanical engineering.

Similar examples are readily found in the literature on reduction. For instance, 'a correct use of grammar does not account for good style,'[160] the sequence of nucleotides in a DNA molecule is extraneous to the laws of chemistry,[161] and 'the chemical pathways by which genes become phenotypically manifest . . . are not implicit in chemical theory.'[162] H. H. Pattee summarizes the position of the anti-reductionist: 'Each level has its own laws or rules which control the behavior within each level.'[163] And the laws or rules specific to one

---

[156] Polanyi (1983) 34.
[157] Ibid., 36.
[158] Ibid., 38.
[159] Ibid., 39.
[160] Ibid., 36.
[161] Grene (1971) 18.
[162] Kleiner (1975) 534.
[163] Pattee (1970) 119.

level can not be derived from those specific to a lower level. Thus, these upper-level laws, formulated by the special sciences to describe the behavior of upper-level objects, are irreducible to the laws of chemistry and physics.

In conclusion, the anti-reductionist's argument against scientific reductionism depends on two claims: that the properties and behavior of upper-level objects derive from the form or structure of those objects, and that this form or structure can not be derived from the laws of physics, since those laws are consistent with a manifold of other forms and structures. The lower-level sciences are not sufficient to account for the form or structure of upper-level objects, since the laws of the lower-level sciences underdetermine how the parts of an object will be put together.

This argument possesses a great deal of intuitive appeal. The scientific reductionist can admit many of the claims upon which the argument depends. For instance, the reductionist allows that the properties of upper-level objects derive from the form or structure of the object, and that the properties of the constituent particles underdetermine the form of the object composed of them. But the reductionist denies that these claims provide any reason to think that upper-level objects can not be explained by the laws of physics.

For the reductionist, the properties of upper-level objects are contained in the interactions of the particles. A macro-property is not shown to be irreducible simply because it is not possessed by the individuals. The question that must be answered is whether there is anything more to the object than its interacting constituent parts. The anti-reductionist answers that the object is more than just its parts; it is the collection of particles plus their arrangement or structure. The form is separate from the particles, and can not be explained in terms of the particles, since those particles could be arranged differently. For instance, '[t]he picture on a jigsaw,' Davies writes, 'can only be perceived at a higher level of structure than the individual pieces—the whole is greater than the sum of its parts.'[164]

[164] Davies (1983) 61.

If an upper-level object *really* is more than the sum of its parts, then it truly is irreducible to the parts. But it must be *ontologically* more than the sum of its parts. Thus, in order to avoid reduction, the anti-reductionist must show that something ontologically new has come into being from the interaction of the particles, which has causal powers not explainable as the outcome of the interaction of the parts. With respect to Davies jigsaw puzzle, we must ask whether the picture that can be perceived only at a higher level of structure is an ontologically emergent thing, whose macro-properties can not be explained in terms of the parts and their micro-properties. If the upper-level object does not have ontologically emergent causal powers, then it *is* nothing more than its parts, however interesting we may regard it.

Davies argues, for example, that an ant colony has 'emergent holistic features': 'To say that an ant colony is nothing but a collection of ants is to overlook the reality of colonial behaviour.'[165] But the reality—that is, the ontological reality—of the colonial behavior is precisely what the reductionist denies, because once we have provided a description of the behavior of each ant in the colony, and how each ant is affected by its interactions with other ants, we have accounted for all of the causal powers of the colony. The ant colony does not have any additional causal powers that are inexplicable in terms of the actions of the ants. And without these additional causal powers, which confer ontological respectability upon an object, the ant colony *is* nothing but a collection of ants. The upper-level colonial description of the ants is therefore redundant, since the behavior of the colony can, in principle, be derived from the behavior of the ants that compose it.

Having discussed in the previous chapter how the reductionist explains structure and form, I shall not repeat that discussion here, except to say that the form or structure of an object is not an ontologically emergent thing or property. The patterns that we find in nature, Faber notes, 'arise[ ] . . . as the outcome of the actions of the particles carrying on their

---

[165] Ibid., 63.

small affairs according to their individual natures.'[166] Once we have specified the relative locations of the various particles that make up an object, we have given its shape, and showing how the particles interact with one another accounts for the structure of the object. But let us consider specifically how the reductionist explains the square peg, the Möbius band, and the operation of machines.

How does the reductionist explain why a square peg will not fit into a round hole? The peg is nothing more than a collection of particles, since it has no inexplicable causal powers that can not be traced to the causal powers of its particles. The causal powers of the peg are determined by the properties of its constituent particles. We can explain the behavior of each individual particle, and its interactions with its neighbors, in terms of the laws of physics, and since the peg is nothing more than a collection of interacting particles, we can explain all that there is to explain about the peg. The anti-reductionist's complaint that an explanation in terms of individual particles makes no mention of peg, shape, or hole, is dealt with by denying that these refer to ontologically real things. A true ontological description would include no mention of pegs, shapes, or holes—only particles and their interactions with one another. What we are calling the peg and the hole are just clusters or groupings of particles, while shape is nothing but a summation of the relative locations of the individual constituent particles. Unless these objects have novel causal powers that can not be explained in terms of the properties of their components, then they are nothing but those components. The anti-reductionist has made no showing of any novel causal powers attaching to the peg or hole.

How does the reductionist explain the strange properties of the Möbius band? Like the square peg, the Möbius band is nothing over and above the particles of which it is composed, nothing ontologically new having come into existence from our giving the name Möbius band to a particular collection of particles. The macro-properties that accrue to a collection of particles by virtue of its shape are readily accounted for, without supposing that

[166] Faber (1986) 136.

these properties have mysteriously emerged. Every shape has properties that do not belong to the individual parts. For instance, the collection of particles that constitute the object we call a snowflake has the property of being hexagonal or six-sided; 'even the stars are grouped in triangles, Latin crosses, half-circles, and so on.'[167] But this shape is not something in addition to the collection of particles. The shape of the Möbius band or of a constellation of stars is nothing more than the relative location of each particle or star considered collectively. The macro-property of shape is 'contained' in the collection of particles; it is not ontologically emergent. Therefore, the Möbius band is not irreducible, as the anti-reductionist suggests.

And finally, what account can the reductionist give of a machine? What has been said with respect to the square peg and the Möbius band can also be said about a machine: unless the machine has ontologically emergent properties, which emerge mysteriously from the interacting components as a genie arises from a lamp, then the machine is nothing more than a collection of particles. The question at issue is not whether the machine can do something that the individual particles can not, which of course it can, but whether the machine has properties and causal powers that can not be explained in terms of the whole collection of interacting particles. The anti-reductionist has offered no evidence that it can.

Instead of arguing that the machine has ontologically emergent properties, the anti-reductionist makes two other claims, which purport to show the irreducibility of machines. First, a description of a machine in terms of interacting material particles makes no mention of the distinguishing characteristics of a machine; namely, its structure and its function, or the purpose for which it was built. What distinguishes a machine from a non-mechanical object, such as a stone, is that the parts of a machine are arranged in such a wise as to perform, or to be capable of performing, useful work, such as raising a load, or converting heat into mechanical energy. But reducing a machine to its components ignores those character-

[167] Ibid.

istic features, so the reduction is incomplete; the reduction has not fully captured what it is to be a machine.

Second, the lower-level laws of chemistry and physics are insufficient to explain the structure of an upper-level object, such as a machine, since those laws undetermine the structure. The same laws apply to and describe a working machine, a broken machine, and a non-machine, so it can not be those laws that explain the structure and operation of a machine. In order to explain why one machine operates more efficiently than another, we must look to upper-level laws, such as those of engineering, just as we must appeal to laws of biological form and organization to explain the distinction between living organisms and non-living objects, since both follow the laws of chemistry and physics.

But despite the plausibility of these two claims, the anti-reductionist has not refuted reductionism, because, as we have said, unless the object has ontologically emergent properties that can not be explained in terms of the collection of particles that constitute the object, then the object really is just a collection of particles. Nonetheless, let us consider each of these two objections to scientific reduction. First, the anti-reductionist complains that reducing a machine to a collection of particles will ignore the structure and function of the machine, its distinguishing characteristics. As we have said, the structure of an object is nothing more than the relative locations of the interacting individual particles taken collectively. Once we have specified the position and interactions of each particle, there is nothing called 'structure' left to explain. And what we call the function of a machine is simply the causal power for which it is specially designed or actually used; for instance, the function of a thermostat is to regulate temperature by activating cooling or heating devices. But we can explain this causal power in terms of the properties of the constituent particles, so there is nothing mysterious, inexplicable, or irreducible about the function of a machine.

Second, the anti-reductionist argues that lower-level laws underdetermine the operation of upper-level objects. For example, the sequence of amino acids in a particular protein is not determined by the laws of physics or chemistry, since other sequences are also consistent with those same laws. The anti-reductionist is correct, of course, that the sequence of amino acids in a protein is not determined solely by the properties of those amino acids. In order to explain the sequence, we must take into account the interactions of those amino acids with other particles that are not part of the protein, and we must consider the causal history of the protein, since its present sequence is largely determined by events that occurred before it came into existence. We can not explain the current interaction of a group of particles without taking into account both past and present particle interactions.

But the anti-reductionist is incorrect in thinking that we need higher-level laws to explain, for example, the sequence of nucleotides in a molecule of DNA. The interactions of material particles are described by the laws of physics and chemistry, and it is these physical and chemical interactions, many of them occurring in the past, that determine the sequence of nucleotides in a molecule of DNA, or the structure of any object whatsoever; there are no extra higher-order principles involved. Even though the physical and chemical properties of material particles in an object do not by themselves determine how they will be organized, it does not follow that the laws of physics and chemistry are insufficient to explain that organization. This is the anti-reductionist's mistake. For example, Polanyi writes that '[i]f we [acknowledge] the principle that the operations of a higher level can never be derived from the laws governing its isolated particulars, it follows that none of [the upper level] operations can be accounted for by the laws of physics and chemistry.'[168] But in order to explain the form or structure of an object, we must simply take into account how the constituent particles are affected by the physical and chemical properties of other particles outside the object. Thus, if we regard upper-level objects as collections of interacting particles that are open to outside influences, then the laws of physics and chemistry can explain the structure, behavior, and operation of upper-level objects.

[168] Polayni (1983) 37.

Conclusion

In this final section, I want to summarize a few of the conclusions of this chapter. Scientific reductionism is the claim that all the laws, theories, and generalizations of the special sciences can be reduced to the laws of physics. The anti-reductionist rejects this claim. While most anti-reductionists accept that particular things are nothing more than the material parts of which they are composed, they nonetheless deny that the laws and theories of the special sciences can be reduced to the laws and theories of physics. This rejection of scientific reductionism ultimately rests on the intuition that there is something real, important, legitimate, and non-arbitrary about the terms and generalizations posited by biology, psychology, economics, and other special sciences. This intuition stands behind Fodor's argument that the predicates of the special science are natural kinds that can not be translated by artificial disjunctive sets; for Fodor, it is inconceivable that there is not some truth in the division and conception of the world represented by the special sciences.

The same intuition informs the argument by Polanyi and others that upper-level structures can not be explained by the laws of physics. For Polanyi, the special sciences paint 'a picture of the universe filled with strata of realities.'[169] Even though a brick house, a brick sidewalk, and a mere pile of bricks are all composed of bricks, in reducing each of them to a collection of bricks, we leave out what it is that fundamentally distinguishes them: their form or structure. This form must be regarded as real, because, intuitively, we see that these upper-level objects are 'real entities.'[170]

Thus, the resistance to scientific reductionism, I contend, is fundamentally a dissatisfaction with ontological reductionism, with the claim that the familiar objects of experience really are nothing more than their components, and ultimately nothing more than collections of material particles. Ontological reductionism offends our basic understanding of reality and our belief in the existence of such objects. However, the anti-reductionist does not

[169] Ibid., 35.
[170] Ibid., 33.

wish to challenge ontological reductionism directly, because the alternative seems to commit him to mysterious vital forces, entelechies, or breaches of physical causation, which he regards as anathema to his materialist worldview. So he argues instead for the autonomy of the special sciences, which take seriously the existence of upper-level objects. In this way, he hopes to avoid ontological reductionism indirectly.

Alas, his escape from reductionism is illusory. For if the objects of experience are really nothing more than collections of particles, then the laws of physics, which explain and describe the properties and behavior of each particle, provide a complete account of the objects. Unless the interaction of the constituent particles creates an ontologically new thing, with novel causal powers that can not be explained in terms of the constituent particles, there is nothing left for the special sciences to explain. All the explanatory work is done at the level of the particles by the laws of physics. Faber describes the anti-reductionist position as 'conced[ing] . . . more than seems possible without also surrendering unconditionally,'[171] and he concludes correctly that the concession is 'generally fatal'[172] to that position. The battle must be fought at the ontological level.

For Rosenberg, we are all good ontological reductionists, so the 'denial of reductionism . . . is an epistemological point, not a metaphysical one.'[173] This statement is mistaken, I think, on two counts. First, he overstates the extent of commitment to ontological reductionism, and second, he mischaracterizes the position of many of those who oppose scientific reductionism. For many anti-reductionists, opposition to scientific reductionism is not primarily epistemological. For instance, I take Fodor seriously when he writes 'that there are special sciences not because of the nature of our epistemic relation to the world, but because of the way the world is put together.'[174] And Polanyi certainly does not believe that greater knowledge of physics would enable us to reduce the operation of machines to the

[171] Faber (1986) 122.
[172] Ibid., 133.
[173] Rosenberg (1986) 90.
[174] Fodor (1995) 439.

laws of physics. Such philosophers object to scientific reductionism for ontological reasons: they want to affirm the commonsense belief in the existence of the objects of experience. If the anti-reductionist can succeed in showing the irreducibility of the special sciences, he believes that he can obtain some respectability for the objects posited by those sciences, without denying that such objects are really nothing more than collections of particles. But this admission crucially undermines his position, because if all upper-level objects are just collections of particles, then not only will 'the lower-level theory . . . be nearer the truth in every individual case,'[175] but the objects posited by the special sciences will have no independent ontological reality. Why anyone would consider such a position worth fighting for, I can not conceive.

In the final chapter, I will consider issues relating to reduction in biological science.

[175] Faber (1986) 134.

CHAPTER 4

ORGANISMIC REDUCTIONISM

In the two previous chapters, I sought to unravel some of the tangled and complicated issues relating to metaphysical and scientific reduction. I tried to show, among other things, that ontological emergence, despite the numerous criticisms to which it has been subject, remains a viable metaphysical alternative to reductive materialism, and that commitment to reductive materialism, which starkly conceives of reality in terms of particles and forces, entails a belief in scientific reductionism, according to which all of science really is physics. Ontological emergence offers a potential avenue of escape from reductive materialism, but the proponent of reductive materialism will find no escape from scientific reductionism.

With each chapter in this work, my focus on reductionism has become narrower and more concrete. I began at the highest level of generality by discussing the metaphysical reduction of reality to materialism, then stepped down, so to speak, to scientific reduction, the reduction of the upper-level sciences to chemistry and physics, and in this final chapter, I continue this descent by addressing a particular type of reduction in biology, which I shall unartfully denominate organismic reduction. This is the view, popularly associated with the work of zoologist Richard Dawkins, that biological organisms can be reduced, in some sense, to their genes, or as Dawkins himself puts it, organisms are merely 'survival machines—robot vehicles blindly programmed to preserve the selfish molecules known as genes.'[1] I think it is preferable to call this view organismic reduction, rather than genetic reduction, in order to forestall any possible confusion with another reductionist debate involv-

[1] Dawkins (1989) v.

ing genes, which concerns 'the reducibility or not of nonmolecular, Mendelian genetics to molecular biology.'[2] I shall have very little to say here about the reduction of genes to segments of DNA; rather, the focus of this final chapter will be an examination and critique of the 'gene's-eye view of nature.'[3]

Accordingly, this chapter will be divided into two parts. In the first, I shall explain the selfish gene view of organisms and evolution, and then present, uninterrupted by the challenges of opponents, Dawkins' two main arguments in support of that view. In the second section, I shall subject the selfish gene perspective to a number of criticisms that I consider utterly decisive against any attempt to treat the struggle between selfish genes as 'an up-to-date version of the orthodox Darwinian theory of evolution.'[4]

The Selfish Gene View of Nature

A. Introduction

The main object of Richard Dawkins' two most famous books, *The Selfish Gene* and *The Extended Phenotype*, the latter directed primarily at Dawkins' 'professional colleagues,'[5] is to defend the thesis that '[t]he selfish gene theory is Darwin's theory, expressed in a way that Darwin did not choose.'[6] Dawkins proposes to update Darwin's theory with all that we have learned about genetics in the years since Darwin first proposed natural selection as 'the main but not exclusive'[7] engine of evolutionary change. But '[r]ather than focus on individual organisms,'[8] as Darwin had done, the selfish gene theory tries to account for evolution in terms of natural selection acting on individual genes, each of which is supposed to engage in the familiar Darwinian struggle to survive and reproduce itself. Whereas Darwin had regarded evolution as a process that occurs in populations of organisms, as individuals with

[2] Rosenberg (1986) 90.
[3] Dawkins (1989) x.
[4] Mackie (1978) 455.
[5] Dawkins (1982) v.
[6] Dawkins (1989) x.
[7] Darwin (1998) 7.
[8] Dawkins (1989) x.

favorable variations displace 'the less improved forms of life,'[9] the selfish gene view proposes that natural selection acts on genes, which differ in their ability to 'out-propagate each other,'[10] and that '[e]volution is a process by which some genes become more numerous and others less numerous in the gene pool.'[11] Thus, Dawkins attempts to show that evolution by natural selection can be translated into the language of genes, that evolution can be understood as the result of a competition among selfish genes to leave copies of themselves in the gene pool.

But as a number of commentators and even Dawkins himself have recognized, this project of showing that Darwinian evolution can be characterized as a competition among genes, rather than among organisms, leaves unanswered the question of the status of these two ways of representing evolution. Dawkins' main thesis can be understood in three incompatible ways. First, Dawkins might be claiming that the selfish gene view of evolution is actually the correct view of evolution; second, he might mean that the two views are different, but equally correct, ways of describing the facts of evolution; or third, he might be arguing that some biological phenomena are better described as a competition among genes, while other phenomena are perhaps better represented as a struggle between organisms. But anyone who would try to discover which of these three interpretations Dawkins has set himself the task of defending soon makes the unhappy discovery that support for each of them can be found in his writings.

In *The Extended Phenotype*, the book that Dawkins calls the 'pride and joy' of his 'professional life,'[12] he explains that what he is 'advocating is a point of view, a way of looking at familiar facts and ideas, and a way of asking new questions about them.'[13] The selfish gene view of nature 'is not a new theory, not a hypothesis which can be verified or falsified, not a model which can be judged by its predictions,' but a different 'way of seeing biological

[9] Darwin (1998) 6.
[10] Dawkins (1982) 133.
[11] Dawkins (1989) 45.
[12] Ibid., 234.
[13] Dawkins (1982) 1.

facts,' [14] which does not deny the truth contained in other ways of seeing those same facts. Dawkins explains his position by means of the 'well-known visual illusion called the Necker Cube,' 'a line drawing' that can be perceived in 'two possible orientations,' both of which are 'equally compatible with the two-dimensional image on the paper,' but 'neither of [which] . . . is the correct or "true" one. They are equally correct.'[15] Likewise, the selfish gene view 'is not provably more correct than the orthodox view,' though Dawkins 'sus-pect[s]' that 'it provides a deeper understanding'[16] of nature.

