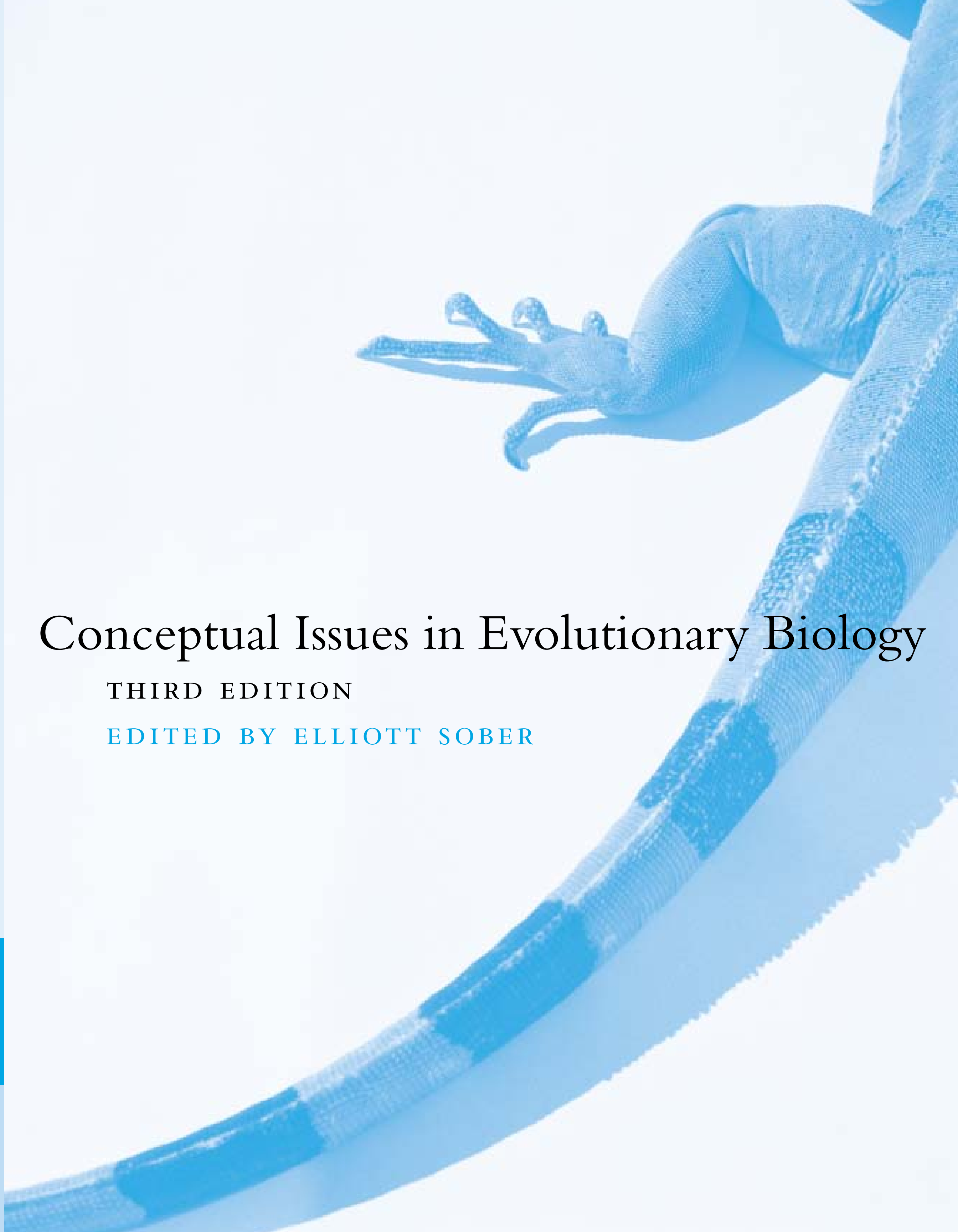


Conceptual Issues in Evolutionary Biology

THIRD EDITION

EDITED BY ELLIOTT SOBER

Conceptual Issues in Evolutionary Biology

Conceptual Issues in Evolutionary Biology

Third Edition

Edited by Elliott Sober

Bradford Books
The MIT Press
Cambridge, Massachusetts
London, England

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Stone Serif and Stone Sans on 3B2 by Asco Typesetters, Hong Kong, and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Conceptual issues in evolutionary biology / edited by Elliott Sober. — 3rd ed.
p. cm.
Includes bibliographical references (p.).
ISBN-13: 978-0-262-19549-2 (alk. paper) — 978-0-262-69338-7
ISBN-10: 0-262-19549-6 (alk. paper) — 0-262-69338-0 (pbk. : alk. paper)
1. Evolutionary (Biology) 2. Evolution (Biology)—Philosophy. 3. Biology—Philosophy.
I. Sober, Elliott.

QH366.2.C64 2006
576.8—dc22 2006044934

10 9 8 7 6 5 4 3 2 1

Contents

IV Women in the Evolutionary Process

V Evolutionary Psychology

VI Laws in Evolutionary Theory

VII Reductionism

VIII Essentialism and Population Thinking

IX Species

X Phylogenetic Inference

XI Race–Social Construction or Biological Reality?

XII Cultural Evolution

XIII Evolutionary Ethics

Preface

This anthology brings together essays written by philosophers and scientists that address conceptual issues that arise in the theory and practice of evolutionary biology. This third edition differs substantially from the previous two, which were published in 1984 and 1994. Four new sections have been included (on women in the evolutionary process, evolutionary psychology, the existence of laws in biology, and race) and some additions and subtractions have been made in chapters present in previous editions.

Each part in this collection contains two or three chapters that develop opposing views of the problem at hand. Evolutionary biology is a living, growing discipline, and the same is true of the philosophy of evolutionary biology. One sign that a discipline is growing is that there are open questions, with multiple answers still in competition. I hope the clash of ideas presented here will be useful to evolutionary biologists and to philosophers of biology, both in their teaching and in their research.

In the remainder of this preface, I will briefly indicate what some of the major issues are that animate the chapters that follow.

I Fitness

Darwin accorded a preeminent role to the process of natural selection in his account of how life has evolved. Central to the concept of natural selection is the idea of fitness: if the organisms in a population undergo a selection process, they must differ from each other in terms of their abilities to survive and reproduce. Herbert Spencer coined the slogan "survival of the fittest" to describe Darwin's theory and Darwin adopted this description phrase, thinking that it might help readers avoid misunderstanding what he mean by his own term, "natural selection." Had he realized the confusions that would ensue, maybe Darwin would have distanced himself from this slogan. For once the theory is summarized by the phrase "the survival of the fittest," it invites the following line of criticism: Who survives? Those who are fit. And who are the fit? Those who survive. If the theory of natural selection comes to no more than this, then the "theory" is no theory at all. It is a piece of circular reasoning, an empty truism, which masquerades as a substantive explanation of what we observe.