In this suspicion that the selfish gene view may in fact provide a deeper understanding of nature, Dawkins provides the first hint of his dissatisfaction with his own Necker Cube analogy. Within a few pages of his discussion of the Necker Cube, he concedes that the analogy 'may be too timid and unambitious.'[17] It turns out that 'the Necker Cube [actually] expresses [his] *minimum* hope for [the] book.'[18] This minimum goal is to show that the selfish gene view 'is at least as satisfactory' as the orthodox view; while '[i]n many cases the two ways of looking at life will, indeed, be equivalent,' in other cases biological phenomena may 'make no sense at all if we keep our mental gaze firmly on . . . the selfish organism.'[19] Some biological facts are 'inexplicable on certain hypotheses and easily explicable on others.'[20] This understanding of the selfish gene theory leaves open the possibility, though Dawkins himself does not acknowledge it, that there may be biological phenomena that can be explained only in terms of the struggle between selfish organisms, but that receive no account in terms of selfish genes.

Dawkins also presents what he calls the 'strong form' of his argument, according to which 'the selfish organism view is [not] necessarily wrong, but . . . it is looking at the mat-

[14] Ibid.
[15] Ibid.
[16] Ibid.
[17] Ibid., 7.
[18] Ibid., emphasis in the original.
[19] Ibid.
[20] Ibid., 52.

ter the wrong way up.'[21] Addressing the proponent of 'individual selectionis[m],' Dawkins writes that the two views of nature 'really do almost agree'; '[i]t is just that you *see* it wrong!'[22] To say the least, his meaning has not been expressed as conspicuously as could be wished. For if Dawkins truly regards the selfish organism view as the wrong way of looking at nature, then it is hard to credit him, except with a pinch of salt, when he writes that the two views are in many cases equally correct accounts of nature. If the orthodox view sees nature wrongly, then the Necker Cube analogy is not merely 'too cautious,' it is, as Dawkins himself comes to appreciate, downright 'misleading because it suggests that the two ways of seeing are equally good.'[23] But having conceded, as we saw above, that his selfish gene view 'is not a hypothesis which can not be verified or falsified,' but merely a new 'way of looking at familiar facts and ideas,' Dawkins apparently fails to realize that this concession critically undermines any suggestion that the selfish gene perspective is in fact the proper way of looking at nature.

Nonetheless, Dawkins does, at numerous points in his writings, emphatically assert that the selfish gene view is the correct view of evolution and organisms. Consider one final example. According to the orthodox Darwinian view, genes may be regarded as an organism's way of making another organism; genes, 'like any other aspect of an organism, [are] selected because [they do] the organism some good.'[24] But this way of looking at nature, Dawkins writes, actually 'turn[s] the truth upside down,'[25] for 'we [must] deeply imbibe the fundamental truth that an organism is a tool of DNA, rather than the other way around.'[26] Dawkins charges us to 'remember that the phenotypic effects of a gene are the tools by

[21] Ibid., 6.
[22] Ibid.
[23] Dawkins (1989) ix.
[24] Dawkins (1982) 158.
[25] Dawkins (1989) 238.
[26] Dawkins (1982) 158.

which it levers itself into the next generation,'[27] and reminds us that '[i]t requires a deliberate mental effort to turn biology the right way up.'[28]

Dawkins never adopts a single unequivocal argumentative stance with respect to the status of the selfish gene view, and while he is plainly aware of the need to clarify his position, he never seems to resolve the fundamental tension at its heart. He appears unable to decide whether selfish genes provide merely a fruitful perspective on nature, which can not be distinguished from the selfish organism view in point of its correctness, or whether the selfish gene view is in fact a more accurate way of looking at the biological world. He calls it his 'wildest daydream' to show that 'all interactions between and within organisms . . . can be illuminated in new ways by th[is] doctrine,'[29] but he frequently tempers his enthusiasm for the truth of the selfish gene perspective, suggesting only that 'the paradigm of the selfish individual, if not actually incorrect, can lead to difficulties.'[30]

'As is the way with advocates,' however, Dawkins tries 'to make the strongest case [he]can,'[31] which means arguing for the *truth* of the selfish gene view of nature. In consequence, he seeks to show that there are biological phenomena that can not be explained in terms of selfish organisms, but 'whose explanation is lucidly written'[32] as a story involving competing selfish genes, and that the orthodox Darwinian account of evolution, as a process caused by the natural selection of fitter organisms, can likewise be described as the result of natural selection acting on individual selfish genes. I do not think it is possible to reconcile everything Dawkins says about how we should understand his 'view of life,'[33] and so no such attempt will be made; but in what follows, I propose to present and evaluate his arguments for the claim that the selfish gene view explains the features of life at least as well as, and in some cases better than, the orthodox position involving selfish organisms.

[27] Dawkins (1989) 238.
[28] Ibid., 265.
[29] Dawkins (1982) 7.
[30] Ibid., 8.
[31] Ibid., 7.
[32] Ibid.
[33] Darwin (1998) 396.

In order to defend the thesis that the 'gene's-eye view of Darwinism' can 'make everything about nature fall into place,'[34] Dawkins must show that genes rather than organisms are the main actors on the evolutionary stage. This he attempts to do by means of two arguments, which unwind throughout his writings, but are not always kept tidily distinct. First, he argues that organisms are merely vehicles created by genes 'for their [own] continued existence'[35]; 'DNA is not working for the good of the cell [or organism] but for the good of itself.'[36] Second, he argues that the 'fundamental units of natural selection, the basic things that survive or fail to survive, that form lineages of identical copies with occasional random mutations,'[37] are genes; 'natural selection favours [genes] that are good at building survival machines, genes that are skilled in the art of controlling embryonic development.'[38]

Charles Darwin had regarded evolution as the product of the 'struggle for life' among organisms, in which struggle 'any [heritable] variation, . . . if it be in any degree profitable to an individual of any species, . . . will tend to the preservation of that individual, and will generally be inherited by its offspring.'[39] But for Dawkins, the struggle for life that results in evolution is waged by genes, with natural selection preserving the fitter genetic variants, the genes that are better able to construct bodies to transmit the genes into future generations.

Another way of distinguishing these two views is in terms of the way that each treats the phenotypic characteristics of an organism. Both sides regard natural selection, in Darwin's words, as nothing other than the 'preservation of favourable variations and the rejection of injurious variations,'[40] but they disagree about precisely what it is that is benefited or harmed by these variations. For the orthodox Darwinian, it is the organism that profits from adaptive traits, and is harmed by maladaptive characteristics; it is the fitness of the individual that is judged by natural selection. But the selfish gene theorist, on the other hand,

[34] Dawkins (1989) ix.
[35] Ibid., 19.
[36] Dawkins (1982) 162.
[37] Dawkins (1989) 254.
[38] Ibid., 24.
[39] Darwin (1998) 52.
[40] Ibid., 68.

treats the gene as the beneficiary of favorable variations, because it is genes that give rise to phenotypic traits, and genes that are actually transmitted from parent to offspring in reproduction; 'natural selection . . . favour[s] organs and behaviour that cause the individual's genes to be passed on.'[41]

Thus, in order to support this selfish gene view of nature, which regards the gene as the 'only . . . entity whose point of view matters in evolution,'[42] Dawkins, as I have said, argues that it is genes that produce organisms, and genes that are the units of selection. We turn to these arguments now.

B. Genes Create Organisms

There are, Dawkins notes with apparent bemusement, '[s]ome biologists [who] go so far as to see DNA as a device used by organisms to reproduce themselves, just as an eye is a device used by organisms to see!'[43] But these biologists commit 'an error of great profundity,'[44] because they fail to recognize that it is the genes, 'the individual hidden determinants of surface characters,'[45] that create organisms, and not organisms that create genes. This latter view Dawkins characterizes as 'the truth turned crashingly on its head.'[46]

Dawkins variously describes an organism as a 'survival machine' built by genes that 'manipulat[e] it by remote control,'[47] a 'proxy'[48] for the genes, the collective 'phenotypic effects' of a genome, [49] and 'a colony of genes.'[50] These are merely different colorful ways of expressing the view that genes create organisms, and by means of these organisms the genes indirectly show themselves to, and interact with, the external world. Natural selection may act directly on the organism, but inasmuch as the genes 'indirectly control the manufacture

[41] Dawkins (1982) 185.
[42] Dawkins (1989) 137.
[43] Ibid., 237.
[44] Ibid.
[45] Rose (1997) 102.
[46] Dawkins (1989) 237.
[47] Ibid., 20.
[48] Dawkins (1982) 4.
[49] Dawkins (1989) 235.
[50] Ibid., 46.

of bodies,'[51] selection acts indirectly on the genes, whose fitness is determined, at least in part, by the fitness of the bodies they create. Consequently, genes 'ensure their [own] survival by means of phenotypic effects on the world,'[52] these 'phenotypic effects of a gene [being] the tools by which it levers itself into the next generation.'[53]

To prove that genes create organisms, as a means of ensuring their own survival, Dawkins appeals to a number of related arguments. First, genes provide indirectly the instructions to make bodies. Although the only thing that 'genes can really influence directly is protein synthesis,'[54] which they do by acting as templates that determine the sequence of amino acids that constitute particular proteins,[55] this protein synthesis is, significantly, 'the first small step' to make a body, since '[p]roteins not only constitute much of the physical fabric of the body[,] they also exert sensitive control over all the chemical processes inside the cell, selectively turning them on and off at precise times and in precise places.'[56] Thus, by influencing the manufacture of protein, the genes initiate a cascade of causal events that terminate in the complete organism: '[t]he gene determines a protein sequence that influences *X* that influences *Y* that influences *Z* that eventually influences[, for example,] the wrinkliness of [a] seed or the cellular wiring up of the nervous system.'[57] Admittedly, most of the paths from genes to phenotype, 'are long and tortuous,'[58] but this does not diminish the rôle of genes as initiators of the causal process that leads to the creation, in each generation, of a new organism. Genes, Dawkins concludes, 'come first, in importance as well as in history.'[59]

---

[51] Ibid., 23.
[52] Dawkins (1982) 5.
[53] Dawkins (1989) 238.
[54] Ibid., 240.
[55] 'The coded message of the DNA, written in the four-letter nucleotide alphabet, is translated in a simple mechanical way into . . . the alphabet of amino acids which spells out protein molecules.' Ibid., 23.
[56] Ibid.
[57] Ibid., 240.
[58] Dawkins (1982) 197.
[59] Dawkins (1989) 265. Dawkins means that genes are first both in phylogenetic and ontogenetic history; the molecular forerunners of genes arose before organisms in the history of life, and the genome precedes the organism in the process of embryological development.

A second related argument directs our attention to, and draws much significance from, the fact that '[e]ach new generation starts from scratch.'[60] If we recognize '[a]n organism [a]s the physical unit associated with one single life cycle,'[61] then 'each act of reproduction involves a new developmental cycle' that begins from a single cell, 'a return "back to the drawing board" . . . in every generation.'[62] Within that single cell, 'endowed with one master copy of the architect's plans,'[63] are found all the instructions necessary for the construction of the adult organism. These instructions are contained within the genes, or sequences of DNA nucleotides, which code for proteins, and regulate the expression of other genes. Therefore, since an adult organism develops from a single cell, based on the genetic 'recipe'[64] contained within that cell, we can say that the genes create the organism, and control its ontogeny. And by creating fit organisms, which successfully transmit their genes to their offspring, genes ensure their own survival. Consequently, Dawkins concludes that organisms are vehicles used by genes to convey themselves into future generations.

But would it not be just as correct to adopt the alternative, orthodox view that genes are the means by which organisms reproduce themselves, and thus to treat genes as the tools of organisms? After all, the single cell that begins a life cycle, and develops into an adult organism, is derived from the parental organisms during the process of reproduction; the complement of genes that provide the developmental instruction for the newly formed organism likewise proceed from the parents. So it would seem that we can regard, with equal veracity, either the gene or the organism as the tool of the other.

Dawkins, however, will have none of this alternative perspective. As he frames the debate, organisms are merely vessels in which long-lasting genes reside for a time, and by which the genes are transported into the future. This conclusion follows for Dawkins from

[60] Ibid., 23.
[61] Dawkins (1982) 259. As Dawkins recognizes, equating an organism with a life cycle is far more plausible with respect to sexually reproducing organisms, for 'there may be no important distinction between growth and asexual reproduction.' Ibid.
[62] Ibid.
[63] Dawkins (1989) 23.
[64] Dawkins (1981) 567.

the claim that genes endure across generations, and organisms do not. In sexually reproducing organisms, 'one individual body is just a temporary vehicle for a short-lived combination of genes, . . . but the [individual] genes themselves are potentially very long-lived'; indeed, '[o]ne gene may be regarded as a unit that survives through a large number of successive individual bodies.'[65] It is this assertion, that genes remain relatively unchanged for generations, that constitutes the 'central argument'[66] of Dawkins' view of nature.

We shall examine the argument in detail in the next section of this chapter, but for now let us note the significance that Dawkins attaches to genetic 'longevity-in-the-form-of-copies,'[67] and to the claim that individual genes are transmitted largely intact from parent to offspring. Since individual genes may flow unchanged through generations of organisms, and individual organisms each last for only a single generation, it is a mistake to regard the enduring gene as the instrument of the transitory organism. Indeed, sexually reproducing organisms can not truly be said to reproduce themselves at all, inasmuch as the offspring is not in fact a copy of the parents, but a new individual made from a unique combination of their genes. But individual genes, though subject to splitting during the process of crossing over, can nevertheless 'expect to survive for a large number of generations in [an organism's] descendants'; '[t]he average life-expectancy of a genetic unit can conveniently be expressed in generations.'[68] Consequently, since genes flow unchanged through generations of organisms, it can not be the case that the genes are the products of the organisms, and the genes must be regarded as the creators of organisms, and not their instruments.

The third argument to prove that genes create organisms is based on the observation that a change at the genetic level may lead to a change at the phenotypic level of the organism. The fact that a genetic change may be attended by a phenotypic change strongly suggests that it is the genes that cause and control the phenotypic characteristics of the organ-

[65] Dawkins (1989) 25.
[66] Ibid.
[67] Ibid., 29.
[68] Ibid.

ism. '[G]eneticists are concerned with [the] "one gene-difference one animal-difference" [relation],'[69] while Dawkins himself is concerned with the significance of this relation for his selfish gene view of nature.

When geneticists speak of a gene 'for' some phenotypic trait, Dawkins explains, this is merely 'a convenient way of talking about "the genetic basis of [that trait]".'[70] This way of speaking, which 'has been standard practice in population genetics for over half a century,'[71] does not imply that the gene 'for' some trait is solely responsible for creating that trait. 'A gene "for brown eyes" is not a gene that, alone and unaided, manufactures brown pigment.'[72] Rather, '[i]t is a gene that, when compared with its alleles (alternatives at the same chromosomal locus), in a normal environment, is responsible for the difference in eye colour between individuals possessing the gene and individuals not possessing the gene.'[73] All other things being equal, if having a particular gene leads to a different phenotypic trait than not having the gene, then we can speak of the gene as being 'for' the trait in question. This is true even if the phenotypic effect of the gene depends upon the gene's interactions with, and the contributions of, many other genes, and even if we are speaking of genes for complex behaviors, such as genes for reading or tying shoelaces.[74] When a geneticist postulates a gene for reading, or for homosexuality,[75] all he means is that the difference between a person who has that trait and another person who does not have it can 'be traced to . . . a simple antecedent difference[ ]'[76] in their genes. It is this genetic difference that we refer to when we speak of a gene for some phenotypic trait.

---

69 Dawkins (1981) 567.
70 Dawkins (1982) 29.
71 Ibid.
72 Dawkins (1981) 566.
73 Ibid.
74 Dawkins (1982) 23.
75 Ibid., 38.
76 Dawkins (1989) 37.

As we have said, '[a]ll genes are fundamentally "genes for making proteins."'[77] But even though the developmental pathway from protein synthesis to the complete organism is exceedingly complex and little understood, 'a geneticist . . . need not care about the detailed pathway from gene to phenotypic effect'; his concern is strictly with 'differences between alleles in their effects on end products.'[78] Indeed, although 'it would [admittedly] be nice to know how phenotypes are made,' geneticists properly 'treat[ ] embryonic development as a black box.'[79] If a genetic difference can be linked to a phenotypic difference, 'even [though] we haven't the faintest idea of the chemical chain of embryonic causes leading from gene to [phenotype],'[80] then the geneticist can be sure that the gene, in concert with other genes in the genome, is responsible for the difference. Thus, genes can be regarded as 'master programmers . . . programming for their lives,'[81] since their success in creating fit organisms partly determines whether they survive as copies in later generations.

Before we turn in the next part of section to the argument that the gene is the unit of selection, it may be helpful to summarize our results to this point. Dawkins is trying to prove that the selfish gene view of nature is Darwin's view of nature, and this requires him to show that the Darwinian struggle for existence between organisms is better represented as a struggle for existence between genes. This he attempts to do by arguing that organisms are really nothing more than the by-products of genes, vehicles built by genes to ensure their own survival. We have considered three arguments for this claim. First, by acting as templates for the production of proteins, genes initiate the causal sequences that terminate in the construction of a new organism. Second, the embryological development of an organism begins with a single cell, which contains, in the form of genes strung along chromosomes, all the instructions necessary to manufacture the organism, and these genes flowing relatively unchanged through generations of organisms, can not be said to be products of the

[77] Dawkins (1981) 566.
[78] Dawkins (1982) 197.
[79] Ibid., 22.
[80] Dawkins (1989) 61.
[81] Ibid., 62.

organisms. And third, a simple change at the genetic level, such as the substitution of one nucleotide for another, may result in a change in the phenotype, thus strongly signaling that the phenotype is a product of its genes, and since genes are not affected by the body in which they are contained, Lamarckian mechanisms being ruled out, genes must be regarded as the causes of organisms, and not the reverse.

C. The Gene is the Unit of Selection

A 'central theoretical problem' of evolutionary theory is to determine 'the nature of the entity for whose benefit adaptations may be said to exist.'[82] It may be stated uncontroversially that natural selection preserves particular traits, such as improved camoflage or resistance to disease, *because* they are beneficial, but '[a]re they for the benefit of the individual organism, for the benefit of the group or species of which it is a member, or for the benefit of some smaller unit inside the individual organism?'[83]

Traditional accounts of Darwinian evolution, beginning with Darwin's own explication of his theory, have unquestioningly regarded the individual organism as the beneficiary of such adaptations. As Darwin writes, 'natural selection acts by life and death,—by the preservation of *individuals* with any favourable variation, and by the destruction of those with any unfavourable deviation of structure.'[84] Elsewhere he explains that 'every slight modification, which in the course of ages chanced to arise, and which in any way favoured the *individuals* of any species, by better adapting them to their altered conditions, would tend to be preserved.'[85] And in distinguishing artificial selection from natural selection, Darwin observes, 'Man selects only for his own good; Nature only for that of the being [i.e., the individual organism] which she tends.'[86] Natural selection preserves traits because of their benefit to the organisms that possess them.

[82] Dawkins (1982) 81.
[83] Ibid.
[84] Darwin (1998) 159, emphasis added.
[85] Ibid., 68, emphasis added.
[86] Ibid., 69.

Dawkins, however, 'argue[s] that the fundamental unit of selection, and therefore of self-interest, is not the species, nor the group, nor even, strictly, the individual. It is the gene, the unit of heredity.'[87] Although organisms are undeniably benefited by adaptations, which enable the organisms to leave offspring, it is actually the gene for whose benefit the adaptations exist. Properly speaking, fitness attaches to genes, rather than organisms, and therefore, natural selection acts on genes, preserving a particular gene because it is successful in 'helping to program the bodies in which it finds itself to survive and to reproduce.'[88] Under the influence of selection, the gene pool comes to be dominated by 'genes whose replicas in previous generations were successful in getting themselves copied . . . by means of [their] influence on the development of bodies.'[89] Thus, evolution is the story of genes struggling to make their way in the world, and prospering or perishing, at least in part, according to their ability to make a living by building fit organisms.

But this claim that genes are the units of selection raises two important questions? First, why does it matter to Dawkins that the gene rather than the organism be regarded as the unit of selection in this 'theoretical' debate; and second, how does he propose to demonstrate that the gene does in fact fill this rôle in evolutionary theory? As for the first of these questions, Dawkins is trying to show, as we have seen, that evolution by natural selection is a process best represented as a competition among genetic units to survive and leave copies of themselves. Evolution occurs, in part, because of natural selection; consequently, if Dawkins can show that selection actually operates at the level of the genes, then he can reasonably conclude that evolution is really a process that occurs at the genetic level. If the gene is the unit of selection, because genes 'have some effect on the world, which influences their chances of surviving,'[90] then evolution can indeed be represented as the survival of the fittest genes, as Dawkins has set out to demonstrate.

---

[87] Dawkins (1989) 11.
[88] Ibid., 88.
[89] Dawkins (1981) 571.
[90] Dawkins (1982) 95.

The orthodox Darwinian view regards evolution as the modification over time of lineages of organisms by the action of natural selection, which differentially favors the fitter organisms over those organisms less well adapted to 'the constellation of environmental conditions they encounter[ ] during their lifetime.'[91] In order to replace this view of evolution with his selfish gene perspective, Dawkins must show that organisms are merely the by-products of genes, and that natural selection is properly described as choosing among genes. In the previous section, I presented three arguments for the organisms-as-by-products conclusion; in the balance of this section, I will attempt to explain Dawkins' assortment of arguments for the position that genes, rather than organisms, are the true targets, or units, of natural selection.

The first argument derives from the results of the previous section, in which Dawkins concludes that genes create organisms for their own benefit. Both sides in this debate agree that natural selection acts *directly* on organisms, preserving or eliminating organisms according to their fitness in a particular environment. As Dawkins admits, whatever the *ultimate* unit of selection may turn out to be, 'the proximal unit of selection is usually regarded as something larger [than the gene], usually an individual organism.'[92] However, if the organisms are themselves merely the by-products of a competition among genes to preserve themselves, then selection is really choosing among the genes, which are simply judged according to their skill at constructing organisms. The organisms are the representatives of the genes, as it were, in the struggle for survival and reproduction. Thus, genes are the ultimate target of natural selection, which acts indirectly on the genes by choosing directly among their proxies, the organisms constructed by the genes for their own benefit.

The second argument, though briefly stated, will require some time to carry through to its conclusion. The argument depends upon two premises. It declares, first, that an entity must possess certain properties to qualify as a unit of selection, and second, that only

[91] Mayr (1988) 96.
[92] Dawkins (1982) 116.

genes have these properties. Simply speaking, organisms and other higher units do not have what it takes to be the unit of selection.

'[W]e begin by identifying the properties that a successful unit of selection must have[;] . . . these are longevity, fecundity, and copying-fidelity.'[93] In order for natural selection to change a lineage, to bring about the evolution of a population or species, something must survive from generation to generation, largely unchanged. If natural selection is understood, in Darwin's words, as the 'principle [that] each slight variation, if useful, is preserved,'[94] then the unit of selection must be something that endures, an entity that is preserved from parent to offspring, through generations of the evolving lineage. This is Dawkins' requirement of *longevity*.

Selection leads to evolution because adaptive traits, which arise by chance, are heritable. Darwin, noting the importance of heredity in his explanation of evolution, writes that individuals with useful variations 'will have the best chance of being preserved in the struggle for life; and from the strong principle of inheritance they will tend to produce offspring similarly characterised.'[95] Natural selection transforms a population by differentially preserving certain variations over others, but these favorable variations can not spread throughout the population unless they are heritable. Moreover, there can be no 'descent from a few created forms with subsequent modification,'[96] as Darwin understood evolution, unless there is something that survives from generation to generation, transmitted from parent to offspring, which permits the accumulation of modifications.

Dawkins gives the name of 'replicators'[97] to the long-lived heritable entities that pass from parent to offspring, surviving generation after generation in the form of copies, and 're-

[93] Dawkins (1989) 35.
[94] Darwin (1998) 52.
[95] Ibid., 104-05.
[96] Ibid., 376.
[97] Dawkins (1989) 15.

tain[ing] their structure largely intact.'[98] Since these are the properties required of the unit of selection, we can identify the unit of selection as the replicator, a 'potential[ly] immor- tal[ ]'[99] entity capable of making accurate copies of itself; or, in Dawkins' words, it is an entity with *longevity*, *fecundity*, and *copying fidelity*. Indeed, '[n]atural selection may usually be safely regarded as the differential survival of replicators.'[100]

To complete his argument that the gene is the unit of selection, Dawkins must show that 'small genetic units have these properties [while] individuals, groups, and species do not.'[101] 'The cornerstone of the argument . . . [is] the assumption that genes are potentially immortal, while bodies and all other higher units are temporary.'[102] Only individual genes have sufficient longevity, fecundity, and fidelity to qualify as units of selection. Indeed, Dawkins has '*defined* the gene in such a way that [he] cannot help being right!'[103] Accordingly, '[a] gene is defined as any portion of chromosomal material that potentially lasts for enough generations to serve as a unit of natural selection.'[104]

Let me try to explain what he means. There is much disagreement among biologists and philosophers of biology about the proper definition of a gene, but all parties to the debate acknowledge that a gene is a sequence of nucleotides, a fragment of DNA, some 'portion of a chromosome'; the disagreement concerns 'how big a portion' of the chromosome we want to treat as the gene.[105] Dawkins is not particularly interested in joining this debate: 'We can define a word how we like for our own purposes, provided we do so clearly and unambiguously.'[106] Rather, he is concerned to identify the portion of the chromosome that embodies the properties required of the replicator, the unit of selection. He seeks to identify

---

[98] Dawkins (1982) 84, quoting David Hull, 1980, 'Individuality and Selection,' *Annual Review of Ecology and Systematics*, 11, 311-332.
[99] Dawkins (1989) 36.
[100] Dawkins (1982) 87.
[101] Dawkins (1989) 33.
[102] Ibid., 40.
[103] Ibid., 33.
[104] Ibid., 28.
[105] Ibid., 29.
[106] Ibid., 28.

the genetic unit with sufficient longevity, fecundity, and copying-fidelity to serve as the unit of selection. The portion of the chromosome that possesses these properties he defines, for his purposes, as the gene, or in the title of his book, the selfish gene. A gene, then, on his definition, is nothing more than 'a replicator with high copying-fidelity.'[107]

Dawkins proceeds to distinguish this unit, the selfish gene with high copying-fidelity, from other possible units of selection, at and above the genetic level. At one time or another, he dismisses as non-replicators, or as evolutionarily insignificant replicators, the single nucleotide, the chromosome, the genome, the individual organism, the group of organisms, the population, and the species. To understand his preferred choice of the gene for the unit of selection, we must consider his arguments against these other candidates.

First, the single nucleotide, a component of DNA and RNA, would seem, on its molecular face, to be the ideal candidate for the replicator; it appears to satisfy, in every respect, the requirements that Dawkins sets forth for the unit of selection. Nucleotides are very long-lived entities that copy true. Nucleotides form the links, or bases, in the long molecular chains of DNA, which constitute genes, so a particular nucleotide can be copied for countless generations as part of a successful gene. Additionally, a nucleotide can be copied indefinitely as part of a non-coding region of the genome,[108] and thereby survive in generations of bodies even without producing any phenotypic effects. And finally, a nucleotide that is a part of one gene can be relocated, by means of crossing over during meiosis, to another gene, and outlive even the gene of which it was originally a part.

So why should we not regard the individual nucleotide as the unit of selection? After all, within a given genetic environment, the individual nucleotide can exert phenotypic effects. As Dawkins concedes, '[i]t is changes at the single nucleotide level that are responsible for evolutionary significant phenotypic changes, although of course the unvarying remain-

[107] Ibid.

[108] Much of the DNA in a genome has no known phenotypic effects. For example, by some estimates, 'well over 90 per cent of the DNA of the human genome has no known function. Much of this DNA consists of repeating sequences of bases, and hence has been called repetitive DNA.' Rose (1997) 123.

der of the genome is necessary to produce a phenotype at all.'[109] Perhaps Dawkins should have entitled his book '*The Selfish Nucleotide*?'[110] He concludes that such a work is unnecessary, because 'while it may not be strictly wrong to say that an adaptation is for the good of the nucleotide, i.e., the smallest replicator responsible for the phenotypic differences concerned in the evolutionary change, it is not helpful to do so.'[111] The reason it is not helpful to identify the nucleotide as the unit of selection is that 'the nucleotide only exerts a given type of power when embedded in a larger unit,'[112] which Dawkins calls the gene. Since this gene is also a replicator, it is 'much more useful'[113] to treat this larger entity as the unit of selection.

Of course, the phenotypic effect of a gene also depends upon its genetic environment, but the phenotypic effect exerted by a nucleotide is, in addition, critically dependent upon its location, or 'its sequential context,' within the larger unit. A gene, however, 'is large enough to have a consistent phenotypic effect, relatively, though not completely, independently of where it lies on the chromosome.'[114] We can ascribe a phenotypic effect to a particular gene, but '[i]t is meaningless to speak [for instance] of the phenotypic effect of adenine,'[115] one of the four nucleotide bases in DNA. We must 'speak [instead] of the phenotypic effect of substituting adenine for cytosine at a named locus within a named [gene].'[116] Hence, the phenotypic effect of a nucleotide is tied to the position of the nucleotide within the gene, whereas the phenotypic effect of a gene is less sensitive to its location on the chromosome, and in relation to other genes.

Dawkins, unfortunately, does not develop this argument, but I think it may be elaborated as follows. The unit of selection, whatever else it is, is an independent actor that

[109] Dawkins (1982) 90.
[110] Ibid.
[111] Ibid., 91.
[112] Ibid.
[113] Ibid.
[114] Ibid., 92.
[115] Ibid., 91.
[116] Ibid.

strives, by means of its effects on the world, to leave descendants, or replicas of itself. But the single nucleotide, because of its greater dependence upon its location in the genetic unit, is less independent, so to speak, than the gene in its ability to bring about phenotypic effects, and thereby influence its own destiny. The single nucleotide is rather like a constituent of an independent actor, than an evolutionary agent in its own right. Therefore, it is preferable to recognize the replicating fragment of DNA, instead of its individual component bases, as the unit of selection.

Second, Dawkins argues that the chromosome is not the unit of selection. It is not a replicator, which survives in copies through generations of bodies, because '[t]he life-span of a chromosome is [but] one generation.'[117] As biologist Ernst Mayr explains, 'chromosomes are not permanently inviolable structures because genes are separated from each other in every generation owing to the process of crossing-over during meiosis.'[118] Therefore, since chromosomes are 'almost bound to be split by crossing-over before the next generation,' they are 'not even *potentially* long-lived,' and 'their [copying] "fidelity" is zero.'[119] Chromosomes do not have sufficient longevity or long-term structural integrity to qualify as the unit of selection.

But the fact of crossing-over tells against not only the chromosome as the unit of selection, but also against any long segment of the chromosome. This is 'simply because [longer segments of the chromosome] are more likely to be broken by crossing over' than shorter segments.[120] Indeed, since the point (called the chiasma) at which the chromosome is 'cut' during crossing-over can occur anyplace along the chromosome, the likelihood that a particular portion of a chromosome will be cut during crossing over is proportional to the length of that portion. At one extreme, the whole chromosome is broken by every instance

---

[117] Dawkins (1989) 30.
[118] Mayr (1988) 428. During *meiosis*, the process of cell division that gives rise to *gametes*, or sex cells, each chromosome pairs with its *homologous* counterpart, and may randomly swap segments in an exchange called *crossing over*, thus resulting in a *recombination* of genes on each chromosome that participated in the exchange.
[119] Ibid., emphasis in the original.
[120] Ibid.

of crossing over; and at the other, the nucleotide, the base unit of DNA, is entirely unaffected, the chiasma always falling between nucleotides, so to speak. Thus, 'a long portion of chromosome, even if successful in terms of its phenotypic effects, will not be represented in many copies in the population,'[121] and can not be the unit of selection.

In order to identify the proper unit of selection, '[w]e must choose [a] . . . portion of chromosome . . . that . . . is small enough to last, at least potentially, for many generations before being split by crossing-over; small enough to have a 'frequency' [of copies] that can be changed by natural selection';[122] but large enough, as we saw above, to have a phenotypic effect somewhat independent of its genetic context. This unit, larger than the single nucleotide, and very much smaller than a whole chromosome, Dawkins denominates the gene. Accordingly, he defines a gene 'as a piece of chromosome which is sufficiently short for it to last, potentially, for *long enough* for it to function as a significant unit of natural selection.'[123] Although Dawkins admits there is no 'hard and fast answer' to what counts as 'long enough,' he concludes that '[t]he largest practical unit of natural selection—the gene—will usually be found to lie somewhere on the scale between cistron and chromosome.'[124] It is perhaps not possible to be more specific.

Third, Dawkins 'reject[s] the whole sexual genome as a candidate replicator, because of its high risk of being fragmented at meiosis.'[125] Like the chromosome, the genome does not endure for long enough to be the unit of selection, since the sexual genome is lost in every generation, and it is not passed on intact from generation to generation. 'Our [individual] genes may be immortal but the *collection* of genes that is any one of us is bound to crumble away.'[126] As G. C. Williams vividly explains, 'Socrates' genes may be with us yet,

[121] Ibid., 88.
[122] Ibid., 88-89.
[123] Dawkins (1989) 35-36, emphasis in the original.
[124] Ibid., 36. A cistron is a segment of DNA that codes for a *polypeptide*, a chain of amino acids.
[125] Dawkins (1982) 91.
[126] Dawkins (1989) 199, emphasis in the original.

but not his genotype, because meiosis and recombination destroy genotypes as surely as death.'[127]

Moreover, Dawkins stresses repeatedly that although genes must coöperate with one another in order to exert their phenotypic effects, genes are nevertheless selected individually. Genes may act as a group, but they are selected as individuals. 'Tempting as it is, it is positively wrong to speak of . . . genes being selected "as a group." Each [gene] is selected as a separate selfish gene, [even though] it flourishes only in the presence of the right set of other genes.'[128] After all, '[t]he genotype may be a "physiological team," but we do not have to believe that that team was necessarily selected as a harmonious unit in comparison with less harmonious rival units'; instead, genes are selected individually for their ability to 'prosper in each others' presence.'[129] In order for genes to create a successful organism that will ensure the transmission into future generations of gene copies, the genes must coöperate with one another. '[G]enes are highly gregarious' entities that 'have phenotypic effects on bodies, but they do not do so in isolation.'[130]

To illustrate this distinction between genes as coöperative actors, working together to build an organism, and genes as 'self-seeking agents of life,'[131] seeking to preserve only themselves, Dawkins employs his well-known analogy of oarsmen competing for seats on a rowing team. A single oarsman can not win a boat race without the help of teammates. 'Rowing the boat is a coöperative venture, but some men are nevertheless better at it than others.'[132] Dawkins imagines a rowing coach trying to choose his best crew from the pool of candidates trying out for the team. The coach proceeds by randomly assembling each day several different trial crews, and racing these against one another, keeping careful records of the results. 'After some weeks of this it will start to emerge that the winning boat often

[127] Williams (1966) 24, cited in Sober (1984) 250. Genotype is the genetic composition of an organism.
[128] Dawkins (1989) 258.
[129] Dawkins (1982) 240.
[130] Ibid., 113.
[131] Dawkins (1989) 38.
[132] Ibid.

tends to contain the same individual men[;] . . . [o]ther individuals seem consistently to be found in slower crews.'[133] The coach then puts together his team from the individuals who tended '*on average* . . . to be in the winning boat[s].'[134]

Of course, in this illustration '[t]he oarsmen are genes,'[135] and the point of the analogy is that the oarsmen are selected individually for their ability to perform well in a crew composed of other oarsmen. The successful oarsman is one with 'a good team spirit.'[136] But notice that the coach does not select his oarsmen as a group; rather, he chooses individuals who are good team players, who coöperate with their teammates. 'Similarly, the fact that genes are selected for mutual compatibility does not necessarily mean we *have* to think of groups of genes as being selected as units.'[137] Natural selection acts on the several genes, not genomes.