In the first chapter in part I, "The Propensity Interpretation of Fitness," Susan Mills and John Beatty address this criticism by clarifying the probabilistic character of the concept of fitness. Their goal is to describe how the concept of fitness figures in contemporary biology and to show that the charge of circularity is entirely misguided. They formulate an interpretation of the fitness concept that draws on more general ideas concerning how the concept of probability might be understood; if fitnesses are probabilities, perhaps the propensity interpretation of probability (defended by Karl Popper in his influential book *The Logic of Scientific Discovery*) will help clarify what fitness is.

In the second chapter, I take up some of the questions that Susan Finsen (née Mills) and Beatty subsequently raised about their proposal. An organism has an expected number offspring, an expected number of grandoffspring, and a probability of having descendants that exist 1000 generations hence. Which of these probabilistic concepts is the organism's fitness? And is it always true that an organism's prospects for reproductive success can be represented as a probabilistic expectation? These questions must be addressed if the adequacy of the propensity interpretation of fitness is to be evaluated.

II Units of Selection

Human beings are organisms, so it may strike us as entirely natural to think that the parts of organisms exist in order to benefit the organisms that contain them. We have hearts, so we naturally think that hearts exist in order to help organisms circulate their blood. And each of our cells contains genes, so we naturally think that genes exist in order to help organisms transmit traits from parents to offspring.

If we are prepared to think that hearts and genes have the functions of helping organisms to perform various tasks, why not frameshift this idea up a level and conclude that organisms have the function of helping the groups in which they live? If hearts help organisms survive, why not also say that organisms have hearts to help the species to which those organisms belong to avoid extinction? Conversely, if parts can have the function of helping the wholes in which they exist, why can't the opposite relationship also obtain? Why not think of organisms as devices that have the function of guaranteeing the survival and reproduction of the genes they contain?

These questions are central to what is now called the problem of the units of selection. Let us assume that a trait—the opposable thumb, for example—evolved because it was good for the things that possessed it. But which objects should we regard as the relevant beneficiaries? Did the opposable thumb evolve because it helped the species to avoid extinction, or because it helped organisms survive and reproduce, or because it helped genes coding for opposable thumbs to make their way into successive generations?

Darwin usually thought of natural selection as a process in which traits evolve because they benefit individual organisms. The two most famous exceptions to this pattern of thinking were his discussion in *The Origin of Species* of sterile castes in the social insects and his discussion in *The Descent of Man* of human morality. In both instances, Darwin argued that a trait sometimes evolves because it benefits the group and in spite of the fact that it is deleterious to the individuals possessing it. Later on, evolutionary biologists came to call such traits ''altruistic.'' The idea that some traits are group adaptations was a standard part of biological thinking during the heyday of the Modern Synthesis (1930–1960). For example, in his 1937 book *Genetics and the Origin of Species*, Theodosius Dobzhansky suggested that sexual reproduction may be a group adaptation, its function being to ensure that a species is genetically diverse, so that the species is less likely to go extinct if the environment suddenly changes.

Even though group selection thinking was for many years a standard part of the evolutionary biologist's toolkit, its fortunes plummeted after 1966; this was the year in which George C. Williams published his book *Adaptation and Natural Selection*. The most prominent message in Williams's book was that group selectionism is a kind of sloppy thinking. Biologists had uncritically talked about traits existing ''for the good of the species'' even though more parsimonious explanations can be provided at lower levels. Although Darwin almost always favored individual selection over group selection, Williams's critique of group selection thinking did not conclude that the classical Darwinian picture was the best way to think about adaptation. Rather, he suggested that we descend to a lower level still. The real unit of selection, Williams argued, is not the group, nor even the individual, but the gene. Thus was born the view of natural selection that Richard Dawkins later popularized in his book *The Selfish Gene*.

The first chapter in part 2 consists of excerpts from Williams's *Adaptation and Natural Selection*. Williams emphasizes the importance of not assuming that a trait that benefits a group must be a group adaptation. The benefit to the group may be a side-effect, and not the reason the trait evolved. Williams also argues that hypotheses of group adaptation are less parsimonious than those that posit adaptations at lower levels. In defense of the genic point of view, Williams contends that genes have a longevity that gene combinations and whole organisms do not. Williams also argues, repeating an argument that R. A. Fisher made in his book *The Genetical Theory of Natural Selection*, that group selection must be a weak force, compared with individual selection, because groups usually go extinct and found colonies at a slower rate than the rate at which organisms die and reproduce.

In the second chapter in this section, ''Levels of Selection: An Alternative to Individualism in Biology and the Human Sciences,'' David Sloan Wilson argues that the selfish gene theory—the idea that the gene is the one and only unit of selection—rests on a fallacy. Wilson does not urge a return to uncritical group selectionism; rather, he defends a pluralistic conception of selection, according to which adaptations can evolve for a variety of reasons. According to Wilson, the living world contains selfish genes, but it also contains well-adapted organisms and well-adapted groups; genic selection occurs, but so too do individual selection and group selection.

III Adaptationism

Although Darwin accorded a central role to the process of natural selection, he allowed that other processes can influence the course of evolution. How important these other processes are, and how we should endeavor to understand the features of the living things that we observe, are the main issues in the controversy concerning adaptationism.

When a biologist studies a complex characteristic of morphology, physiology, or behavior, the first impulse is often to ask ''What is this trait for?'' The mind searches for the trait's adaptive significance. We observe the dorsal fins on a dinosaur and immediately set to work thinking about whether the fins were for self-defense, or for thermal regulation, or to attract mates. If one hypothesis of adaptive advantage does not pan out, we discard it and invent another.

In the first chapter in part III, ''The Spandrels of San Macro and the Panglissian Paradigm: A Critique of the Adaptationist Programme,'' Stephen Jay Gould and Richard C. Lewontin argue that this impulse toward adaptationist thinking has led biologists to neglect the possibility that many traits may be present for reasons that have nothing to do with their adaptive significance. Adaptationism, they suggest, is an assumption that is as pervasive as it is unproven. In the part's second chapter, ''Optimization Theory in Evolution,'' John Maynard Smith replies to this criticism by outlining the elements that figure in adaptationist explanations; he defends the adaptationist approach by suggesting how particular adaptationist explanations can be tested.

Two questions float through the debate between adaptationists and anti-adaptationists. The first concerns what is true in the natural world: how important has natural selection been in evolution? Is it true that virtually all traits have evolved because of natural selection, or is this true only for some significantly smaller subset? The second question concerns how nature should be studied: is it sound methodology to formulate and test adaptive hypotheses? Should alternative, nonadaptive explanations, also be contemplated? That these are separate questions can be seen by considering the fact that one might be agnostic about what is true in nature, but still maintain that evolutionary inquiry requires the formulation and testing of hypotheses about natural selection. This distinction between biological and methodological questions is important to bear in mind in thinking about the adaptationism debate.

IV Women in the Evolutionary Process

Gould and Lewontin's critique of adaptationism was motivated in large measure by the rising popularity of sociobiology. Biologists had been interested in the evolution of behavior (human and nonhuman) for a long time, but Edward O. Wilson's publication in 1975 of his book *Sociobiology—The New Synthesis* was a rallying cry; it called on biologists to apply recent advances in evolutionary biology to the project of systematically understanding human mind and culture. For Gould and Lewontin, sociobiology was a symptom of a pervasive deficiency in evolutionary thinking. For adaptationists, sociobiology was a new opportunity.

Gould and Lewontin find it highly significant that adaptationists are happy to invent a new adaptive hypothesis to explain a trait if an old explanation is refuted by the evidence. Does this represent a flaw in adaptationism, or is it an unobjectionable consequence of the fact that adaptationism is a flexible research program? The two chapters in the present section embody opposing answers to this question in the context of discussing how evolutionary biology should address questions concerning female sexuality.

Sarah Blaffer Hrdy, in her article "Empathy, Polyandry, and the Myth of the Coy Female," writes as a sociobiologist who found herself, starting in the 1970s, increasingly dissatisfied with the received wisdom on female sexual behavior. In his book *The Descent of Man, and Selection in Relation to Sex*, Darwin wrote that females tend to be "coy" and males tend to be "eager" in their mating behavior. By this he meant that females are more choosy in deciding with whom they will mate, and males are more promiscuous. A. J. Bateman defended this general thesis and attempted to give it an evolutionary explanation in a paper he published in 1948. In his laboratory work on fruitflies, Bateman observed that almost all the females reproduced, but about a fifth of the males did not. He argued that males increase their fitness by mating with multiple females, but that females do not increase their fitness by mating with multiple males. Males, unlike females, have something to gain from seeking new mating opportunities. This is why males are promiscuous and females are choosy. Hrdy's dissatisfaction with this paradigm grew out of a growing body of observational evidence that female primates often have multiple partners. These observations led her to seek out plausible adaptive explanations of this pervasive pattern. One of them, the "manipulation hypothesis," says that females gain protection and resources for their infants by having multiple partners. In addition to defending a new adaptationist explanation of female mating behavior, Hrdy also speculates about why biology largely ignored female sexual behavior until the 1970s. She conjectures that the growing representation of women in primatology, and the empathy that women brought to studying female primates, may have been contributing factors.

In some respects Elisabeth Lloyd is on the same wave length as Hrdy, but in other respects, she is not. In her chapter ''Pre-theoretical Assumptions in Evolutionary Explanations of Female Sexuality,'' Lloyd agrees with Hrdy that male bias has impeded the understanding of female sexuality. But whereas Hrdy is reacting against a conception of females as essentially different from males (the ''coy'' versus ''eager'' contrast), Lloyd takes issue with the assumption that female sexuality is to be understood is if it were a carbon copy of male sexuality. Her particular subject is female orgasm. If male orgasm has a reproductive function, must the same be true of female orgasm? Lloyd argues against this assumption. Following a suggestion made by Donald Symons in his 1979 book *The Evolution of Human Sexuality*, Lloyd suggests that female orgasm may be what Gould and Lewontin call a spandrel. Females have orgasms for the same reason that males have nipples. Neither trait was selected for; rather, each evolved because of a developmental correlation. Males and females both have nipples because nipples are the outcome of developmental processes present in both sexes; this means that selection for nipples in females leads male nipples to evolve as well, even though nipples have no evolutionary function in males. Lloyd suggests that the same may be true of female orgasm; it has no evolutionary function, but evolved because there was selection for male orgasm. Lloyd concludes her chapter with a more general lesson—that philosophers of science need to pay closer attention to the social assumptions and prior commitments that influence the scientific process.

V Evolutionary Psychology

Evolutionary psychology is a more recent adaptationist project than sociobiology. Whereas sociobiologists tend to focus on behaviors and try to explain them by ferreting out their adaptive significance, evolutionary psychologists tend to think of cognitive mechanisms, not behaviors, as the fundamental subject for evolutionary theorizing. Evolutionary psychologists also emphasize an idea that Maynard Smith mentions in his chapter in part III—that natural selection often occurs slowly, so that the traits we observe in a present population are not the ones that would be optimal for its present circumstance. Rather, the traits we observe now are often vestiges of traits that were adaptive in earlier, ancestral, environments. Evolutionary psychologists suggest that the way to understand the human mind as it presently is constituted is to understand the adaptive problems our ancestors faced; we are adapted to past conditions, not to present conditions.

This position is developed in the first chapter in this section, ''Toward Mapping the Evolved Functional Organization of Mind and Brain,'' by John Tooby and Leda Cosmides, who are two of the main founders of evolutionary psychology. Tooby and Cosmides argue that the human mind should be viewed as a collection of different adaptive mechanisms, each evolved to address a different adaptive problem. They oppose the idea that the mind guides behavior by deploying an all-purpose learning strategy; rather, it is a tool box filled with a large number of special-purpose cognitive tools. Tooby and Cosmides also argue that the complex adaptations that the human mind possesses tend to be species-typical universals. Whereas sociobiologists have often been open to the possibility that different individuals in the same species might have different adaptive features, evolutionary psychologists have been less friendly to the hypothesis of within-species adaptive variation; the only context in which they think this is plausible concerns differences between the sexes. In the second chapter in this section, "Evolutionary Psychology: A Critique," David Buller raises questions about the main tenets of evolutionary psychology. He agrees with Cosmides and Tooby that the mind has evolved, and that adaptive hypotheses about its features need to be considered; Buller's objections concern the details of evolutionary psychology, not its broad goals.

VI Laws in the Evolutionary Process

During the "bad old days," philosophy of science was dominated by logical empiricist ideas and physics was the science that dominated philosophical thinking about what science is. Both the influence of logical empiricism and the fixation on physics encouraged the idea that scientific inquiry is first and foremost the search for general laws. When philosophers of science took an occasional (and usually brief) look at biology, they often took away the impression that biology contains no laws, and just as often, they drew the conclusion that biology is deficient as a science. Echoing Kant, philosophers often concurred that there can be no "Newton of the blade of grass." Biology is not just different, it is inferior.

With the demise of logical empiricism and the growing realization that philosophy of science has to be more than philosophy of physics, the idea took hold that there are different kinds of science. One version of this idea holds that some sciences aim to find general laws while others seek to reconstruct the histories of particular events. Relativity theory and quantum mechanics are sciences of the first type, but evolutionary biology is a historical science. It is different, but not inferior.