However, as Dawkins recognizes, there is one wrinkle in this pat dismissal of the genome as a possible unit of selection. As we have said, the chief difficulty with regarding the genome as a replicator is that it is fragmented in every generation by meiosis. But this is not strictly true. For only the genomes of organisms that reproduce sexually are actually subject to meiosis. Asexual organisms reproduce without forming gametes.[138] 'Aside from the occasional mutational change,'[139] the asexual genome is transmitted unchanged generation after generation. Thus, Dawkins admits that a 'genome [may be a replicator] . . . if it is reproduced asexually.'[140] 'If there is no sex we can . . . treat the entire genome of an asexual organism as a replicator.'[141] Presumably, then, though I do not find that Dawkins ever says as

[133] Ibid.
[134] Ibid., emphasis in the original.
[135] Ibid.
[136] Dawkins (1989) 84.
[137] Ibid., 85.
[138] Asexual reproduction 'involves a single parent organism splitting, budding, or fragmenting to make one or more new individuals' which inherit the parent's genome intact. Snyder and Rodgers (1995) 305.
[139] Ibid., 310.
[140] Dawkins (1982) 109.
[141] Ibid., 95.

much, the asexual genome *can* be regarded as a unit of selection. He would probably be inclined, however, to discount its importance in the evolutionary drama.[142]

Fourth, Dawkins argues that, contrary to the traditional understanding of Darwinian evolution, the individual organism is not the unit of selection. '[T]he organism itself is *not* a replicator.'[143] Dawkins gives two arguments for this conclusion. The first argument, which should be quite familiar to us by now, we need not belabor. Since the individual organism does not survive in the form of copies for more than a single generation, it lacks the longevity and copying fidelity required of the unit of selection. Dawkins predictably touts 'the fragmenting effects of meiosis as a reason for not regarding sexually reproduced organisms as replicators.'[144] Rather than stable entities that survive largely intact generation after generation, 'individuals and groups [of individuals] are like clouds in the sky or dust-storms in the desert. [As] . . . temporary aggregations or federations[,] . . . they are not stable through evolutionary time.'[145] An organism is nothing more than 'a communal *vehicle* for [the true] replicators,' the genes, which create and 'travel about . . . inside [this] . . . tool of replicator propagation.'[146]

The second reason not to regard the organism as a replicator, and hence a possible unit of selection, is that to do so would be to violate 'the "central dogma" of the non-inheritance of acquired characteristics.'[147] Natural selection works by 'scrutinizing . . . every variation, [and] . . . rejecting that which is bad, preserving and adding up all that is good.'[148] Accordingly, in order for an entity to be the unit of selection, it must be capable of passing on to the next generation the changes that happen to occur to it; such changes are reflected in all the subsequent copies. As Dawkins explains, if we imagine an unbroken lineage of

---

[142] Dawkins also admits, as we shall see later in this section, that the species may be a replicator, but he concludes that, as a unit of selection, it is probably not as evolutionarily significant as the replicating gene.
[143] Dawkins (1982) 95, emphasis in the original.
[144] Ibid., 97.
[145] Dawkins (1989) 34.
[146] Dawkins (1982) 112.
[147] Ibid., 97.
[148] Darwin (1998) 70.

replicators, with each new generation being copied from the previous generation, any 'blemish [that] appears somewhere along [the way] will be passed on to all subsequent links in the chain.'[149] But the accidents of life that organisms experience are not inherited by their offspring. A parent's scars and broken bones, for example, are not passed on to its young. This is what is meant by the non-inheritance of acquired characteristics. However, a change '(mutation) to a gene in the germ-line[150] . . . *can* be passed on to the next generation.'[151] Consequently, a gene is a replicator, but '[a]n organism . . . is not a replicator, not even a crude replicator with poor copying fidelity.'[152]

Finally, we have to consider whether the unit of selection might be identified with any 'larger units, groups of organisms, species, communities of species, etc.'[153] I shall be brief in presenting Dawkins' reasons for discounting these as units of selection, partly because Dawkins himself spends little time discussing them, but also because even staunch defenders of the orthodox Darwinian position have expressed doubts about whether selection acts above the level of the individual organism. For example, biologist Ernst Mayr, who 'favor[s] the individual' as the unit of selection, [154] has concluded that 'group selection among animals . . . is not supported by any evidence.'[155] And Dawkins finds, with a swipe at supporters of group selection, that 'group-selection theory now commands little support within the ranks of those professional biologists who understand evolution.'[156]

We need not be concerned here with the actual extent or source of professional dissatisfaction with group selection. Dawkins' own dissatisfaction stems from familiar grounds. Group selection, Dawkins explains, 'is the hypothetical process whereby natural selection chooses among whole groups of organisms, as opposed to choosing among indi-

---

[149] Dawkins (1982) 97.
[150] I.e., the line of cells that produce the *gametes*, or sex cells.
[151] Dawkins (1982) 99.
[152] Ibid.
[153] Ibid.
[154] Mayr (1988) 101.
[155] Mayr (1988) 79.
[156] Dawkins (1989) 8. For a contemporary defense of some group selection ideas, see Elliott Sober and David Sloan Wilson, 1998, *Unto Others: The Evolution of Altruism* (Cambridge, Mass.: Harvard University Press).

viduals.'[157] But leaving aside some practical difficulties with the theory,[158] into which we need not enter here, a group of organisms can not be the unit of selection because, in the evolutionary drama, it is not 'even a *candidate* for the replicator rôle.'[159] As a result of 'immigration and emigration, the destruction of the integrity of groups by the movement of individuals into and out of them,' groups lack the 'stab[ility] through evolutionary time' [160] required of a unit of selection. Additionally, although groups or populations of organisms 'may last a long while, . . . they are constantly blending with other populations and so losing their identity.'[161] Therefore, even if there were no other problems with group selection, the fact that groups do not preserve their identity largely intact generation after generation is sufficient to discount them as units of selection.

But what of species?[162] Can a whole species be the target of selection? First, we must clarify the meaning of species selection. During the process of speciation, one or more daughter species develop from an existing species; this occurs when a population, the incipient species, 'diverge[s] so much from the parent population that interbreeding can no longer occur between them,'[163] and each evolves on its own. As Mayr has emphasized, this divergence typically requires the isolation of the daughter population. 'What actually seems to happen almost universally is that the first step is a split of the population into two, owing to the establishment of a founder population beyond the species border or owing to a splitting of the population following the origin of a new geographic barrier.'[164] Each new species can

---

[157] Dawkins (1981) 563.
[158] Dawkins writes that 'except under very special conditions, biologists now agree that group selection cannot work in nature.' Ibid., 559.
[159] Dawkins (1989) 254.
[160] Dawkins (1982) 99-100.
[161] Ibid., 100.
[162] A species, it turns out, is confoundedly difficult to define. It would draw us too far afield here to discuss the different definitions that have been proposed, but for a wide-ranging presentation of attempts to find a satisfactory definition, see Ereshefsky, M., ed., 1992, *The Units of Evolution: Essays on the Nature of Species* (Cambridge, Mass.: MIT Press).
[163] *A Dictionary of Biology*, 1996, 3rd edition (Oxford: Oxford University Press) 477.
[164] Mayr (1988) 144.

then spawn daughter species of its own, forming separate lineages of species through evolutionary time.

Now, it may be the case that in a particular lineage of species the 'new species differ from their predecessors randomly with respect to major trends';[165] for example, in a lineage of horse species, 'new species of horses [might] arise equally often at sizes smaller and larger than their ancestors.'[166] But if, let us suppose, the species of smaller horses are more likely to go extinct than the larger horse species, then the horse lineage will show an evolutionary trend towards larger size. This trend would then be the result of species selection: the differential selection of species, or, in our example, the preservation of species with larger size.[167]

Now, in regard to the possibility of species selection, Dawkins grudgingly admits it, with the proviso that species selection almost certainly lacks the 'power to put together complex adaptations such as eyes and brains,'[168] 'because it is too slow.'[169] He also allows 'that there may be a case for regarding the gene-pool of a reproductively isolated group, such as a species, as a replicator.'[170] A gene pool, on some accounts, should be considered 'a coadapted unit, homeostatically buffered against change,'[171] which, like a replicator, endures relatively intact over very long stretches of time. As Mayr explains, '[w]hat characterizes most of the well-documented cases [of speciation] . . . is the relative stability of the parental species and also of the neospecies once it has completed the speciation process';[172] indeed, 'in most cases there is very little change of the phenotype over millions of years.'[173] In addition, the gene pool of a species resembles a replicator in that it preserves and transmits

[165] Dawkins (1982) 105.
[166] Ibid., citing Stephen Jay Gould, 1980, 'The Promise of Paleobiology as a Nomothetic, Evolutionary Discipline,' *Paleobiology* 6, 96-118.
[167] The important point about species selection is that the 'favoured qualities at the species level may in theory have nothing to do with the qualities that are favored by selection within species.' Dawkins (1982) 294.
[168] Dawkins (1982) 106.
[169] Ibid., 108.
[170] Ibid., 109.
[171] Ibid., 108.
[172] Mayr (1988) 415.
[173] Ibid., 109. Note that the long-term stability of the phenotype, which paleontologists can trace in the fossil record, provides strong evidence of the stability of the genotype, and thus of the gene pool, out of which the genotype coalesces.

into the future any changes that happen by chance to arise in it. However, even if the gene pool of a species is taken to be a replicator, it is by no means a very significant one, for species selection is 'unlikely to explain complex adaptations.'[174] Thus, the selfish fragment of DNA remains 'our archetypal replicator.'[175]

This completes the two arguments for the conclusion that the gene is the unit of selection. Let me summarize our results. First, the gene must be accounted the unit of selection because genes create organisms, and natural selection, by preserving some organisms and eliminating others, is ultimately targeting the genes, which give rise to the organisms. As the products of genes, organisms are merely the form in which genes happen to expose themselves to the life or death judgements of natural selection. Another way of explaining the argument is this. According to the orthodox understanding of evolution by natural selection, the origin and modification of species is caused by the differential survival of the organisms that constitute the species; that is, by natural selection acting on individual organisms. But if genes create these organisms, then it is actually more correct to say that '[g]enes cause evolution, because they cause something else (i.e., phenotypes) which themselves cause evolution.[176]

Second, the gene is the unit of selection, because the unit of selection must have the properties of what Dawkins calls a 'replicator,' and of all the candidates proposed as the unit of selection, the gene 'is the most convenient approximation'[177] of the replicator. Evolution by natural selection is a cumulative process, in which a population, whether of organisms or genes, changes over time because of the differential survival of individuals in the population. But 'in order for this selective death to have any [evolutionary] impact on the world,' selected individuals must be capable of perpetuating themselves in the form of cop-

[174] Dawkins (1982) 109.
[175] Ibid.
[176] Sober (1984) 229.
[177] Dawkins (1981) 569.

ies that can exist 'for a significant period of evolutionary time.'[178] To put this another way, natural selection is the survival of the fittest, but selection can not bring about evolution unless the fittest have the potential to survive generation after generation in the form of copies, passing on any changes that happen to arise.

The next step in this argument, then, is to consider which of the candidates for the unit of selection best satisfies the requirements of a replicator. Dawkins admits that the nucleotide, the gene, the asexual genome, and the reproductively isolated species may all be regarded as replicators. But there are reasons for preferring the gene to each of the others.[179] The gene is more independent than the nucleotide in its ability to cause phenotypic effects and influence its own survival; genic selection is more important to evolution than selection of asexual genomes; and selection of the gene has greater power to drive evolution than selection of the species.

In this first section of the chapter, I have tried to explain why Dawkins regards the gene as the most important actor on the evolutionary stage, and why he thinks that Darwinian evolution is best represented as a competition among selfish genes to leave descendents. Now we turn to an evaluation of this selfish gene view of nature.

Criticisms of the Selfish Gene View of Nature

A. Introduction

As with many controversial theories, Dawkins' gene's eye view of nature has attracted its share of criticism. Numerous arguments, some more successful than others, have been advanced to show that, notwithstanding Dawkins' efforts, evolution by natural selection is not in fact best represented as a competition among selfish genes. In a chapter of this size, it would be possible either to present brief summaries of a large number of these arguments, or to treat in greater detail a handful of the more forceful and cogent arguments. In

[178] Dawkins (1989) 33.
[179] Unfortunately, Dawkins does not expatiate upon the reasons for preferring the gene to each of these other replicators.

this section, I propose to pursue the second of these options, concentrating on the arguments that seem to me most decisive against Dawkins' view of evolution.

By way of introducing these criticisms of the selfish gene position, let us briefly distinguish it from the traditional Darwinian view. The simplest way of distinguishing the two views of life is that the traditional Darwinian, whom Mayr calls a naturalist, 'plac[es] the organism, rather than the gene, at the centre of life.'[180] Accordingly, the naturalist denies that organisms can be 'reduc[ed] to "nothing but" machines for the replication of [their] DNA,'[181] and regards genes as 'no more and no less than an essential part of the toolkit with and by which organisms construct their own futures.'[182] He is convinced that 'the individual and not the gene must be considered the target of selection.'[183] It is natural selection acting on individual organisms that drives evolution.

Naturalists regard Dawkins' 'genocentric biology,'[184] which they have variously labeled 'ultra-Darwinism,'[185] 'neo-Darwinism,'[186] and 'beanbag genetics,'[187] as merely an 'excessive preoccupation with one aspect of [biological] reality.'[188] The selfish gene view, by treating the gene as 'the currency of evolution,'[189] understands evolution, natural selection, and organisms in terms of genes. Thus, evolution is 'defined as a change in gene frequencies, the replacement of one allele by another';[190] the gene is 'consider[ed] . . . to be the target of selection';[191] and 'organisms are not what they seem to be—integrated entities with lives and natures of their own—but complex molecular machines controlled by the genes

---

[180] Rose (1997) x.
[181] Ibid., 6.
[182] Ibid., 137.
[183] Mayr (1988) 101.
[184] Goodwin (1994) 29.
[185] Rose (1997) 209.
[186] Collier (1981) 339.
[187] Mayr (1988) 449.
[188] Goodwin (1994) viii.
[189] Mayr (1988) 100.
[190] Ibid., 423.
[191] Ibid., 405.

carried within them.'[192] In this selfish gene picture of nature, '[o]rganisms have disappeared as the fundamental units of life[,] . . . [replaced by] genes, which have taken over the basic properties that used to characterize living organisms.'[193] Even life itself is characterized in terms of genes. As biologist Steven Rose portrays the ultra-Darwinist position, 'the purpose . . . of life is reproduction, reproduction of the genes embedded in the "lumbering robots" which constitute living organisms . . . Every living process is therefore in some way directed towards this grand goal.'[194]

In the first section of this chapter, I described the selfish gene view of nature in terms of a main thesis and two supporting sub-theses. The main thesis is that evolution by natural selection is best represented as a competition among genes to leave descendants. The two sub-theses are, first, that organisms are created by genes, a sort of 'reduction from phenotype to genotype'; and, second, 'that the gene is the unit of selection.'[195] If we can remove these two stays of the main thesis, and show that genes do not create organisms and that the gene is not the target of selection, then we shall have done much to collapse Dawkins' genocentric view of evolution.

B. Genes Do Not Create Organisms

For Dawkins, as we have seen, genes contain the instructions for making an organism. He regards the genome as 'a set of instructions which, if faithfully obeyed in the right order and under the right conditions, will result in a body.'[196] Biologist Brian Goodwin describes this as '[t]he idea of a genetic program in an egg that specifies all the details of the organism by the information it contains.'[197] Indeed, Dawkins, observing a seeding willow tree outside his window, exclaims, 'It is raining DNA . . . It is raining instructions out there;

[192] Goodwin (1994) vii-viii.
[193] Ibid., 1.
[194] Rose (1997) 209.
[195] Collier (1981) 339.
[196] Dawkins (1982) 175.
[197] Goodwin (1994) 42.

it's raining tree-growing, fluff-spreading algorithms. This is not a metaphor, it is the plain truth. It couldn't be any plainer if it were raining floppy discs.'[198]

The naturalists, however, deny that this description is in fact the plain truth; it has been characterized as 'misleading in almost every respect.'[199] Any genocentric 'claim[ ] that understanding genes and their activities is enough to explain the properties of organisms . . . is simply wrong,' because 'organisms cannot be reduced to the properties of their genes.'[200] It takes much more than genes to make an organism.

I propose to consider here only four problems with Dawkins' claims that genes create organisms. The first problem is that 'an egg contains more than just . . . DNA,'[201] and these additional contents play an indispensable rôle in the creation of an organism. An organism begins life as a single cell. In the case of sexually produced organisms, this single cell is a fertilized egg, called a zygote. The zygote becomes a mature adult organism through the complex process known as development. 'To develop normally, the egg cell must contain a great array of complex biochemical machines[, including] . . . basal bodies and microtubule organizing centers, cytoplasmic chemical gradients, DNA methylation patterns, and membranes and organelles, *as well as DNA*.'[202] Although the intricate details of development are far beyond the scope of this chapter, and the competence of its author, we can say enough about the process to call into question the unique creative power of genes. Consider just a few of the functions performed by the non-genetic components of the fertilized egg.

First, in addition to DNA, the cell has 'all the cellular apparatus required to bring [the paternal and maternal] DNA together and persuade the otherwise inert fibres to play their part in the cellular orchestra.'[203] The genes contained in the sperm and egg do not unite of their own accord to form the genotype. Second, '[f]rom th[e] moment of concep-

---

198 Dawkins (1986) 111, cited in Rose (1997) 121.
199 Rose (1997) 121.
200 Goodwin (1994) 3.
201 Rose (1997) 129.
202 Sterelny and Griffiths (1999) 96, emphasis added.
203 Rose (1997) 129.

tion on, the maternal cellular machinery is responsible for directing the activation of particular genes and hence the synthesis of specific proteins,'[204] some of these proteins in turn helping to regulate and activate other genes. Third, protein synthesis itself is an elaborate process carried out by cellular machinery.[205] And fourth, 'the egg carries the organization required for accurate replication of the DNA.'[206] '[T]he capacity of DNA to make accurate copies of itself and to produce proteins via mRNA is very much dependent upon a highly organized context: the living cell.'[207] And the contents of this living cell are not simply the products of maternal DNA; as we shall see, much of what 'parents pass on . . . to their offspring [including] genes, cellular chemistry, and other cell structures,'[208] were themselves inherited by the parents from their forebears.

Given the essential rôle played by the cellular machinery in development and reproduction, several preliminary points can be made. First, Dawkins' distinction between replicators and vehicles, upon which his genocentric view depends, becomes difficult to maintain, because genes are not in fact 'autonomous replicator[s]'[209] capable of accurately replicating themselves. DNA is in 'dynamic interaction . . . [with] the cellular system in which it is embedded.'[210] '[T]here is no privileged class of replicators among the many material causes that contribute to the development of an organism.'[211] Second, we must question the claim that genes are independent causal agents that create bodies for their own preservation. Given even this very brief account of the importance of the non-genetic components of the

---

204 Ibid., 130.
205 The first step, in which segments of DNA are converted to messenger RNA (mRNA), is called transcription. In this process, the double-stranded helical DNA molecule is uncoiled and separated, while an enzyme called RNA polymerase assembles the mRNA molecule by traveling along the DNA, pairing free RNA nucleotides with the corresponding bases in the DNA chain. In the second step, called translation, the newly formed mRNA is transported outside of the nucleus to a small body called a ribosome. The ribosome moves along the mRNA chain, assembling a polypeptide (i.e., a chain of amino acids; a protein is a polypeptide usually consisting of 100-300 amino acids) by linking together amino acids that correspond with the sequence of nucleotides in the mRNA molecule.
206 Goodwin (1994) 36.
207 Ibid., 5.
208 Sterelny and Griffiths (1999) 97.
209 Goodwin (1994) 36.
210 Rose (1997) 130.
211 Sterelny and Griffiths (1999) 94.

cell, it is difficult to construe genes as the sole causal agents involved in the creation of an organism, and if DNA is recognized as only one of the causes of development and reproduction, then it is by no means obvious that it is the most important of such causes. Third, genes do not directly create even proteins. Although the gene serves as a template for the synthesis of a protein, it is not clear, looking only at the activities within the cell, why the cell must be thought to serve the genetic template, rather than the template serve the cell. What we see in the cell is an intricate, well-coördinated *system* of interacting components, one of which is a set of nuclear DNA molecules. That all of life can be understood in terms of that single component of the cell, as Dawkins would have us believe, is not a claim that can be substantiated by studying the structure and function of the cell.