John Beatty's chapter, "The Evolutionary Contingency Thesis," develops this view of evolutionary biology. His thesis is that there are no biological laws. When a biological generalization is true, it owes its truth to the fact that contingent evolutionary events turned out one way rather than another. Biological generalizations, since they depend on contingencies, are themselves contingent. And since they are contingently true, they are not laws. Beatty offers two further arguments in favor of this thesis. He argues that evolutionary biology is often given over to disputes about "relative significance." Two examples of what Beatty has in mind were the subjects of previous chapters of this anthology: the debate about adaptationism concerns how important natural selection has been and the units of selection debate concerns how important group selection has been. Beatty contends that the centrality of such disputes to biology is evidence that there are no biological laws. Beatty also argues that evolutionary biology has forsaken the Newtonian ideal of seeking out parsimonious explanations and that this methodological difference between physics and evolutionary biology provides further evidence that there are no distinctively biological laws.

My chapter "Two Outbreaks of Lawlessness in Philosophy of Biology" replies to Beatty's arguments and also to arguments for a similar conclusion advanced by Alexander Rosenberg in his 1994 book *Instrumental Biology or the Disunity of Science*. In reply to Beatty, I suggest that if a contingent evolutionary event *E* causes a biological generalization *G* to be true later on, then we should expect there to be a law linking *E* to *G*. This is consistent with Beatty's claim that if *G* is contingent on *E* (where *E* is contingent), then *G* is not a law. I also argue that the prevalence of relative significance controversies in evolutionary biology is no evidence for biology's lawlessness and that biology has not abandoned the principle of parsimony, when that principle is properly understood. Rosenberg's brief for lawlessness is based on the fact that biological properties are multiply realizable. Take a biological property like "fitness" or "predator;" these predicates apply to organisms that differ vastly from each other in terms of their physical properties. Rosenberg thinks this fact of multiple realizability rules out there being any biological laws (aside from a generalization he calls the principle of natural selection). He also contends that the only objective probabilities discussed in science are the ones found in quantum mechanics. This means that the probabilities discussed in evolutionary biology are mere reflections of our ignorance; biologists use these probabilities to describe evolutionary processes because they don't know enough about the physical details. Rosenberg takes this to show that the probabilistic generalizations used in evolutionary theory do not describe objective lawful relations.

VII Reduction

Philosophers interested in how science changes have devoted a great deal of attention to the issue of theoretical reduction. When a new theory replaces an old one, is this because the later theory shows that the former theory is false or because the new theory captures and supplements the insights of the theory it supercedes? The relation of Mendelian genetics to molecular biology has been an important test case for this question. Does modern molecular theory show that Mendel's ideas were false? Or does it show that Mendel was right? Advocates of the latter position tend to say that Mendelian genetics reduces to molecular biology.

In fact the problem of reduction—of understanding what it means for one theory to reduce to another—is more complicated than this. Sometimes a later theory shows that an earlier theory is true only in a limiting case. Einstein's theory of special relativity is inconsistent with Newtonian mechanics, but the latter can be thought of as a special case of the former; Newtonian theory is a better and better approximation as objects move more and more slowly. This leads some philosophers to maintain that Newtonian mechanics reduces to special relativity. A reducing theory (when supplemented with appropriate ''bridge laws'' that connect the vocabularies of the theories, as Ernest Nagel emphasized in his book *The Structure of Science*) may entail the reduced theory, or the reducing theory may (with bridge laws) entail that the reduced theory is false, though true as a special case. When a successor theory utterly discredits an earlier theory, showing not just that it is false but that it doesn't even hold true in a special case, no reduction relation can be claimed to obtain. Here we have theory replacement, not reduction of any kind. This is arguably what happened when Lavoisier's theory of combustion replaced the phlogiston theory.

In the first chapter in this section, ''1953 and All That: A Tale of Two Sciences,'' Philip Kitcher defends the antireductionist position. He argues that the relationship of current molecular theory to Mendel's ''laws'' is best viewed as one of theory replacement, not theoretical reduction. C. Kenneth Waters takes issue with Kitcher's arguments in his chapter, ''Why the Antireductionist Consensus Won't Survive the Case of Classical Mendelian Genetics.''

The question of how Mendelian genetics is related to molecular biology is an instance of a larger question. Reductionism is often understood as a thesis about the whole of science. The idea is that the social sciences reduce to psychology, psychology to biology, biology to chemistry, and chemistry to physics. Like a nested set of Russian dolls, all sciences above the level of physics reduce to physics, directly or indirectly. Understood in this way, reductionism is a thesis about how the true theories that have been discovered (or will be discovered) in different sciences are related to each other. Truth cannot conflict with truth, so it is obvious that all these true theories are compatible with each other. But *X*'s reducing to *Y* requires something more than just that *X* and *Y* be mutually consistent. One standard idea of what reduction requires is that *Y* must explain everything that *X* explains, and that *Y* must explain why *X* is true (to the extent that it is). Reductionism thus can be associated with the slogan ''in principle, physics explains everything.''

Kitcher's argument against reductionism in the case of Mendelian genetics draws on an influential argument against this general reductionistic thesis. Hilary Putnam and Jerry Fodor have both argued that macro-theories typically fail to reduce to micro-theories. An instance of their claim is the thesis that population biology does not reduce to particle physics. I describe and criticize this antireductionist argument in my chapter ''The Multiple Realizability Argument Against Reductionism.''

VIII Essentialism and Population Thinking

The phrase ''discerning the essence of things'' usually is interpreted to mean that one has identified what is most important. However, the phrase's familiarity should not lead us to forget that the word ''essence'' is part of a substantive philosophical doctrine, one that has exercised a considerable influence on metaphysics, both ancient and modern. Essentialism is a doctrine about natural kinds. Gold is a kind of thing; there are many samples of gold and they differ from each other in numerous respects. In the light of this diversity among gold things, how should science endeavor to understand what gold is?

The essentialist replies that science aims to discover properties that are separately necessary and jointly sufficient for being gold. If all and only the specimens of gold have a particular atomic number, then that atomic number may be the essence of gold. Discovery of essences is not an idle exercise but is fundamental to science's search for explanation. We understand what gold is by seeing what all gold things have in common. Understood in this way, the existence of variation within a kind is a distraction from what is important. Gold things vary from each other, but this variation needs to be set aside. Variation is a veil that we must see through if we are to see what is important.

In "Typological versus Population Thinking" Ernst Mayr argues that evolutionary biology has rejected essentialism in favor of an alternative philosophical framework, which he terms population thinking. I try to provide further clarification of Mayr's thesis in my chapter "Evolution, Population Thinking, and Essentialism." According to Mayr, essentialism may be a good research strategy when chemists try to understand gold, but it would be disastrous as a strategy for investigations in population biology. From a Darwinian point of view, the most important thing to understand about a species is how the individuals in it vary. This variation is what permits the species to evolve. A deep understanding of populational phenomena does not require that we brush aside the variation and discern what all members of a population have in common; rather, we have to characterize the variation and understand how it affects the ways in which the population will change.

IX Species

When philosophers try to cite examples of natural kinds, they often mention chemical elements and biological species. I just exemplified half this pattern by describing gold as a natural kind whose essence is its atomic number. But is it correct to think of species in the same way? Is the house mouse (*Mus musculus*) a kind of creature in the same way that gold is a kind of stuff? And when we speak of "human nature," are we supposing that there is some property (an essence) that all and only the members of our species possess?

In "A Matter of Individuality," David Hull articulates a position that he and the biologist Michael Ghiselin have developed. Picking up on Mayr's anti-essentialism, they argue that species are individuals, not natural kinds. Species are born and die and have a significant degree of internal cohesiveness while they persist. Hull defends the idea that a species is an integrated gene pool; organisms belong to the same species not because they are similar to each other, but because they reproduce with each other. Individualists such as Hull and Ghiselin have therefore endorsed Ernst Mayr's biological species concept, which defines a species as a group of populations that interbreed with each other but not with other populations.

Baum and Donoghue, in their chapter ''Choosing Among Alternative 'Phylogenetic' Species Concepts,'' oppose the biological species concept and other species concepts that are, in their terminology, ''character-based.'' This is not because they embrace the idea that species are natural kinds. Rather, they favor a ''history-based'' species concept that groups individuals into species according to their genealogical relationships. Consider four contemporaneous organisms *a*, *b*, *c*, and *d*. The fact that *a* and *b* breed with each other, but cannot interbreed with *c* and *d*, does not show that *a* and *b* belong to a species that does not include *c* and *d*. Perhaps *a* and *c* are more closely related to each other than either is to *b* and *d*. If so, it won't be true that *a* and *b* belong to a species that fails to include *c*.

As noted above, Hull argues for the thesis that species are individuals by emphasizing certain similarities between species and individual organisms. Organisms are born and die, and their cells interact in ways that affect their mutual survival; they have a common fate. In similar fashion, species originate in speciation events and go extinct, and their member organisms affect each other's chances for long-term reproductive success by virtue of belonging to a common gene pool. The disagreement between Mayr's species concept and various phylogenetic species concepts arises from a difference between species and many organisms. The cells of an organism interact with each other, thus affecting their future, and those same cells also trace back to an embryo. The cells have a special history unique to them and their current interactions means that their futures are bound together by a common fate. However, not all individuals are like this. Consider a fleet of ships. Their forming a single armada means that their futures are bound together, but the ships may have very different genealogies. Ships a and b may belong to one fleet while ships c and d belong to another, even though a and c were built in one shipyard while b and d were built in another. Grouping by current interactions can cross-cut grouping by genealogy. Are organisms to be grouped into species by their current reproductive interactions, or by their genealogies?

X Phylogenetic Inference

We do not directly observe that human beings are more closely related to mice than either is to snakes. Rather, we observe various characteristics that the three groups display; what we observe are patterns of similarity and difference. Human beings and mice are warm-blooded, but snakes are not. Mice and snakes do not have opposable

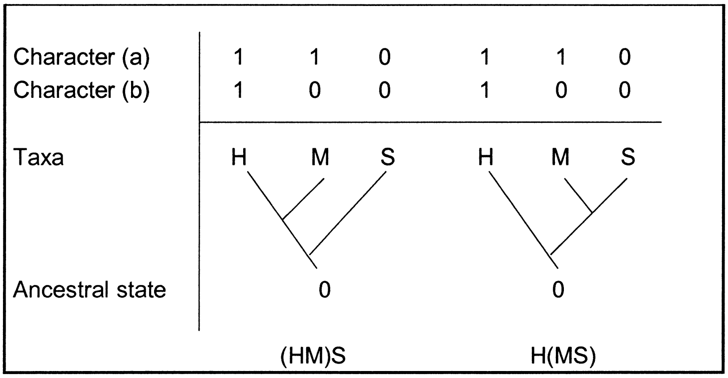


Figure P.1
Two kinds of similarity. (a) When H and M share a derived character (1) that S lacks, (HM)S is more parsimonious than H(MS); (b) when M and S share an ancestral character (O) that H lacks, (HM)S and H(MS) are equally parsimonious.

thumbs, but human beings do. How are we to use observations such as these to infer phylogenetic relationships?

In "The Logical Basis of Phylogenetic Analysis," James Farris argues for a hypothetico-deductive approach. We should choose genealogical hypotheses on the basis of their explanatory power, where the most explanatory hypothesis is the one that is most parsimonious. Farris endorses the cladistic idea that the parsimony of a phylogenetic hypothesis should be measured by counting the number of independent originations of features that the hypothesis requires to explain the data. The fact that human beings and mice are warm-blooded, while snakes are not, favors the (HM)S hypothesis over the H(MS) hypothesis. (HM)S means that humans and mice have a common ancestor that is not an ancestor of snakes. However, the fact that mice and snakes lack opposable thumbs, while human beings have them, is not evidence favoring H(MS) over (HM)S. The reason can be understood by considering figure P.1.

The two characters depicted in the figure are each dichotomous, with "0" in each case representing the state of the ancestor that humans, mice, and snakes share, and "1" representing the derived (= non-ancestral) character state. The distribution of character (a) across the three taxa has humans and mice sharing a derived character state (a "synapomorphy"). If a tree begins in state 0, how many changes must occur in the tree's interior to produce the distribution of character (a) found at the tips? The answer is that (HM)S requires that there be at least one change, while H(MS) requires at least two. Thus (HM)S is the more parsimonious explanation of character (a). Character (b) is different. Here the similarity that unites mice and snakes is ancestral (a "symplesiomorphy"). Notice that (HM)S and H(MS) can each explain the distribution of character

(b) by postulating that a single change has occurred; the two trees are equally parsimonious. This explains why cladistic parsimony involves a rejection of the idea that overall similarity is to be taken as evidence of phylogenetic relatedness; some similarities provide such evidence, but others do not. If you are inclined to think that the warm-bloodedness of humans and mice indicates that they are more closely related to each other than either is to snakes, you are agreeing with the dictates of cladistic parsimony. And if you also think that the absence of an opposable thumb in mice and snakes is not evidence that they are more closely related to each other than either is to humans, you are here agreeing with cladistic parsimony as well.