The second problem with Dawkins' view that genes create organisms is something I referred to above: an organism begins life as a single cell. For Dawkins, reproduction amounts to the transmission from parent to offspring of genetic information. But what is actually transmitted from generation to generation is not merely a set of genetic instruction, but a complete cell, one of whose components, albeit an essential one, is a set of DNA molecules. As Sterelny and Griffiths note, 'the idea that nuclear genes are all an organism inherits in the cells carrying the gametes is simply out of date.'[212] 'This is why Dawkins' claim that his garden willow tree is simply 'raining DNA' is so biochemically wide of the mark.'[213] More precisely, the tree is raining seeds, tiny potential willow trees, each of these seeds containing an embryo and a supply of nutritive tissue. Under the right environmental conditions, the embryo—with its measure of DNA and RNA, its ribosomes and energy-producing mitochondria, and its assortment of other cellular machinery—may in time develop into a willow tree.

In reproduction, '[p]arts of the adult create a new whole organism by a process in which an initially simple structure develops into the adult form, which then produces the

[212] Ibid., 96.
[213] Rose (1997) 130.

parts that will start the next generation.'[214] Indeed, we can 'describe[ ] reproducing organisms as systems having the property that a part can produce a whole, so that they can regenerate their own natures.'[215] Thus, a life cycle begets another life cycle; that is to say, '[e]ach life cycle is initiated by a period in which the functional structures characteristic of the lineage must be reconstructed from relatively simple resources.'[216] The genetic material in the nucleus of the cell is only *one* of these resources. Reproduction, like development, is a process in which earlier stages construct later stages, because the 'resources that construct later stages . . . are constructed by earlier stages.'[217] Thus, to single out genes as the predominant actors in this process seems thoroughly arbitrary, and to describe reproduction simply as the transmission of genetic information is incomplete and deeply misleading.

The third difficulty with Dawkins' position concerns the relation between genes and the phenotypes that they are supposed to create. Dawkins treats each gene as an independent actor that can 'produce a single phenotypic characteristic of selective importance';[218] 'single genes cause[ ] the different phenotypes that selection acts upon.'[219] The phenotypic characteristics thus represent, or act as proxies for, the genes. But in order for phenotypic traits to stand for individual genes, there must be 'a direct and relatively unmodifiable line between gene and adult phenotype.'[220] However, 'there is not and cannot be a simple one-to-one relationship between any given gene and the phenotypic expression at the level of the mature, fully developed organism.'[221] To the very slight extent that genes can be said to create organisms at all, they do not act individually. Genes work together to produce pheno-

[214] Goodwin (1994) 78-79.
[215] Ibid., 197.
[216] Griffiths and Gray (1994) 304.
[217] Ibid., 285.
[218] Sober (1984) 312.
[219] Ibid., 313.
[220] Rose (1997) 215.
[221] Ibid., 222.

typic effects. '[T]his is the phenomenon of *polygenic effects*, in which a given phenotype is the result of an interaction among an ensemble of genes.'[222]

As Stephen Jay Gould explains, '[b]odies cannot be atomized into parts, each constructed by an individual gene. Hundreds of genes contribute to the building of most body parts,'[223] and even though geneticists often speak 'misleading[ly]'[224] of genes for particular traits, '[i]t is still a truism that there is no interesting sense in which a gene builds a trait.'[225] Dawkins professes to agree: 'I am of course not suggesting that small genetic units work in isolation from each other . . . [Genes] have phenotypic effects on bodies, but they do not do so in isolation.'[226] He 'insist[s] on an atomistic view of units of selection, in the sense of units that actually survive or fail to survive, while being whole-heartedly interactionist when it comes to the development of the phenotypic *means* by which they survive.'[227]

But how can this be? If genes actually work together to construct the organism, how can they be selected individually? 'How can the single gene be the unit of selection, if genes build organisms only in elaborate collaboration with each other?'[228] On Dawkins' view, genes are selected indirectly by way of their phenotypic effects. Since genes are supposed to create the phenotype, '[s]election for or against the phenotype . . . impl[ies] selection for or against the gene[s]'[229] that created it. Therefore, if genes create as a team, then it would seem to follow that they should be selected as a team.

Dawkins tries to avoid this result by explaining in what sense an individual gene *can* be said to exert an effect upon the phenotype, without being the sole cause of any phenotypic characteristic. Dawkins suggests that while genes are not solitary causes, they are individual difference makers. All else being equal, the substitution of one allele for another on

[222] Sober (1984) 313.
[223] Gould (1977), cited in Midgley (1979) 454, and Dawkins (1982) 116.
[224] Rose (1997) 115.
[225] Sterelny and Griffiths (1999) 87.
[226] Dawkins (1982) 113.
[227] Ibid., 113-14.
[228] Sober (1984) 231.
[229] Ibid., 312.

a chromosome can make a difference in the phenotype, even though the allele itself does not cause the trait; and because of this difference-making ability, the gene can be selected, and become fixed in the population. Thus, 'Dawkins needs to show that we can sensibly speak of alleles having . . . effects, effects in virtue of which they are selected for or selected against.'[230] He tries to do this by emphasizing the 'meaningfulness of talk of "genes for" indefinitely complex morphological and behavioral traits.'[231]

For example, the difference between a radio that produces sound and one that does not may be due to something simple, such as whether or not a single wire is connected. Thus, according to Dawkins' way of speaking, we may call this wire a component 'for' sound, because its connection makes a difference to the production of sound. But the connected wire does not, of course, single-handedly cause the sound. Clearly, the sound is produced by the interaction of the set of components, including the connected wire, that constitute a working radio.

This way of speaking, which is apparently 'faithful to the practice of classical geneticists,'[232] derives from the geneticists' discovery that 'a change in one gene can make a big difference to the shape of an organism, or indeed to any other inherited property,'[233] even though the gene in question does not single-handedly cause the phenotypic property. It is common practice, for instance, to speak of a gene for blue or brown eyes. This does not mean, of course, 'that a pair of appropriately chosen stretches of DNA, cultured in splendid isolation, would produce a detached eye of the pertinent color. Rather, the intent is to indicate the effect that certain changes at a locus would make against the background of the rest of the genome.'[234] Thus, to speak of a gene for eye color means that 'there is a [genetic] difference in the biochemical pathways that lead to brown and to blue eyes,'[235] and this genetic

[230] Sterelny and Kitcher (1988) 348.
[231] Ibid.
[232] Ibid.
[233] Goodwin (1994) 15.
[234] Sterelny and Kitcher (1988) 348.
[235] Rose (1997) 115.

difference at a single locus we refer to as a gene for the eye color; '[a] certain allele in humans is an "allele for brown eyes" because, in standard environments, having that allele rather than alternatives typically available in the population means that your eyes will be brown rather than blue.'[236] This is true even though the allele does not single-handedly create the eye color, because the effects of a gene depend upon the environment in which it finds itself. '[G]enes are *context-sensitive* difference makers.'[237] But how exactly does this way of speaking of genes for some trait, this 'convenient shorthand,'[238] support the position that genes create organisms for their own benefit?

Dawkins' argument seems to run as follows. Everything else being the same, changing a gene at a particular locus has an effect upon the phenotype; therefore, the gene must be causing that effect; and if the gene causes a phenotypic effect, then natural selection can judge the gene by that effect. However, we saw above that Dawkins concedes that individual genes do not make body parts by themselves: '[t]he manufacture of a body is a cooperative venture of such intricacy that it is almost impossible to disentangle the contribution of one gene from that of another.'[239] If it is true that genes must work together to produce phenotypic traits, then a gene 'for' some trait, such as brown eyes, is actually only *one* of the causes of that trait. And if the gene is only one of the causes of the trait, then we seem to reach the conclusion that natural selection can not judge the gene separately from the other genes that cause the trait.

Dawkins tries to have it both ways. On the one hand, he admits that genes do not cause traits by themselves; but on the other hand, he tries, by means of the concept of genes 'for' some phenotypic trait, to associate single genes with particular phenotypic traits, as a way of giving selection access to individual genes. As Gould explains, '[i]f bodies were unambiguous maps of their genes, [with parts of the body standing for individual genes,] then

[236] Sterelny and Griffiths (1999) 87.
[237] Ibid., 87-88.
[238] Rose (1997) 115.
[239] Dawkins (1989) 24.

battling bits of DNA would display their colors externally and selection might act upon them directly.'[240] But there is no one-to-one correlation between parts of the body and individual genes; particular phenotypic traits do not stand for individual genes. 'The fact that a genotypic change will probably have an effect on the phenotype does not mean that a particular trait is reducible to a particular gene.'[241] And if the fitness of a phenotypic trait is not reducible to a gene, then the gene can not stand for the trait.

Let us consider Dawkins' argument further. When Dawkins says that the gene is selected because of its effect on the phenotype, what we want to know is why the gene is selected. Here it will be helpful to mention Sober's distinction between selection *of* and selection *for* some trait.[242] What Sober means is that an object with a certain property can be selected even though the property was not the cause of the selection. There is selection 'of' the property but not selection 'for' it. As Sober explains, '"[s]election of" pertains to the *effects* of a selection process, whereas "selection for" describes its *causes*.'[243] For example, Sober describes a children's sorting toy[244] that sorts marbles of four different sizes, the smallest marbles descending to the lowest level of the toy, the next smallest marbles to the next lowest level, and so forth. Each of the four sizes has a different color, the smallest marbles, let us suppose, being green. Consequently, at the lowest level, the smallest marbles are the objects selected, but 'it is equally true that the green balls are the objects that are selected.'[245] However, since the size of the marbles and not their color is what matters for the selection, there is selection 'of' green marbles, but there is no selection 'for' greenness. As Sober explains,

> To say that there is selection for a given property means that having that property *causes* success in survival and reproduction. But to say that a given sort of object was

[240] Gould (1977), cited in Dawkins (1982) 116.
[241] Collier (1981) 342.
[242] Sober (1984) 97-102.
[243] Ibid., 100.
[244] Ibid., 99.
[245] Ibid., 100.

selected is merely to say that the result of the selection process was to increase the representation of that kind of object.[246]

In Sober's terms, Dawkins' position amounts to the claim that there is selection 'for' individual genes that make a difference in the phenotype, because they cause success in survival and reproduction. But there are reasons for resisting Dawkins' claim, and for concluding that although there is selection 'of' the individual gene, there is not selection 'for' it, because the individual gene is not in fact the cause of the fitness of the organism.

We begin by noting that 'for a given gene at a particular locus . . . [to be] a positive causal factor in survival and reproduction' means that 'the allele must not decrease fitness in any (developmental) context, and must raise it in at least one.'[247] But since the effect of a gene upon the phenotype depends upon the developmental environment in which it finds itself—the other genes in the genome and the non-genetic environment—the gene can not be said to have a univocal causal rôle. Given a particular background, a gene may make either a harmful or beneficial difference in the organism. To take a simple instance, the case of heterozygote superiority[248] demonstrates that the genetic environment determines whether a given allele will be harmful or beneficial to the organism. The gene, so to speak, is not anything in itself; it does not have any causal power independent of the collection of developmental resources that constitute the organism. The gene possesses the potential to do much good and much ill, depending, so to speak, upon the company it keeps.

In some ways, a gene is like a letter of the alphabet, and the fitness of an organism like the meaning of a word or sentence. Individual letters are generally thought not to have any meaning of their own. The words or sentences, in which meaning resides, are, of course, composed of letters, and a substitution of one letter for another in a word almost al-

---

[246] Ibid.
[247] Ibid., 302.
[248] Heterozygote superiority means that the heterozygote (i.e., an organism having different alleles at a given locus on homologous chromosomes) is superior in fitness to the homozygote (i.e., an organism having, at a given locus, two alleles that are the same). Thus, 'the *Aa* heterozygote is fitter than both the *AA* and the *aa* homozygotes. This means that one dose of the *a* allele is better than two, and is also better than zero; the same holds for the *A* allele.' Ibid., 302-03.

ways results in a change in the meaning, but the meaning attaches to the world as a whole, or the *collection* of letters, and not to the individual letters themselves. In the same way, fitness attaches to the organism, or the collection of developmental resources, and the individual genes do not have any fitness or causal power apart from the collection in which they function. If the collection of developmental resources is fit, then it will survive and reproduce, and the gene, as a member of the collection, may survive in turn. So the fact that the gene makes a difference in the phenotype does not show that there is selection 'for' the gene; rather, there is selection for the fit combination of developmental resources, and selection of the individual components of that collection. What is evaluated by natural selection is the collection of developmental resources; the genes derive their properties from the collection, or from their participation in it. Consequently, when we refer to a gene 'for' some trait, what we mean is that certain combinations of genetic and non-genetic factors, which include this particular gene, give rise to that trait.

As I see it, Dawkins is mistaken in his suggestion that because a single gene can make a difference in the phenotype there must be selection for the gene. There is selection 'for' the collection of developmental resources that give rise to useful traits and fit organisms, and selection 'of' the individual genes in that collection. The fact that a difference-making gene survives and is replicated, does not prove, as Dawkins seems to believe, that there was selection for the gene. The gene is not the cause of fitness, even if the gene makes a difference in fitness; rather, fitness, which selection evaluates, is caused by having the right collection of genes—and everything else--which includes the gene in question.

The fourth problem with the view that genes create organisms is that it treats genes as *the* source of information in the construction of organisms. '[T]he variety of metaphors for DNA and its genetic functions have grown almost out of hand—it has been likened to a codebook, a blueprint, a recipe and a telephone directory, to name but four of the more prosaic comparisons.'[249] For Dawkins, '[t]he genetic code in not a blueprint for assembling a body from a set of bits; it is more like a recipe for baking one from a set of ingredients.'[250] 'It is a set of instructions which, if faithfully obeyed in the right order and under the right conditions, will result in a body.'[251] On this view, if the genome is compared to a recipe, then the other resources that contribute to the making of the organism, comprising those resources within and outside the cell, must be likened to ingredients. Thus, Dawkins 'presume[s] that the key to understanding development is to understand the interaction of two classes of developmental resources---genes and the rest.'[252] However, this 'dichotomous approach to development' is untenable, because, as we have suggested, '[t]he genes are just one resource that is available to the development process . . . The role of the genes is no more unique than the role of many other factors.'[253]

No one, certainly not Richard Dawkins, believes that genes create organisms on their own. As Dawkins puts it, a 'body . . . is the consequence of the obeying of a series of instructions . . . *in the right environment*.'[254] '[G]enes control embryonic development,'[255] but they can do so 'only . . . in some specified environment.'[256] This way of describing embryonic development suggests that genes are the true causal movers in development, and the non-genetic resources are merely necessary conditions for the genes' causal activity. Or, to put this another way, genes provide the essential information about how to construct an organism; the other resources apply this information in actually constructing it.

But '[i]n fact, there are . . . two ways to make sense of the notion of information in development. First, the entire set of developmental resources . . . may be said to contain

---

[249] Rose (1997) 33.
[250] Dawkins (1981) 567.
[251] Dawkins (1982) 175.
[252] Griffiths and Gray (1994) 277.
[253] Ibid.
[254] Dawkins (1982) 175, italics added.
[255] Dawkins (1989) 23.
[256] Dawkins (1982) 23.

information about evolved developmental outcomes.'[257] Molecules of DNA bear part of the information required to make a body, with the remainder located in the other developmental resources. Dawkins seems to acknowledge this: 'Genetic causes and environmental causes are in principle no different from each other.'[258] The second way to understand developmental information 'is to embed the information in one resource by holding the state of the other resources fixed as [background] conditions.'[259] In this way, any of the resources may be arbitrarily chosen as the base of the information, which creates an organism in the presence of the other resources. Accordingly, we may say, as Dawkins does, that genes create organisms if they are in the right environment; or, 'we can talk with equal legitimacy of cytoplasmic or landscape features coding for traits in standard genic backgrounds.'[260] 'It makes no more and no less sense to say that the other resources "read off" what is "written" in the genes than that the genes read off what is written in the other resources.'[261]

Dawkins is clearly disposed to assign the leading causal rôle in development to the genes, which constitute for him 'a set of plans for building a body.'[262] But what reason does he give for this assignment? Why does he privilege the genes over the other resources? For Dawkins, '[t]he special status of genetic factors rather than non-genetic factors is deserved for one reason only: genetic factors replicate themselves, blemishes and all, but non-genetic factors do not.'[263] Considering the difficulties that we have identified with Dawkins' view, there are two obvious replies to this distinction between the genetic and non-genetic resources. In the first place, even if we assume that only genes replicate, this assumption provides no support for Dawkins' claim that genes create organisms. Whether or not genes replicate is entirely irrelevant to the issue of their rôle in development. As Dawkins concedes, 'when we are talking about *development* it is appropriate to emphasize non-genetic as well as

[257] Griffiths and Gray (1994) 282.
[258] Dawkins (1982) 13.
[259] Griffiths and Gray (1994) 282.
[260] Ibid., 283.
[261] Ibid., 284.
[262] Dawkins (1989) 23.
[263] Dawkins (1982) 99.

genetic factors. But when we are talking about units of selection . . . [the] emphasis [must be] on the properties of replicators.'[264] Therefore, since we are asking a question about development, about why we should privilege genes over other resources in the construction of organisms, we must conclude that '[n]o basis has been provided for privileging the genes over other developmental resources.'[265]

Our second response is to challenge Dawkins' assertion that only genes have the power to replicate, 'that is to say they make copies of themselves.'[266] I shall return to this in the next section, concerning the unit of selection, but as we explained above, '[a] segment of DNA isolated from the cytoplasmic machinery of ribosomes and proteins has no such power'[267] to replicate itself. Genes do not make copies of themselves. Rather, they are copied, along with the other cellular structures, as part of the process by which organisms reproduce themselves, 'making [new] wholes from parts'[268] of themselves. Given the nature of development, it is far more accurate to ascribe the power of replication to 'total developmental processes or life cycles.'[269] 'The life cycle includes genes, environmental influences, and the generative field in a single process that closes on itself and perpetuates its nature generation after generation.'[270] Thus, attempts to ascribe unique significance to genes fly in the face of the facts of development.

To conclude these criticisms of Dawkins' view that genes create organisms for their own benefit, let us simply note that the fundamental problem with attributing a unique creative rôle to genes is that they are but one of the causal components of the developing organism. Some of the developmental resources, including the complement of genes, are provided by the parents (or parent, in the case of asexually reproducing organisms); others are derived from the environment. But all of these resources make ineliminable contributions to

[264] Ibid., 98.
[265] Griffiths and Gray (1994) 283.
[266] Dawkins (1989) 23.
[267] Griffiths and Gray (1994) 299.
[268] Goodwin (1994) 176.
[269] Griffiths and Gray (1994) 278.
[270] Goodwin (1994) 176-177.

the construction of the organism, such that we can not single out one of them as being *the* cause of the organism. An organism is judged by natural selection according to its phenotypic characteristics, and these characteristics being the products of a developmental system, of which genes are but one component, it is really the complete developmental system that is being judged by selection.

C. Genes Are Not the Unit of Selection

As we have said, Dawkins' thesis that evolution is best represented as a competition among selfish genes depends upon two sub-theses: that genes create organisms, and that genes are the unit of selection. Having completed our discussion of the first of these sub-theses, in this final section we shall develop and defend a handful of criticisms of the second sub-thesis.

The controversy over which entity should be regarded as the unit of selection involves a great many issues and arguments. Some of these have been touched upon in our discussion of the relation between genes and organisms, but I want to explain in more detail here a few of the reasons why it is erroneous and misleading to identify, as Dawkins does, the individual gene as the target of selection. Accordingly, I shall discuss Dawkins' position in terms of two main issues: the unity of the organism and the genotype, and Dawkins' distinction between replicators and vehicles.

Our first concern is whether Dawkins is correct to treat individual genes as independent actors. As Mayr explains, proponents of the genocentric view have 'a reductionist-atomistic conception of the genotype,' which treats 'the genotype [as] an aggregate of separate, independent genes.'[271] This view is frequently 'derided as "beanbag" genetics precisely because it depend[s] . . . on the assumption that each gene [i]s an isolated unit which [can] be shaken, shuffled and selected like one bean in a beanbag independently of all others.'[272]

[271] Mayr (1988) 101.
[272] Rose (1997) 216.

In Dawkins' view, each of these independent genes has an individual effect upon the phenotype, by means of which natural selection can judge the fitness of the gene.

The alternative 'organocentric' position, on the other hand, might be understood in terms of its own two theses. First, 'an organism is an integrated whole and not a collection of independent genes.'[273] And second, 'natural selection does not operate with separate "traits".'[274] As Mayr explains, the genocentric position involves an 'inappropriate . . . understanding of natural selection,' because it would 'give[ ] every gene a separate fitness value,' and 'dissect[ ] the phenotype into the greatest number of separate features,'[275] with natural selection supposedly acting on each part independently. Or as Gould puts it, the problem with Dawkins' view is that 'parts [of organisms] are not translated genes, and selection doesn't even work directly on parts. It accepts or rejects entire organisms.'[276]

How does Dawkins defend his claim that genes are selected individually, and not collectively? As we have seen, he offers three reasons. First, he argues that because individual genes can make a *difference* in the phenotype, selection can judge that difference; selection can act on a gene 'for' some trait, even if the gene does not create the trait on its own. Having already rejected this argument, I shall have nothing more to say about it here. Second, he argues that individual genes but not whole genotypes can be replicators,[277] and that only replicators have a legitimate claim to be the unit of selection. I shall address this argument later in the chapter. And third, he offers his rowing analogy in order to demonstrate that what appears to be selection of a team is actually selection of the individuals that constitute the team. It is this argument that I want to consider now.

The problem for Dawkins, stated in his vocabulary, is that 'replicators (genes) combine to build vehicles (organisms) and the effect of a gene is critically dependent on the

---

[273] Mayr (1988) 139.
[274] Ibid., 152, citing T. Dobzhansky, 1956, 'What is an Adaptive Trait?' *Amer. Nat.*, 90: 337-347, 340.
[275] Ibid., 139.
[276] Gould (1977), cited in Midgley (1979) 454.
[277] To be more precise, Dawkins claims only that the genotypes of sexually reproducing organisms are not replicators; asexual genotypes may be replicators.

company it keeps.'[278] Genes work together to build organisms, so it would appear that it is the collection of genes that creates the organism, and consequently the collection of genes that is favored or discriminated against by natural selection. But Dawkins wants to persuade us that this appearance is false. Like rowers competing for seats in the boat, genes compete against their alleles to occupy the loci on chromosomes. Genes are selected individually, but they 'compromis[e] for the moment in co-operation with other genic cohabitants of the organism,[279] because coöperation is in their self-interest. Thus, the rowing 'coach may scrutinize the relative times of different teams but the competition can be analyzed by investigating the contributions of individual rowers in different contexts.'[280]

In the rowing analogy, it may be remembered, Dawkins asks us to imagine a rowing coach who is trying to assemble his best crew from the available candidates. He chooses the crew by randomly assigning the rowers to different teams every day, and racing them against one another, recording the results of each trial. After a large number of such races, the coach can identify the best rowers for his team, because the winning boats tend to contain the same individuals. Of course, sometimes a poor rower will be found in the winning boat, and an excellent rower in a losing boat, 'either because of the inferiority of the other members, or because of bad luck,'[281] '[b]ut by definition luck, good and bad, strikes at random,'[282] and the better oarsmen will tend, in the long run, to have more success than the poorer rowers. Thus, Dawkins concludes, what appears to be an instance of team selection is properly understood as an instance of individual selection, even though the individuals can succeed only when they work as a team. 'One of the qualities of a good oarsman is teamwork, the ability to fit in and cooperate with the rest of the crew. This may be just as

[278] Sterelny and Kitcher (1988) 341.
[279] Reid (1985) 223.
[280] Sterelny and Kitcher (1988) 341.
[281] Dawkins (1989) 38.
[282] Ibid., 39.

important as strong muscles.'[283] The coach selects coöperative individuals for his team, not a team as a whole.

Dawkins' argument for 'the thesis that all selection is genic selection'[284] depends upon a showing that the coach is indeed selecting individuals, and that nature selects individual genes in the same way that the coach selects individual rowers. But is the rowing analogy in fact a good analogy for the operation of natural selection? I want to suggest that it is not a convincing analogy. Sober calls it 'deeply misleading.'[285] The most significant problem with the analogy is that it grossly misrepresents the relation that exists between genes in an organism; it fails to recognize the extent of 'interaction among genes, the integration of the genotype.'[286] Dawkins' view, which treats 'each gene . . . as if it were quite independent of all others,' is contradicted by 'all sorts of phenomena' 'such as the linkage of genes, epistasis, pleiotropy, and polygeny.'[287]

The genocentric theorists are, of course, 'fully aware of pleiotropy, polygeny, and other processes that produce the phenotype . . . [but] they ignore these processes in their evolutionary interpretations.'[288] Let us consider one example of how Dawkins ignores genetic interaction in his rowing analogy. As everyone agrees, '[g]enes are capable of epistatic interactions.'[289] Loosely, this means that genes can affect the expression of other genes. One gene may completely mask or suppress the expression of another gene, or merely control the phenotypic effects of the other gene. For instance, a so-called 'recessive' gene will

---

[283] Ibid.
[284] Sterelny and Kitcher (1988) 341.
[285] Sober (1984) 307.
[286] Mayr (1988) 423.
[287] Ibid. Linkage is the tendency of different genes on the same chromosome to be inherited together, though linked genes can be separated by crossing over. Epistasis is a type of gene interaction in which a gene at one locus affects the expression of a gene at another locus; the suppression of one gene by another is an example of epistasis. Pleiotropy is the phenomenon by which one gene has a number of different phenotypic effects; e.g., a mutation in a single gene can lead to Marfan's syndrome, a condition characterized by the elongation of the long bones and defects in the heart and eye. And polygeny is the phenomena by which a set of genes collectively control the expression of a quantitative or continuously varying trait, such as height; each polygene exerts a cumulative effect on the trait.
[288] Mayr (1988) 449.
[289] Sober (1984) 307.

be expressed and affect the phenotype only if the genotype includes two copies of the recessive gene; however, if its allele is 'dominant,' then the recessive gene will be muted, and will have no effect upon the phenotype.[290] As a result of epistatic interactions, the phenotypic effect of a particular gene depends largely upon other genes in its genetic environment.

But Dawkins' rowing analogy fails to take account of the interaction among genes, and once we adjust the model to include interactions among the rowers, the claim that the coach is selecting only individuals becomes implausible. As Dawkins explains epistatic interactions, '[t]echnically the interactions are non-additive which means, roughly, that the combined effects of the two genes is not the same as the sum of their separate effects.'[291] Now, in order to accommodate this concept of non-additive effects into the rowing analogy, let us assume that among the candidates competing in the racing trials, there are two rowers who are no more likely to be in a winning boat than in a losing boat, *unless* they are paired together, in which case they both become, under their mutual motivation, above-average rowers, and tend to be in the winning boat. Using Dawkins' terms, we may say that the combined efforts of the two rowers is greater than the sum of their separate efforts. Accordingly, neither of them will be selected alone, since separately they are merely average rowers, but they may be selected as a pair, because, in the presence of the other, each becomes a good rower. The coach may select them together, but not individually; it is both or neither. Therefore, when we incorporate epistatic relations into the rowing analogy, 'we can see how the coach may well be selecting for combinations of rowers, not for single rowers,'[292] as Dawkins would have us believe.

We can make the same point in another way. In Dawkins' view, the gene and the rower are selected for their causal powers. A gene is selected because of its causal effects

[290] Strictly speaking, the suppression of the recessive allele by the dominant allele is not an example of epistasis, which refers to the control or suppression of one gene by another gene at a *different* locus; but it is a familiar way of making the point about the phenotypic effect of a gene being dependent upon other genes in the genotype.
[291] Dawkins (1982) 286, glossary. Note that epistatic interactions may involve more than two genes.
[292] Sober (1984) 307.

upon the phenotype, a rower because of his causal effects upon the speed of the boat. If we suppose, with Dawkins, that individual rowers are independent of one another in their effects upon the speed of the boat, that is to say, if there are no 'epistatic' interactions among the rowers, then each rower may be judged separately from the others; the coach may simply follow Dawkins' proposal and choose the rowers who tend to finish in the winning boats. But if the actual rowing ability of an oarsman is dependent upon the causal influences of other rowers in the boat, then the rower can not be judged apart from those influences, inasmuch as he may be a good rower with one set of teammates, and a bad rower with another set. His very skill as a rower will then depend on his interactions with the other rowers in the boat.

In the rowing analogy, Dawkins seems to suppose that the skill of each rower is invariant with respect to the other occupants of the boat,[293] for he writes that '[r]owing the boat is a cooperative venture, but some men are nevertheless better at it than others.'[294] This suggests that the rower, so to speak, carries his skill with him from race to race, his rowing ability remaining unaffected by his interactions with his teammates. But as an analogy for the interactions of genes in the genotype, Dawkins' independent rowers will not do, because genes affect the phenotypic effects of one another. As Mayr points out, [g]ene exchange at any locus may have an impact on the selective value of genes at other loci.'[295] Consequently, we must incorporate into the analogy epistatic interactions among the rowers, and once we do so it becomes clear why the coach is not selecting only individuals. If the ability of a rower to affect the speed of the boat is epistatically determined by his teammates, then the rower can not even be said to have a context-free rowing ability; whether he helps or hurts his team in a race depends on his interactions with the other members of the team.

---

[293] In point of fact, of course, whether or not a person rows well *is* largely independent of the influences of his teammates; his skill as a rower is not determined by the occupants of his boat. But in order to capture the epistatic interactions of genes, we must assume that the ability of a rower is determined by his crewmates.
[294] Dawkins (1989) 38.
[295] Mayr (1988) 449.

Therefore, what determines the speed of the boat is not the individual rowers, who have no effect apart from the team dynamic, but the *group* of interacting rowers.

It is these epistatic interactions among genes that cause Mayr to insist that 'genes perform as teams,'[296] and Sober to conclude 'that ensembles [of genes] may have determinate causal roles in selection processes, even when single genes do not.'[297] And if it is genetic ensembles that have effects upon the phenotype, in the same way that groups of interacting rowers affect the speed of the boat, then it is the ensembles, and not the individuals, that are selected.

With that we come at length to the last argument of this last chapter, the argument that the gene, and not the individual organism, is the unit of selection, because the gene is a replicator. Dawkins' argument is based on a particular understanding of 'what characteristics an entity must have to be a unit of natural selection.'[298] And these characteristics, Dawkins argues, are possessed by genes, but not by individuals. Therefore, the gene is the entity for whose benefit adaptations exist; it is the entity that survives, or fails to survive, in the form of copies; it is the entity that is ultimately judged by natural selection. Organisms, on the other hand, are 'nothing but the vehicles for genes,'[299] 'just packaging for the hereditary essence,'[300] a means 'for the transmission of [genetic] instructions to the next generation.'[301]

Dawkins' argument is rather straightforward. Evolution by natural selection is a cumulative 'process that typically spans many generations.'[302] 'For a selection process to span many generations, the variants being selected must exist over that length of time,'[303] and in order for the process to be cumulative, the enduring entity 'must be capable of pass-

---

[296] Ibid.
[297] Sober (1984) 314.
[298] Collier (1981) 339.
[299] Goodwin (1994) 28.
[300] Ibid., 33.
[301] Ibid., 27.
[302] Sober (1984) 250.
[303] Ibid., 251.

ing the changes in state it acquires along to descendants.'[304] Thus, the unit of selection must have sufficient 'permanence and the capacity to transmit acquired changes to descendants.'[305] Individual genes have these qualities, Dawkins argues, because genes are capable of replication during reproduction, and any mutations that do occur in the genes are transmitted when they are copied. But organisms do not have either of these characteristics: an organism is not copied during reproduction, and, *pace* Lamarck, it can not inherit characteristics acquired by its parents during their lifetimes. Consequently, the single gene must be identified as the unit of selection.

There are a number of problems with this argument that could be mentioned, but I will examine only a few of them here. The first problem is that 'it is unclear how Dawkins' argument against the claim that organisms are replicators is to proceed.'[306] Consider the two arguments Dawkins offers to show that organisms are not replicators. He argues, first of all, that organisms do not survive from generation to generation; 'it is not the bodies that survive; they reproduce their genes and die.'[307] But if we regard an organism as a developmental system, in which 'genes are just one resource that is available to the developmental process,'[308] then an organism can be a replicator, because 'the developmental system consists of the resources that produce the developmental outcomes that are stably replicated in the lineage.'[309] As we have said, reproduction is a process by which a part of an organism becomes a whole organism; the offspring inherits much more from the parents than just genes, including the machinery, resources, and organization of the single cell from which the new organism develops. Thus, the life cycle/organism reproduces itself, using resources or information derived from the environment and from the parents, and by means of this process

[304] Ibid., 253.
[305] Ibid., 254.
[306] Ibid., 253.
[307] Dawkins (1981) 570.
[308] Griffiths and Gray (1994) 277.
[309] Ibid., 278.

of reproduction, the life cycle can span many generations, satisfying one of Dawkins' requirements for the unit of selection.

His second argument for the claim that organisms are not replicators is the familiar charge against Lamarckism, namely, that organisms can not pass on acquired characteristics, as genes can transmit their mutations. For instance, a spider that loses one of its leg to a predator does not produce offspring that are similarly deformed, but a genetic mutation that causes a spider to be born without a leg may be passed on to the offspring. However, if genes are part of an organism, that is, one of the developmental resources that give rise to the organism, then a genetic mutation *is* a heritable change in the organism.[310] As Sober asks, '[w]hy shouldn't a [genetic] mutation count as something that happens to an organism as well as to the gamete in which it occurs? If this cannot be prohibited . . . then organisms *can* count as replicators in Dawkins' sense.'[311] Moreover, as numerous writers have pointed out, 'it isn't genes themselves which survive, but gene types.'[312] But an organism likewise reproduces its type; a duck gives birth to a duck, and a duck-billed platypus has, well, it has a little egg-laying mammal with a mélange of body parts. When organisms reproduce, 'it is the same *kind* of organism which is being reproduced,' the same 'pheno*type* which is being replicated.'[313] In brief, Dawkins has not explained why an organism, properly understood as a life cycle that grows and develops from a single cell, can not be treated as a replicator. For understood in this way, an organism satisfies Dawkins' requirements for the unit of selection: it has longevity (a lineage of organisms can survive for generations), fecundity (organisms reproduce themselves and pass on their traits to their descendants), and fidelity (the offspring resemble the parents).

---

[310] I do not consider here the uncontroversial cases in which changes to the phenotype are actually inherited by the offspring, as when experimental, non-genetic changes are made to the body of a *paramecium* and the changes are inherited by the progeny. In Goodwin's words, 'here is a case of a mutation, produced by an operation, that is inherited via body structure, not via the genes.' Goodwin (1994) 14. For a well-argued defense of the importance of neo-Lamarckian mechanisms in evolution, see Eva Jablonka and Marion J. Lamb, 1995, *Epigenetic Inheritance and Evolution: The Lamarckian Dimension* (New York: Oxford University Press).
[311] Sober (1984) 253-54.
[312] Collier (1981) 344.
[313] Ibid., italics in the original.

A second problem with Dawkins' claim that the gene, and not the organism, is a replicator, is something we mentioned earlier in rejecting Dawkins' other claim that genes create organisms. A gene does not have the power to replicate itself. It depends upon 'the organization of the cell . . . for its replication.'[314] What truly can be said to replicate itself is the developmental system or life cycle, which 'only replicates itself because of the presence of all the developmental resources.'[315] The genes are one of the resources that contribute to the development and reproduction of the life cycle. Thus, Goodwin can correctly write that '[t]he capacity to reproduce is a property of the whole *organism*, not a special replicating part that is distinct from the rest of the reproducing body.'[316]

Finally, the third criticism of Dawkins' view is a subtle one. I have argued that the organism, properly understood as a life cycle or developmental system, can in fact be a replicator, which satisfies Dawkins' requirements for the unit of selection. But I want to suggest now that these requirements carelessly misrepresent what it means for an entity to be a target of selection. Dawkins argues that an entity must have sufficient longevity to qualify as a target of selection, because evolution by natural selection is a historical process that occurs over long periods of time. Darwin described evolution as a process of descent with modification from ancestral forms. We can trace this descent with modification through the phylogenetic lineage of a species. But there can be no descent from ancestral forms, no evolution, unless something survives from generation to generation, some entity that can accumulate changes and pass them on. Therefore, the unit of selection, the entity that is preserved because of its adaptedness to the particular environment in which it finds itself, must be such a long-lived entity, which can transmit acquired changes to its descendants.

I want to suggest that this argument is mistaken, because it subtly confounds natural selection with evolution. It is true, of course, that we can not have a phylogenetic lineage without generational continuity; each generation must give rise to the next generation.

---

314 Goodwin (1994) 38.
315 Griffiths and Gray (1994) 299.
316 Goodwin (1994) 36.

Thus, stability of type or kind is required for evolution, which can not occur without heredity. But natural selection is merely the fact that some individuals, because they have more of what it takes to survive than other individuals, tend to make a better living than their more poorly adapted conspecifics. Natural selection judges variations; as Darwin puts it, '[i]t may metaphorically be said that natural selection is daily and hourly scrutinizing, throughout the world, every variation, even the slightest; rejecting that which is bad, preserving and adding up all that is good.'[317] Thus, '[v]ariation, rather than permanence, is the principal prerequisite for selection.'[318] Or consider this passage from Darwin, which shows the relation between selection and inheritance:

> [From the] recurring struggle for existence, it follows that any being, if it vary however slightly in any manner profitable to itself, under the complex and sometimes varying conditions of life, will have a better chance of surviving, and thus be *naturally selected*. From the strong principle of inheritance, any selected variation will tend to propagate its new and modified form.[319]

Here we see that 'natural selection' plus 'the strong principle of inheritance' will tend to produce evolution, a 'new and modified form.' But Darwin does not conflate natural selection and inheritance. He does not claim that selection itself, which is the preservation of favorable variations, requires the transmission of something enduring across generations. Rather, selection occurs, in a sense, *prior to* the transmission of favorable traits to the offspring. So when Dawkins says that something must endure from generation to generation in order to qualify as the unit of selection, he is mistaking a requirement of evolution (i.e., favorable variations must be inherited by the offspring) for a requirement of natural selection. This is why Mayr 'rightly rejects the claim of the genic selectionists that "the unit of selection must have a high degree of permanence".'[320] As I should put it, permanence or stability of type is a requirement of evolution, not of selection, which is merely a filter through which each generation of an evolving lineage must pass.