Whereas Farris thinks of phylogenetic inference from a hypothetico-deductive point of view, Joseph Felsenstein's approach is statistical. In his article ''Cases in which Parsimony and Compatibility Methods Will Be Positively Misleading,'' Felsenstein addresses a question about the property of statistical consistency. A method of inference is statistically consistent if it is bound to converge on the true hypothesis as the data set is made large without limit. Felsenstein assumes that an acceptable method of phylogenetic inference must be consistent and then constructs a simple example in which parsimony fails to converge on the truth. In his own article, Farris discusses Felsenstein's argument and points out that the assumptions Felsenstein makes about the evolutionary process are unrealistic; Farris concludes from this that Felsenstein's argument does not demonstrate any defect in the parsimony criterion. For himself, Felsenstein agrees that the assumptions he makes that allow him to derive the result about statistical inconsistency are unrealistic, however, he takes this to show that parsimony must make substantive assumptions about the evolutionary process. Parsimony is not a ''purely methodological'' inference criterion, but makes sense only to the extent that its implicit assumptions about the evolutionary process are correct.

One of the principal issues that divides Farris and Felsenstein concerns how much one must know about the evolutionary process before one can make inferences concerning genealogical relationships. Farris maintains that relatively modest assumptions suffice to justify the method of cladistic parsimony; Felsenstein holds that the assumptions required for one to use cladistic parsimony, or any other method, are far more substantive.

XI Race—Social Construction or Biological Reality?

Does group selection occur? Is adaptationism true? These questions, addressed in previous chapters, can be answered only after one clarifies what group selection is and what adaptationism asserts. The question ''do races exist?'' conforms to the same pattern.

How should the concept of race be defined? One possibility, discussed by Kwame Anthony Appiah in his chapter ''Why There Are no Human Races,'' is provided by the thesis of racialism, which is a version of essentialism (the subject of part VIII). This is the idea that we can "divide human beings into a small number of groups, called 'races,' in such a way that the members of these groups shared certain fundamental, heritable, physical, moral, intellectual, and cultural characteristics with each other that they did not share with members of any other race." Appiah argues that if this is what races are, then races do not exist. They are like phlogiston and witches—things thought at one time to exist, but which later scientific inquiry has discredited. When people talk of races, they are imposing on reality a set of categories of their own devising, rather than describing nature as it really is.

Racial essentialism is not the only way to answer the question of what a race is. Perhaps races can be conceptualized as biological subspecies. If so, the doubts that biologists have had about the reality of subspecies in general will be relevant to assessing the reality of human races in particular. And conversely, if subspecies are legitimate taxonomic categories, human races may be so as well. One way to approach this question is via the taxonomic philosophy of pheneticism. Pheneticists define taxa by forming similarity clusters. The members of a taxon need not all possess a set of characteristics that is unique to them (as the essentialist demands). Rather, they must be more similar to each other than they are to other individuals not in the taxon. If we cluster human beings in this way, will the resulting taxonomy approximate what human beings call races? If clustering yielded this result, phenetic races would exist, even if essentialist races do not. But what if the clusters are miles away from what people in a country (the United States, for example) call "races?"

Here we must attend to a distinction that Appiah discusses, and which also figures prominently in Robin Andreasen's chapter "A New Perspective on the Race Debate." If similarity groupings of human beings don't correspond much to what people ordinarily call races, then the pheneticist will conclude that common sense races do not exist. But this leaves open the possibility that phenetic races exist. It's just that people have often misconceived what races are. Here's an analogy: at one time people thought that whales are fish. Scientific taxonomy eventually rejected this idea, but did not conclude that fish is not a proper taxon. Rather, the conclusion was drawn that fish comprise a taxon whose membership differs from what people had thought. Even if common sense races do not exist, it is an open question whether scientific races exist.

Andreasen introduces a third way of clarifying the question of whether races exist, one that she thinks is superior to both essentialism and pheneticism. This is provided by cladistics. Cladism is a taxonomic philosophy according to which a scientifically legimate taxon is a monophyletic group. A monophyletic group is comprised of an ancestor and all its descendants. Birds constitute a real taxon, not because they are similar to each other, but because Aves is monophyletic. If a bird gave birth to an organism that lacked wings, that wingless organism would still be a bird. According to the cladistic point of view, if the human species contains distinct lineages that rarely if ever interbreed, those lineages comprise cladistic races. Andreasen argues that human races were at one time a biological reality, but more recently the reproductive separation of lineages started to break down. For Andreasen, races once existed but now they are on the way out.

XII Cultural Evolution

Do cultures evolve in a way that is structurally similar to the way that biological populations evolve? In the biological process of natural selection, organisms participate in a struggle for existence; evolution occurs because fitter organisms tend to outsurvive their less fit competitors. In a process of cultural selection, the ideas in a culture compete with each other in a marketplace, where fitter ideas are the ones that tend to attract more adherents. By this process, the mix of ideas in a culture changes. One difference between these cases concerns the mode of transmission. Biological traits are transmitted from parents to offspring by the passing along of genes, while ideas are transmitted from teachers to students by learning. But this point aside, the analogy seems clear-cut.

In their chapter "Does Culture Evolve?" Joseph Fracchia and Richard C. Lewontin pose a number of challenges for the project of modeling cultural change as an evolutionary process. At the most general level, they doubt that a single "transhistorical law or generality" can explain the dynamics of all cultural change and they also doubt that the evolutionary approach will be able to capture "the particularity, the uniqueness of all historical phenomena." They also have more specific objections. Fracchia and Lewontin note that recent evolutionary models of cultural change (as opposed to nineteenth-century progressivist theories of cultural development) depict culture in terms of the traits (ideas) of individuals. The state of a culture is given by the frequencies of different ideas in it. Fracchia and Lewontin object to this reductionist approach—a country's military power and its gross domestic product cannot be represented at the level of individuals, so the evolutionary approach will fail to capture how these social facts are causes and effects in social change. The problem of methodological individualism is a standard part of the philosophy of the social sciences, and here we see it making itself felt in the philosophy of biology.

In my chapter "Models of Cultural Evolution," I try to explain how selectional models of cultural evolution and selectional models of biological evolution are related to each other by focusing on how each understands the concepts of fitness and heredity. I discern three types of model, not just two, and then discuss the distinction between source laws and consequence laws that I developed in my book *The Nature of Selection*. Given a set of heritable traits that differ in fitness, a selectional model will be able to compute the consequences of those initial conditions by describing how the population is apt to change. However, such models are often silent on why the traits have the fitness values they do. It is one thing to describe the consequences of fitness differences, another thing to describe their sources. Models of cultural evolution have mainly focused on describing the consequences of fitness differences, while saying little about the sources. For example, the decline in birth rate that occurred in nineteenth-century Italy can be described as a process of cultural selection in which the trait of having fewer children had higher cultural (not biological) fitness than the trait of having more. There is nothing false about this description, but it does leave unanswered the question of why preferences about family size suddenly changed, or why women suddenly had the power to control their own reproduction in ways that were not possible before.