---

[317] Darwin (1998) 70.
[318] Mayr (1988) 124.
[319] Darwin (1998) 6.
[320] Mayr (1988) 123-24.

With that I conclude this discussion of Dawkins' selfish gene vision of nature. I think that it is a distorted vision. I think that there is more truth (and grandeur) in Darwin's 'view of life,'[321] which sees organisms struggling to survive, than in Dawkins' hyper-exaggerated emphasis on one type of macro-molecule in those organisms.

[321] Darwin (1998) 396.

CHAPTER 5

CONCLUSION

In this concluding final chapter, I propose, in a very few pages, to recapitulate some of the findings of the previous chapters, and to give some indication of their philosophical significance. I began this discussion of reductionism, back in the second chapter, by referencing the question that lies at the heart of metaphysics: What is there? I want to return to that question now, and suggest that the debate over reductive materialism is really a debate about what exists, about what should be included in an ontological inventory of real things.

It is remarkable, when one stops to consider it, how many issues in philosophy turn on the existence or non-existence of some proposed entity. In the philosophy of religion, the disputed entities include God and souls; in the philosophy of mind, they are minds and consciousness; in ethics and aesthetics (axiology), we debate the existence of objective values; in the philosophy of science, one of the fundamental disputes concerns the existence of unobservable objects, such as sub-atomic particles; and numerous other philosophical issues revolve round the existence and status of such things as universals, numbers, sets, propositions, species, fictional characters, historical persons, sense data, and possible worlds.

Ontological reduction, as we have seen throughout these pages, is a relation between upper-level objects and their lower-level constituents. In this relation, the upper-level object being nothing more than a collection of constituent parts, the object is eliminated as a real thing in its own right; it is reduced, we might say, out of existence. Consequently, in the debate over reductive materialism, the central question is the nature of the objects of experience: animals, plants, cells, genes, stars, machines, eyeglasses, books, and so forth. Are these real things? The reductive materialist, in opposition to the anti-reductionist, and in defiance of common sense, answers in the negative: they are not real things. They are only collections of real particles to which, for convenience' sake, we assign names. Only the fundamental particles themselves possess ontological reality.

The chief dissatisfaction with reductive materialism, which many people apparently feel, is that it denies the reality of upper-level objects. Such objects, which we can observe and touch, and which include ourselves, would seem, to common sense, to have at least as high a claim to reality as the theoretical, unobservable entities of particle physics. Thus, all the basic criticisms of reductive materialism, including those that we considered in chapters two and three, are really reactions to this denial of upper-level objects. The anti-reductionists, whether they realize it or not, are making metaphysical claims about what exists.

But while the anti-reductionist wants to affirm the reality of upper-level objects, he faces a considerable difficulty, of which many philosophers and scientists appear to be unaware. The problem is that denying reductive materialism requires that the anti-reductionist supply a metaphysical system of his own in which upper-level objects can find their reality. It is not sufficient merely to affirm the existence of both macro-objects and their micro-components. In addition, the anti-reductionist must supply a metaphysical account of these macro-objects and of the interaction between the upper and lower levels. If a macro-object, such as a bacterium or a fountain pen, is not just a collection of interacting particles, then what more is it? What are our criteria for distinguishing collections of particles that constitute real things from collections of particles that do not? Is there a point at which a group of atoms cross the ontological threshold and become a real thing independent of the atoms? How can we identify such a point? As Marjorie Grene wisely recognizes, the chief problem with anti-reductionism is that it 'breaks through the defenses of a simple, one-level physicalism without providing an alternative metaphysic to take its place.'[1] Consider a few examples of how this problem manifests itself in the accounts of anti-reductionism that we have considered.

[1] Grene (1971) 21.

Biologist Ernst Mayr argues that '[a]ttempts to "reduce" biological systems to the level of simple physico-chemical processes have failed because during the reduction the systems lost their specifically biological properties.'[2] He explains that although '[a]ll processes in organisms . . . strictly obey . . . physical laws, . . . organisms differ from inanimate matter . . . in the organization of their systems and especially in the possession of coded information.'[3] Perhaps Mayr is making only an uncontroversial epistemological point about our current inability to represent all biological phenomena in terms of physics. This is how Alexander Rosenberg understands Mayr's anti-reductionist stance. 'In the case of reductionists,' writes Rosenberg, 'the commitments are metaphysical; in the case of their opponents, the presuppositions seem to be epistemological: Mayr and the other antireductionists do not deny that life is but matter in motion.'[4] If this is all that Mayr means when he writes, for example, that 'the evidence in support of the autonomy of biology has grown exponentially in recent years,'[5] then he is not in fact an opponent of ontological reductionist at all. Biology really is nothing more than physics; the problem is simply that at present we lack the knowledge to translate biology into physics. It may be, however, that Mayr's anti-reductionism is not purely epistemological. When Mayr says that reduction fails because it leaves out the 'specifically biological properties,' he seems to mean that there is something real and ineliminable about such biological properties, because if there were nothing real about them, then it would not be a criticism of reductionism that it leaves them out. But if this is the proper understanding of his position, as I think it is, then Mayr needs a theory of reality that will accommodate the reality of both biological objects and their physical components. I do not find either that he has attempted to provide such a metaphysics, or even that he is sensible of the need to do so.

---

[2] Mayr (1988) 1.
[3] Ibid., 2.
[4] Rosenberg (1986) 88.
[5] Mayr (1988) 14.

Second, as we saw in chapter three, Michael Polanyi argues that upper-level objects can not be explained by physical laws, because those laws do not specify the particular form or structure of such objects. As Polanyi explains, an upper-level object 'relies for its operations on the laws governing the elements of the lower [level] . . . but these operations of it are not explicable by the laws of the lower level,'[6] since a vast multitude of upper-level objects are compatible with the same lower-level laws. For instance, 'a complete physical and chemical topography of an object would not tell us whether it is a machine, and if so, how it works, and for what purpose.'[7] Polanyi's position depends upon his belief that the structure or form of an object gives it a reality independent of the particles of which it is composed. This is an ontological claim about the objects of experience, but Polanyi does not develop a metaphysical system in which both physical particles and the objects made up of them can find their place. Polanyi needs to provide some account of what makes a real thing, because not all of the things to which we give names, such as traffic and constellations, are real things independent of their components. Therefore, how can we decide whether an upper-level object is a real thing in its own right, or simply a collection of more basic constituents that, out of convenience, we label a object? Polanyi does not tell us.

Third, Fodor's rejection of reductionism depends upon a similar view that upper-level objects are, in some sense, real entities. Fodor argues that it is impossible to establish the bridge laws required by reductionism between upper-level objects and their lower-level components, because upper-level objects are multiply realizable at the lower level. We can not translate the terms in the special sciences, such as genes, minds, or ecosystems, into the language of physics, because the vocabulary of physics does not contain equivalents for the upper-level terms. The only way this translation can be accomplished is by means of disjunctive sets. The lower-level equivalent of the upper-level object is the set of all the physical bases necessary to realize that upper-level object. However, Fodor objects to these dis-

[6] Polanyi (1983) 34.
[7] Ibid., 39.

junctive sets because they are not natural kinds; a cobbled together set of physical bases is not scientifically respectable in the way that upper-level objects are. But in claiming that the upper-level objects recognized in the special sciences are natural kinds, Fodor seems to be imputing to these objects a sort of ontological reality that distinguishes them from the artificial disjunctive sets. For Fodor, as for Polanyi and Mayr, reduction fails because the upper-level objects are ontologically real, and thus can not be eliminated as nothing more than their components. And like Polanyi and Mayr, Fodor also fails to develop a metaphysical system that recognizes and renders respectable the reality of upper-level objects.

And finally, with the discussion of organismic reductionism in chapter four, we encounter another version of the same problem. As a reductionist, Richard Dawkins has a philosophical preference for the lower level over the higher. Evolution by natural selection, he believes, occurs, not at the level of the organism, but at the lower level of the genes, since genes are the only things that matter in describing the course and the goals of evolution. This position is a form of reductionism in that it regards the upper-level phenomena (i.e., the structure, function, and behavior of organisms) as nothing more than the macro-consequences of the lower-level competition among genes to replicate themselves. Well-adapted organisms are merely the by-products of this competition. It is genes that are significant for Dawkins, as it the fundamental particles that matter to the reductive materialist.

But Dawkins does not attempt to reduce the competition among genes to the interaction of fundamental particles; he does not seek to reduce all of evolutionary biology to physics. This failure to reduce the genes to their components[8] suggests that Dawkins invests the genes with a kind of irreducible reality, in the same way that Mayr invests organisms with ontological significance. Dawkins, however, does not seem to recognize that his genocentric position, part-way between organisms and fundamental particles, is as untenable as any

[8] I would suggest that Dawkins' intuitive sense of the unity and independence of genes partially explains his reluctance to reduce the competition among genes to a struggle for survival among DNA bases, the molecular components of genes. See Dawkins (1982) 90-91 for a discussion of why 'it is not helpful' 'to say that an adaptation is for the good of the nucleotide,' even though 'it may not be strictly wrong.'

other anti-reductionist position without a supporting metaphysics. What is it that makes genes deserving of ontological significance, whereas organisms are not similarly deserving? Dawkins does not address the ontological questions, and he presumably would deny having the knowledge and ability to do so.

But the questions must be answered by anyone who would abandon 'simple, one-level physicalism,' in favor of real upper-level objects. 'To think anti-reductively,' Grene reminds us, 'demands thinking in terms of hierarchical systems, of levels of reality and the like; but we don't know any longer how to think in that way—and to be told, even to *know*, that [reductive materialism] is absurd does not in itself allow us to embrace wholeheartedly what ought to be the more reasonable alternative.'[9]

Metaphysics is the attempt to frame a description of reality in terms of which all of experience can be comprised, and 'this experience must be taken in its whole range, and must not be arbitrarily limited to the data of perception which intelligence works up into science.'[10] But many people find it impossible to reconcile reductive materialism with many of the entities that seem to populate our world. For example, W. R. Sorley has argued that

> [t]he [objective] facts of morality as they appear in the world, and the ideas of good and evil found in man's consciousness, are among the data of experience. If we overlook them in constructing our theory of reality, we do so at the risk of leaving out something that is required for a view of the whole, and we shall probably find that our base is too narrow for the structure we build upon it.[11]

Even such a confirmed physicalist as J. J. C. Smart, feels compelled, as I mentioned in the second chapter, to graft onto his materialist worldview, 'though they are not material things,'[12] numbers, sets, assertions, Occam's razor, and Platonic Forms. And other philosophers and scientists, including Fodor and Mayr, who believe that everything real is made of

---

[9] Grene (1971) 21.
[10] Sorley (1918) 7. Sorley's book, on the relation between morality, metaphysics, and God, is a neglected gem. Note also James' view that an empirical theory of reality 'must neither admit into its constructions any element that is not directly experienced, nor exclude from them any element that is directly experienced.' James (1996) 42, 'World of Pure Experience.'
[11] Ibid., 1-2.
[12] Smart (1978) 382.

matter, nevertheless reject reductive materialism, because it denies a hierarchical view of reality that includes organisms and other macro-objects. There is in many quarters, I suggest, a discontent with reductive materialism, but what shall we put in its place? We have two choices, writes Grene. Either we 'develop a comprehensive metaphysic consonant with the anti-reductionist view,'[13] or we try to avoid 'the straitjacket of a universal physics [by means of] a different—and sounder—approach to science [that] may liberate us from this cramping restraint.'[14]

Each of these alternatives clearly involves a grandiose philosophical project. However, I did not undertake this work with the hope or expectation that I could produce, at the end, a theory of reality or a novel approach to science that would accommodate anti-reductionist insights. Rather, I regard this work as preliminary to any attempt to develop an anti-reductionist response to reductive materialism. My overarching goal has been to show that the debate over reductive materialism is essentially a metaphysical dispute about the existence of upper-level objects, and that it can be resolved only by means of philosophical reflection. A full-scale vindication of the anti-reductionist position would require three parts, of which this represents only the first. The second part would involve a detailed evaluation of reductive materialism, showing that it is in fact an untenable metaphysical position, because it does not encompass those entities that have as much claim to reality as the fundamental particles of physics; 'it is necessary,' writes Sorley, 'that the basis of our theory of reality should be as broad and complete as possible.'[15] And the third part would involve the construction of a theory of reality that would encompass such entities. But parts two and three are, of course, projects for another time.

[13] Grene (1971) 21.
[14] Ibid., 25.
[15] Sorley (1918) 1.

SELECT BIBLIOGRAPHY

Ager, Tryg A., Jerrold L. Aronson, and Robert Weingard, 1974, 'Are Bridge Laws Really Necessary?,' *Noûs*, 8, 119-134.

Alston, William P., 1991, *Perceiving God: The Epistemology of Religious Experience* (Ithaca, New York: Cornell University Press).

Angeles, Peter A., 1992, *The HarperCollins Dictionary of Philosophy*, 2nd ed. (New York: HarperCollins).

Armitage, Angus, 1947, *The World of Copernicus* (New York: Mentor Books).

Audi, Robert, 1988, *Belief, Justification, and Knowledge* (Belmont, California: Wadsworth Publishing Company).

Audi, Robert, 2003, *Epistemology: A Contemporary Introduction to the theory of Knowledge*, 2nd ed. (London: Routledge).

Balashov, Yuri, 1994, 'Duhem, Quine, and the Multiplicity of Scientific Tests,' *Philosophy of Science*, 61, 608-628.

Barbour, Ian G., 1997, *Religion and Science: Historical and Contemporary Issues* (San Francisco: Harper Collins).

Bartley, W. W., III, 1987, 'Philosophy of Biology versus Philosophy of Physics,' in *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*, eds. Gerard Radnitzky and W. W. Bartley, III (LaSalle, Illinois: Open Court).

Berlinski, David, 1996, 'The Deniable Darwin,' *Commentary*, June, 19-29.

Berlinski, David, et al., 1996, 'Denying Darwin,' *Commentary*, September, 4-39.

Bickle, John, 1992, 'Multiple Realizability and Psychophysical Reduction,' *Behavior and Philosophy*, 20, 47-58.

Bickle, John, 1995, 'Connectionism, Reduction, and Multiple Realizability,' *Behavior and Philosophy*, 23, 29-39.

BonJour, Laurence, 1985, *The Structure of Empirical Knowledge* (Cambridge, Massachusetts : Harvard University Press).

Boyd, Richard, 1995, 'Observations, Explanatory Power, and Simplicity: Toward a Non-Humean Account,' in *The Philosophy of Science*, eds. Richard Boyd, Philip Gasper, and J.D. Trout (Cambridge: MIT Press).

Bradie, Michael, 1994, 'Epistemology from an Evolutionary Point of View,' in *Conceptual Issues in Evolutionary Biology*, ed. Elliott Sober, 2nd edition (Cambridge: MIT Press).

Brittan, Gordon G., 1970, 'Explanation and Reduction,' *Journal of Philosophy*, 67, 446-457.

Brooks, D. H. M., 1994, 'How to Perform a Reduction,' *Philosophy and Phenomenological Research*, 54, 803-814.

Butterfield, Herbert, 1965, *The Origins of Modern Science 1300-1800*, revised ed. (New York: Free Press).

Carnap, Rudolf, 1995, *An Introduction to the Philosophy of Science*, ed. Martin Gardner (New York: Dover).

Cartwright, Nancy, 1979, 'Do Token-Token Identity Theories Show Why We Don't Need Reductionism?,' *Philosophical Studies*, 36, 85-90.

Cartwright, Nancy, 1989, *Nature's Capacities and Their Measurement* (Oxford: Clarendon Press).

Cartwright, Nancy, 1995, 'The Reality of Causes in a World of Instrumental Laws,' in *The Philosophy of Science*, eds. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, Massachusetts: MIT Press).

Cartwright, Nancy, 1999, *The Dappled World: A Study of the Boundaries of Science* (Cambridge: Cambridge University Press).

Caullery, Maurice, 1966, *A History of Biology*, trans. James Walling (New York: Walker and Company).

Causey, Robert L., 1972, 'Attribute-Identities in Microreductions,' *Journal of Philosophy*, 69, 407-422.

Chalmers, A. F., 1982, *What is this thing called Science?*, 2nd ed. (Indianapolis, Indiana: Hackett Publishing Co.).

Chalmers, A. F., 1999, *What is this thing called Science?*, 3rd ed. (Indianapolis, Indiana: Hackett Publishing Co.).

Clark, Stephen R. L., 1990, 'Limited Explanations,' in *Explanation and Its Limits*, ed. Dudley Knowles (Cambridge: Cambridge University Press).

Clark, Stephen R. L., 1993, 'Philosophers and Popular Cosmology,' *Journal of Applied Philosophy*, 10, 115-122.

Cohen, I. Bernard, 1985, *Revolution in Science* (Cambridge, Mass.: Belknap Press).

Cohen, Robert S., 1971, 'Tacit, Social and Hopeful,' in *Interpretations of Life and Mind: Essays around the Problem of Reduction*, ed. Margorie Grene (New York: Humanities Press).

Collier, Rohan, 1981, 'Neo-Darwinism: The Unit of Natural Selection,' *Theoria to Theory*, 14, 339-345.

Craig, William Lane, 1994, *Reasonable Faith*, Revised ed. (Wheaton, Illinois: Crossway Books).

Dampier, William Cecil, 1957, *A Shorter History of Science* (New York: Meridian Books).

Darden, Lindley, and Nancy Maull, 1977, 'Interfield Theories,' *Philosophy of Science*, 44, 43-64.

Darwin, Charles, 1998, *The Origin of Species*, 2nd ed. (Oxford: Oxford University Press).

Davies, P. C. W., 1989, 'The Physics of Complex Organization,' in *Theoretical Biology: Epigenetic and Evolutionary Order*, eds. Brian Goodwin and Peter Saunders (Edinburgh University Press: Edinburgh).

Davies, Paul, 1983, *God & the New Physics* (New York: Simon & Schuster).

Dawkins, Richard, 1981, 'In Defense of Selfish Genes,' *Philosophy*, 56, 556-573.

Dawkins, Richard, 1982, *The Extended Phenotype* (Oxford: Oxford University Press).