One important distinction to bear in mind in thinking about cultural evolution is the difference between the following two questions: Do cultural traits ever evolve by a process of cultural selection? How useful is this way of thinking about cultural change for answering the questions that historians and social scientists wish to address? The intelligibility of an approach and its fruitfulness are separate matters.

XIII Evolutionary Ethics

In *Sociobiology: The New Synthesis*, Edward O. Wilson suggests that "the time has come for ethics to be removed temporarily from the hands of the philosophers and biologicized." In "Moral Philosophy as Applied Science," Wilson and Michael Ruse set forth their reasons for thinking that evolutionary biology has a great deal to contribute to our understanding of morality. Philip Kitcher responds skeptically in his chapter "Four Ways of 'Biologicizing' Ethics."

A central distinction that is relevant to understanding this dispute is that between descriptive and normative statements. A descriptive statement describes what is the case without commenting on whether the facts that obtain are good or bad, just or unjust, and so on. A normative statement makes judges about whether something is good or bad, right or wrong, etc. These two sets of statements are sometimes called is-statements and ought-statements, but it would be a mistake to think that descriptive statements never contain the word "ought" and that normative statements never use the word "is." The statement "many people think that capital punishment ought to be abolished" is descriptive, not normative, while the statement "slavery is wrong" is normative, not descriptive.

When we ask whether evolutionary biology has anything useful to say about ethics, we need to divide this question in two. Does evolutionary biology help us understand why this or that normative statement is true or false? Does evolutionary biology help us understand why this or that descriptive statement about ethics is true or false? It is perfectly possible that evolutionary biology throws light on why human beings have the ethical beliefs and feelings they do, and yet says nothing about which of those ethical beliefs (if any) are true. Wilson and Ruse maintain that evolutionary considerations are relevant to both types of inquiry; it is important to bear in mind that separate arguments are required to establish that both types of relevance exist.

Acknowledgments

I am grateful to Mavis Biss, Holly Kantin, Greg Novack, Eric Saidel, Larry Shapiro, and Karen Strier for their help.

List of Contributors

Robin O. Andreasen
University of Delaware

Kwame Anthony Appiah
Princeton University

David A. Baum
University of Wisconsin-Madison

John H. Beatty
University of British Columbia

David J. Buller
Northern Illinois University

Leda Cosmides
University of California, Santa Barbara

Michael J. Donoghue
Yale University

James Farris
Swedish Museum of Natural History, Stockholm

Joseph Felsenstein
University of Washington

Susan K. Finsen (née Mills)
California State University, San Bernardino

Joseph Fracchia
University of Oregon, Eugene

Stephen Jay Gould*

Sarah Blaffer Hrdy
University of California, Davis

David L. Hull
Northwestern University

Philip Kitcher
Columbia University

Richard C. Lewontin
Harvard University

Elisabeth Lloyd
University of Indiana, Bloomington

John Maynard Smith*

Ernst Mayr*

Michael Ruse
Florida State University

Elliott Sober
University of Wisconsin–Madison

John Tooby
University of California, Santa Barbara

* = deceased

C. Kenneth Waters
University of Minnesota

George C. Williams
State University of New York, Stony Brook

David Sloan Wilson
Binghamton University

Edward O. Wilson
Harvard University

I Fitness

1 The Propensity Interpretation of Fitness

Susan K. Mills and John H. Beatty

The concept of "fitness" is a notion of central importance to evolutionary theory. Yet the interpretation of this concept and its role in explanations of evolutionary phenomena have remained obscure. We provide a propensity interpretation of fitness, which we argue captures the intended reference of this term as it is used by evolutionary theorists. Using the propensity interpretation of fitness, we provide a Hempelian reconstruction of explanations of evolutionary phenomena, and we show why charges of circularity which have been levelled against explanations in evolutionary theory are mistaken. Finally, we provide a definition of natural selection which follows from the propensity interpretation of fitness, and which handles all the types of selection discussed by biologists, thus improving on extant definitions.

The testability and logical status of evolutionary theory have been brought into question by numerous authors in recent years (e.g., Manser 1965, Smart 1963, Popper 1974). Many of the claims that evolutionary theory is not testable, that it parades tautologies in the guise of empirical claims, and that its explanations are circular, resulted from misunderstandings which have since been rebuked (e.g., by Ruse 1969, 1973, and Williams 1970, 1973a, 1973b). Yet despite the skilled rejoinders which have been given to most of these charges, the controversy continues to flourish, and has even found its way beyond philosophical and biological circles and into the pages of *Harpers Magazine*. In the spring of 1976, journalist Tom Bethell reported to the unsuspecting public that:

> Darwin's theory . . . is on the verge of collapse in his famous book, *On the Origin of Species* . . . Darwin made a mistake sufficiently serious to undermine his theory. The machinery of evolution that he supposedly discovered has been challenged, and it is beginning to look as though what he really discovered was nothing more than the Victorian propensity to believe in progress. (1976, p. 72)

Those familiar with the details of evolutionary theory, and with the history of this controversy, will rightfully feel no sympathy with such challenges, and may wonder

From *Philosophy of Science*, 1979, 46: 263–286.

whether it is worth bothering with them. But the fact is that there is a major problem in the foundations of evolutionary theory which remains unsolved, and which continues to give life to the debate. The definition of fitness remains in dispute, and the role of appeals to fitness in biologists' explanations is a mystery. This is a problem which ought to concern biologists and philosophers of science, quite independent of the vicissitudes of the controversy which it perpetuates.

Biologists agree on how to *measure* fitness, and they routinely appeal to fitness in their explanations, attributing the relative predominance of certain traits to the relative fitness of those traits. However, these explanations can and have been criticized on the grounds that, given the definitions of fitness offered by most biologists, these explanations are no more than redescriptions of the phenomena to be explained (e.g., Popper 1974, Manser 1965, Smart 1963). Philosophers have proposed new treatments of fitness designed to avoid these charges of explanatory circularity (e.g., Hull 1974 and Williams 1973a). Unfortunately, none of these interpretations succeeds in avoiding the charges, while providing a definition *useful* to evolutionary theory.

Thus it is high time that an analysis of fitness is provided which reveals the empirical content implicit in evolutionary biologists' explanations. To this end, we propose and defend the *propensity interpretation* of fitness. We argue that the propensity interpretation captures the intended reference of ''fitness'' as biologists use the term. Further, using this interpretation, we show how references to fitness play a crucial role in explanations in evolutionary theory, and we provide a Hempelian reconstruction of such explanations which reveals the precise nature of this role. We answer the charges of explanatory circularity leveled against evolutionary theory by showing how these charges arise from mistaken interpretations of fitness.

The concepts of fitness and natural selection are closely linked, since it is through the process of natural selection that the fittest gain predominance, according to the theory of evolution. Thus it is not surprising to find misinterpretations of fitness paralleled by misunderstandings of natural selection. The propensity analysis suggests a definition of ''selection'' which (unlike previously proposed definitions) accords with all the diverse types of selection dealt with by biologists.