Dawkins, Richard, 1986, *The Blind Watchmaker* (New York: W. W. Norton).

Dawkins, Richard, 1989, *The Selfish Gene*, new ed. (Oxford: Oxford University Press).

Dawkins, Richard, 1995, *River Out of Eden: A Darwinian View of Life* (New York: Basic Books).

Dennett, Daniel C., 1995, *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (New York: Simon & Schuster).

Descartes, René, 1970, *The Philosophical Works of Descartes*, vol. 1, trans. Elizabeth S. Haldane and G. R. T. Ross (London: Cambridge University Press).

Duhem, Pierre, 1954, *The Aim and Structure of Physical Theory*, trans. Philip P. Weiner (New York: Atheneum).

Emerson, Ralph Waldo, 1992, *The Selected Writings of Ralph Waldo Emerson* (New York: The Modern Library).

Faber, Roger J., 1986, *Clockwork Garden: On the Mechanistic Reduction of Living Things* (Amherst: University of Mass. Press).

Feyerabend, Paul, 2001, *Against Method*, 3rd ed. (London: Verso).

Fodor, Jerry, 1995, 'Special Sciences,' in *The Philosophy of Science*, eds. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, Mass.: MIT Press).

Foss, Jeffrey E., 1995, 'Materialism, Reduction, Replacement, and the Place of Consciousness in Science,' *Journal of Philosophy*, 92, 401-429.

Frank-Kamenetskii, Maxim D. , 1997, *Unraveling DNA: The Most Important Molecule of Life*, revised ed., trans. Lev Liapin (Reading, Mass.: Addison-Wesley).

French, Roger, 1994, *Ancient Natural History* (London: Routledge).

Gallagher, Kenneth T., 1992, 'Dawkins in Biomorph Land,' *International Philosophical Quarterly*, 32, 501-513.

Garfinkel, Alan, 1995, 'Reductionism,' in *The Philosophy of Science*, eds. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, Mass.: MIT Press).

Gasper, Philip, 1990, 'Explanation and Scientific Realism,' in *Explanation and Its Limits*, ed. Dudley Knowles (Cambridge: Cambridge University Press).

Gasper, Philip, 1995, 'Causation and Explanation,' in *The Philosophy of Science*, eds. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, Mass.: MIT Press).

Gellner, E. A., 1956, 'Explanations in History,' *Proceedings of the Aristotelian Society*, 30, 157-176.

Giere, Ronald N., 1999, *Science Without Laws* (Chicago: University of Chicago Press).

Girill, T. R., 1974, 'Identity, Causality, and the Regressiveness of Micro-Explanations,' *Dialectica*, 28, 223-238.

Girill, T. R., 1976a, 'Criteria for the Part-Whole Relation in Micro-Reductions,' *Philosophia*, 6, 69-79.

Girill, T. R., 1976b, 'Evaluating Micro-Explanations,' *Erkenntnis*, 10, 387-405.

Godbey, John W., Jr., 1978, 'Disjunctive Predicates and the Reduction of Psychology,' *Mind*, 87, 433-435.

Goodwin, Brian, 1979, 'On Morphogenetic Fields,' *Theoria to Theory*, 13, 109-114.

Goodwin, Brian, 1989, 'Evolution and the Generative Order,' in *Theoretical Biology: Epigenetic and Evolutionary Order*, eds. Brian Goodwin and Peter Saunders (Edinburgh: Edinburgh University Press).

Goodwin, Brian, 1994, *How the Leopard Changed Its Spots: The Evolution of Complexity* (New York: Charles Scribner's Sons).

Gould, Stephen Jay, 1977, 'Caring Groups and Selfish Genes,' Natural History 86 (12), 20-24.

Gould, Stephen Jay, 1990, *Wonderful Life: The Burgess Shale and the Nature of History* (New York: Norton).

Grant, Colin, 1991, 'The Gregarious Metaphor of the Selfish Gene,' *Religious Studies*, 27, 431-450.

Grene, Marjorie, 1971, 'Reducibility: Another Side Issue?', in *Interpretations of Life and Mind: Essays around the Problem of Reduction*, ed. Margorie Grene (New York: Humanities Press).

Gribbin, John, 1999, *Almost Everyone's Guide to Science* (New Haven, Conn.: Yale University Press).

Griffiths, P. E., and R. D. Gray, 1994, 'Developmental Systems and Evolutionary Explanation,' *Journal of Philosophy*, 91, 277-304.

Harré, Rom, 1992, *The Philosophies of Science*, 2nd edition (Oxford: Oxford University Press).

Hempel, Carl G., and Paul Oppenheim, 1953, 'The Logic of Explanation,' in *Readings in the Philosophy of Science*, eds. Herbert Feigl and May Brodbeck (New York: Appleton-Century-Crofts).

Ho, Mae-Wan, and Peter Saunders, eds., 1984, *Beyond Neo-Darwinism: An Introduction to the New Evolutionary Paradigm* (London: Academic Press, Inc.).

Horgan, Terence, 1993, 'From Supervenience to Superdupervenience: Meeting the Demands of a Material World,' *Mind*, 102, 555-586.

Howard, Don, 1979, 'Commoner on Reduction,' *Environmental Ethics*, 1, 159-176.

Huberman, Bernardo A., 1989, 'The Adaptation of Complex Systems,' in *Theoretical Biology: Epigenetic and Evolutionary Order*, eds. Brian Goodwin and Peter Saunders (Edinburgh: Edinburgh University Press).

Hull, David, 1974, *Philosophy of Biological Science* (Englewood Cliffs, New Jersey: Prentice-Hall).

Hull, David, 1998, 'A Clash of Paradigms or the Sound of One Hand Clapping,' *Biology and Philosophy*, 13 (1998), 587-595.

Hume, David, 1975, *An Enquiry Concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Clarendon Press).

Hume, David, 1985, *A Treatise of Human Nature*, ed. Ernest C. Mossner (London: Penguin Books).

Jackson, Frank, 1980, 'A Note on Physicalism and Heat,' *Australasian Journal of Philosophy*, 58, 26-34.

Jackson, Frank, 1982, 'Epiphenomenal Qualia,' *Philosophical Quarterly*, 32, 127-136.

Jaki, Stanley L., 1966, *The Relevance of Physics* (Chicago: University of Chicago Press).

James, William, 1984, *The Essential Writings*, ed. Bruce W. Wilshire (Albany, NY: SUNY Press).

James, William, 1996, *Essays in Radical Empiricism* (Lincoln, Nebraska: University of Nebraska Press).

James, William, 2000, *Pragmatism and Other Writings* (New York: Penguin Books).

Joad, C. E. M., 1957, *Guide to Philosophy* (New York: Dover).

Jones, Steve, 1995, *The Language of Genes* (New York: Anchor Books).

Kauffman, Stuart A., 1989, 'Origins of Order in Evolution: Self-Organization and Selection,' in *Theoretical Biology: Epigenetic and Evolutionary Order*, eds. Brian Goodwin and Peter Saunders (Edinburgh: Edinburgh University Press).

Kenny, Anthony J. P., 1971, 'The Homunculus Fallacy,' in *Interpretations of Life and Mind: Essays around the Problem of Reduction*, ed. Margorie Grene (New York: Humanities Press).

Kim, Jaegwon, 1992, 'Multiple Realization and the Metaphysics of Reduction,' *Philosophy and Phenomenological Research*, 52, 1-26.

Kim, Jaegwon, 1996, *Philosophy of Mind* (Boulder, Col.: Westview Press).

Kinoshita, Joyce, 1990, 'How Do Scientific Explanations Explain?' in *Explanation and Its Limits*, ed. Dudley Knowles (Cambridge: Cambridge University Press).

Kitcher, Philip, Kim Sterelny, and C. Kenneth Waters, 1990, 'The Illusory Riches of Sober's Monism,' *Journal of Philosophy*, 87, 158-161.

Kleiner, Scott A., 1975, 'Essay Review—The Philosophy of Biology,' *Southern Journal of Philosophy*, 13, 523-542.

Kleiner, Scott A., 1996, 'Goldschmidt's Heresy and the Explanatory Promise of Ontogenetic Evolutionary Theory,' *Philosophica (Belgium)*, 58(2), 51-72.

Kuhn, Thomas S., 1970, *The Structure of Scientific Revolutions*, 2nd ed. (Chicago, IL: University of Chicago Press).

Kuhn, Thomas S., 1985, *The Copernican Revolution* (Cambridge, Mass.: Harvard University Press).

Laudan, Larry, 1992, 'A Problem-Solving Approach to Scientific Progress,' in *Scientific Revolutions*, ed. Ian Hacking (Oxford: Oxford University Press).

Levins, R., 1970, 'Complex Systems,' in *Towards a Theoretical Biology*, ed. C. H. Waddington (Chicago: Aldine).

Lewontin, R. C., 1970, 'On the Irrelevance of Genes,' in *Towards a Theoretical Biology*, ed. C. H. Waddington (Chicago: Aldine).

Lipton, Peter, 1990, 'Contrastive Explanation,' in *Explanation and Its Limits*, ed. Dudley Knowles (Cambridge: Cambridge University Press).

Lipton, Peter, 1991, *Inference to the Best Explanation* (London: Routledge).

Longuet-Higgins, C., 1970, 'The Seat of the Soul,' in *Towards a Theoretical Biology*, ed. C. H. Waddington (Chicago: Aldine).

Mackey, J. L., 1978, 'The Law of the Jungle: Moral Alternatives and Principles of Evolution,' *Philosophy*, 53, 455-464.

Maynard Smith, John, 1975, *The Theory of Evolution*, 3rd ed. (Cambridge: Cambridge University Press).

Maynard Smith, John, 1990, 'Explanation in Biology,' in *Explanation and Its Limits*, ed. Dudley Knowles (Cambridge: Cambridge University Press).

Mayr, Ernst, 1982, *The Growth of Biological Thought: Diversity, Evolution, and Inheritance* (Cambridge, Mass.: Belknap Press).

Mayr, Ernst, 1988, *Toward a New Philosophy of Biology* (Cambridge, Mass.: Harvard University Press).

McGinn, Colin, 1995, 'Can We Solve the Mind-Body Problem?' in *Modern Philosophy of Mind*, ed. William Lyons (New York: Everyman).

Midgley, Mary, 1979, 'Gene-Juggling,' *Philosophy*, 54, 439-458.

Midgley, Mary, 1983, 'Selfish Genes and Social Darwinism,' *Philosophy*, 58, 365-377.

Moore, G. E., 1951, *Philosophical Studies* (London: Routledge & Kegan Paul).

Moore, G. E., 1959, *Philosophical Papers* (London: George Allen & Unwin).

Moreland, J. P., 1987, *Scaling the Secular City: A Defense of Christianity* (Grand Rapids, Michigan: Baker Books).

Mumford, Stephen, 1994, 'Dispositions, Supervenience and Reduction,' *The Philosophical Quarterly*, 44, 419-438.

Nagel, Ernest, 1961, *The Structure of Science* (New York: Harcourt, Brace & World).

Nagel, Thomas, 1995, 'What Is It Like to Be a Bat?', reprinted in *Modern Philosophy of Mind*, ed. William Lyons (New York: Everyman).

Oppenheim, Paul, and Hilary Putnam, 1995, 'Unity of Science as a Working Hypothesis,' in *The Philosophy of Science*, eds. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, Mass.: MIT Press).

Owens, David, 1989, 'Levels of Explanation,' *Mind*, 98, 59-79.

Pattee, H. H., 1970, 'The Problem of Biological Hierarchy,' in *Towards a Theoretical Biology*, ed. C. H. Waddington (Chicago: Aldine).

Peirce, Charles Sanders, 1955, *Philosophical Writings of Peirce*, ed. Justus Buchler (New York: Dover).

Perry, Clifton, 1979, 'Holism and the Issue of Causality', *Diálogos*, 34, 81-91.

Petrie, Bradford, 1987, 'Global Supervenience and Reduction,' *Philosophy and Phenomenological Research*, 48, 119-130.

Place, U. T., 1995, 'Is Consciousness a Brain Process?' reprinted in *Modern Philosophy of Mind*, ed. William Lyons (New York: Everyman).

Polanyi, Michael, 1962, *Personal Knowledge: Towards a Post-Critical Philosophy* (Chicago: University of Chicago Press).

Polanyi, Michael, 1983, *The Tacit Dimension* (Gloucester, Mass: Peter Smith).

Polkinghorne, John, 1998, *Belief in God in an Age of Science* (New Haven, Conn.: Yale University Press).

Pollack, Robert, 1994, *Signs of Life: The Language and Meaning of DNA* (New York: Houghton Mifflin Company).

Popper, Karl, 1992, 'The Rationality of Scientific Revolutions,' in *Scientific Revolutions*, ed. Ian Hacking (Oxford: Oxford University Press).

Prigogine, Ilya, 1971, 'Unity of Physical Laws and Levels of Description,' in *Interpretations of Life and Mind: Essays around the Problem of Reduction*, ed. Margorie Grene (New York: Humanities Press).

Putnam, Ruth Anna, ed., 1997, *The Cambridge Companion to William James* (Cambridge: Cambridge University Press).

Quine, W. V., 1953, *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press).

Quine, W. V., 1960, *Word and Object* (New York: Wiley).

Quine, W. V., 1969,*Ontological Relativity and Other Essays* (New York: Columbia University Press).

Quine, W. V., 1970, 'Grades of Theoreticity,' in *Experience and Theory*, ed. Lawrence Foster and J. W. Swanson (Amherst: University of Massachusetts Press).

Quine, W. V., 1991, 'Two Dogmas in Retrospect,' *Canadian Journal of Philosophy*, 21, 265-274.

Radnitzky, Gerard, and W. W. Bartley, III, eds., 1987, *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*, (La Salle, Ill: Open Court).

Reid, Robert G. B., 1985, *Evolutionary Theory: The Unfinished Synthesis* (Ithaca, New York: Cornell University Press).

Rescher, Nicholas, 1980, *Induction: An Essay on the Justification of Inductive Reasoning* (Pittsburgh: University of Pittsburgh Press).

Rose, Steven, 1997, *Lifelines: Biology Beyond Determinism* (New York: Oxford University Press).

Rosenberg, Alexander, 1985, 'Darwinism Today—and Tomorrow—but not Yesterday,' *PSA*, 2, 157-173.

Rosenberg, Alexander, 1986, *The Structure of Biological Science* (New York: Cambridge University Press).

Ross, W. D. 1930, *The Right and the Good* (Oxford: Oxford University Press).

Russell, Bertrand, 1971, *The Problems of Philosophy* (New York: Oxford University Press).

Santayana, George, 1923, *Scepticism and Animal Faith: An Introduction to a System of Philosophy* (New York: Charles Scribner's Sons).

Sarkar, Sahotra, 1998, *Genetics and Reductionism* (Cambridge: Cambridge University Press).

Schaffner, Kenneth F., 1967, 'Approaches to Reduction,' *Philosophy of Science*, 34, 137-147.

Schaffner, Kenneth F., 1969, 'The Watson-Crick Model and Reductionism,' *British Journal of Philosophy*, 20, 325-348.

Scott, William T., 1971, 'Tacit Knowledge and the Concept of Mind,' in *Interpretations of Life and Mind: Essays around the Problem of Reduction*, ed. Margorie Grene (New York: Humanities Press).

Searle, John R., 1992, *The Rediscovery of the Mind* (Cambridge: MIT Press).

Singer, Charles, 1959, *A History of Biology*, 3rd ed., revised (London: Abelard-Schuman).

Smart, J. C. C., 1978, 'Is Occam's Razor a Physical Thing?,' *Philosophy*, 53, 382-385.

Smart, J. C. C., 1990, 'Explanation—Opening Address,' in *Explanation and Its Limits*, ed. Dudley Knowles (Cambridge: Cambridge University Press).

Smart, J. C. C., 1995, 'Sensations and Brain Processes,' reprinted in *Modern Philosophy of Mind*, ed. William Lyons (New York: Everyman).

Snyder, John, and C. Leland Rodgers, 1995, *Biology*, 3rd ed. (Hauppauge, New York: Barron's Educational Series, Inc.).

Sober, Elliott, 1984, *The Nature of Selection* (Chicago: University of Chigago Press).

Sober, Elliott, 1990a, 'Let's Razor Ockham's Razor,' in *Explanation and Its Limits*, ed. Dudley Knowles (Cambridge: Cambridge University Press).

Sober, Elliott, 1990b, 'The Poverty of Pluralism: A Reply to Sterelny and Kitcher,' *Journal of Philosophy*, 87, 151-158.

Sober, Elliott, 1993, *Philosophy of Biology* (Boulder, Colorado: Westview Press).

Sober, Elliott, 1995, 'Evolution and Optimality: Feathers, Bowling Balls, and the Thesis of Adaptationism,' *Philosophic Exchange*, 26, 41-57.

Sorley, W. R., 1918, *Moral Values and the Idea of God* (Cambridge: Cambridge University Press).

Sterelny, Kim, and Paul E. Griffiths, 1999, *Sex and Death: An Introduction to Philosophy of Biology* (Chicago: University of Chicago Press).

Sterelny, Kim, and Philip Kitcher, 1988, 'The Return of the Gene,' *Journal of Philosophy*, 85, 339-361.

Swinburne, Richard, 1990, 'The Limits of Explanation,' in *Explanation and Its Limits*, ed. Dudley Knowles (Cambridge: Cambridge University Press).

Taylor, Charles, 1971, 'How is Mechanism Conceivable?', in *Interpretations of Life and Mind: Essays around the Problem of Reduction*, ed. Margorie Grene (New York: Humanities Press).

Trout, J. D., 1995, 'Reductionism and the Unity of Science,' in *The Philosophy of Science*, eds. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, Mass.: MIT Press).

van Fraassen, Bas, 1995a, 'The Pragmatics of Explanation,' in *The Philosophy of Science*, eds. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, Mass.: MIT Press).

van Fraassen, Bas, 1995b, 'To Save the Phenomena,' in *The Philosophy of Science*, eds. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, Mass.: MIT Press).

Wächtershäuser, Günter, 1987, 'Light and Life: On the Nutritional Origins of Sensory Perception,' in *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*, eds. Gerard Radnitzky and W. W. Bartley, III (La Salle, Ill: Open Court).

White, Nicholas P., 1976, *Plato on Knowledge and Reality* (Indianapolis, Indiana: Hackett Publishing Co.)

Whitehead, Alfred North, 1981, *A Key to Whitehead's Process and Reality*, ed. Donald W. Sherburne (Chicago: University of Chicago Press).

Williams, G. C., 1966, *Adaptation and Natural Selection* (Princeton: Princeton University Press).

Williams, Mary B., and Alexander Rosenberg, 1985, '"Fitness" in Fact and Fiction: A Rejoinder to Sober,' *Journal of Philosophy*, 82, 738-749.

Wilson, Edward O., 1998, *Consilience: The Unity of Knowledge* (New York: Alfred A. Knopf).

Wolpert, L., 1970, 'Positional Information and Pattern Formation,' in *Towards a Theoretical Biology*, ed. C. H. Waddington (Chicago: Aldine).