But before proceeding with the positive analyses just promised, we consider the charge of explanatory circularity which arises from the lack of a satisfactory interpretation of fitness, and the reasons for the inadequacy of the replies so far offered in answer to the charge.

The Charge of Circularity

According to the most frequently cited definitions of ''fitness,'' that term refers to the *actual* number of offspring left by an individual or type relative to the actual contribution of some reference individual or type. For instance, Waddington (1968, p. 19) suggests that the fittest individuals are those which are ''most effective in leaving gametes to the next generation.'' According to Lerner (1958), ''the individuals who have more offspring are fitter in the Darwinian sense.'' Grant (1977, p. 66) construes fitness as ''a measure of reproductive success.'' And Crow and Kimura (1970, p. 5) regard fitness ''as a measure of both survival and reproduction'' (see also Dobzhansky 1970, pp. 101–102; Wilson 1975, p. 585; Mettler and Gregg 1969, p. 93).

These definitions of ''fitness'' in terms of actual survival and reproductive success are straightforward and initially intuitively satisfying. However, such definitions lead to justifiable charges that certain explanations invoking fitness differences are circular. The explanations in question are those which point to fitness differences between alternate types in a population in order to account for (1) differences in the average offspring contributions of those phenotypes, and (2) changes in the proportions of the types over times (i.e., evolutionary changes). Where fitness is defined in terms of survival and reproductive success, to say that type *A* is fitter than type *B* is just to say that type *A* is leaving a higher average number of offspring than type *B*. Clearly, we cannot say that the difference in fitness of *A* and *B* *explains* the difference in actual average offspring contribution of *A* and *B*, when fitness is defined in terms of actual reproductive success. Yet, evolutionary biologists seem to think that type frequency changes (i.e., evolutionary changes) can be *explained* by invoking the relative fitnesses of the types concerned. For instance, Kettlewell (1955, 1956) hypothesized that fitness differences were the cause of frequency changes of dark- and light-colored pepper moths in industrial areas of England. And he devised experiments to determine whether the frequency changes were correlated with fitness differences. Several philosophers have pointed to the apparent circularity involved in these explanations. Manser (1965) describes Kettlewell's account of the frequency differences in terms of fitness differences as ''only a description in slightly theory-laden terms which gives the illusion of an explanation in the full scientific sense'' (1965, p. 27).

The whole idea of setting up empirical investigations to determine whether fitness differences are correlated with actual descendant contribution differences seems absurd, given the above definitions of ''fitness.'' If this type of charge is coupled with the assumption that the only testable claims of evolutionary theory are of this variety, (i.e., tests of whether individuals identified as ''the fittest'' are most reproductively successful), then it appears that evolutionary theory is not testable. As Bethell puts it, ''If only there were some way of identifying the fittest before-hand, without always having to wait and see which ones survive, Darwin's theory would be testable rather than tautological'' (1976, p. 75).

However, as Ruse (1969) and Williams (1973a) have made clear, this latter charge is mistaken. Evolutionary theory embodies many testable claims. To take but one of many examples cited by Williams, Darwinian evolutionary theory predicted the existence of *transitional forms* intermediate between ancestral and descendant species. The saltationist (creationist) view of the origin of species which was accepted at the time when Darwin wrote on *The Origin of Species* predicted no such plethora of intermediate forms. Ruse has called attention to the predictions concerning distributions of types in populations which can be made on the basis of the Hardy–Weinberg law (1973, p. 36).

While these replies are well taken, they fail to clarify the role of fitness ascriptions in evolutionary theory. We agree with Williams and Ruse that evolutionary theory does make testable claims, and that many of these claims can be seen to be testable without providing an analysis of the role of fitness ascriptions. Nevertheless, some claims of evolutionary theory cannot be shown to be empirical without clarifying the role of ''fitness.'' Moreover, our understanding of other straightforwardly empirical claims of evolutionary theory will be enhanced by an explication of the role of ''fitness'' in these claims.

What Fitness Is Not

There are two questions to be clarified in defining fitness: What sorts of entities does this predicate apply to, and what does it predicate of these entities? Both these questions have received disparate answers from various biologists and philosophers. Fitness has been claimed to apply to types (e.g., Dobzhansky 1970, pp. 101–102; Crow and Kimura 1970) as well as individuals (Lerner 1958, Waddington 1968, p. 19). As will become apparent in the course of the positive analysis, the question of what sorts of entities ''fitness'' applies to should not be given a univocal answer. Fitness may be predicated of individual organisms, and (in a somewhat different sense) of phenotypes and genotypes. In this section we will only consider the question of what one is predicating of individuals and types in ascribing them a fitness value, according to the various proposals under scrutiny.

Before moving on to alternatives to the definition of ''fitness'' in terms of actual survival and reproductive success, we need to consider the acceptability of this definition, independent of the criticism that it leads to explanatory circularity. This criticism alone is obviously not sufficient to show that the interpretation is incorrect. For proponents of this definition can reply that fitness is actual reproductive success, since that is the way biologists use the term, and there is no other feasible definition. The fact that references to fitness lead to explanatory circularity just shows that fitness has no explanatory role to play in evolutionary theory. In fact, Bethell (1976, p. 75) makes this latter claim, and even maintains that biologists have abandoned references to fitness in their accounts of evolutionary phenomena. This is a scandalous claim.[1] A survey of evolutionary journals like *American Naturalist* and *Evolution* reveals that fitness ascriptions still play a major role in explanations of evolutionary phenomena. Indeed, the current literature on evolutionary theory reveals that the notion of fitness is of tremendous concern. Rather than abandoning the notion, modern evolutionary biologists have chosen to refine and extend it. Levins (1968) has raised the problem of fitness in changing environments. Thoday (1953) has pointed to the distinction between shortterm and long-term fitness. Analysis of and evidence for ''variable fitness'' or ''frequency dependent fitness'' was given by Kojima (1971). The effects of ''overdominance with regard to fitness'' on the maintenence of polymorphisms continue to be studied. And one very promising model of sociobiological evolution has been developed via an extension of traditional notions of fitness (the new notion is one of ''inclusive fitness'' [cf. Hamilton 1964]. As we will argue below, biologists are well advised *not* to abandon references to fitness, for such references play a crucial role in explanations of evolutionary phenomena.

Fortunately, we do have grounds quite independent of the issue of explanatory circularity for deeming inadequate definitions of ''fitness'' in terms of actual survival and reproductive success. For such definitions conflict with biologists' usage of the term, as is demonstrated by the following considerations. Surely two organisms which are genetically and phenotypically identical, and which inhabit the same environment, should be given the same fitness value. Yet where fitness is defined in terms of actual number of offspring left, two such organisms may receive radically different fitness values, if it happens that one of them succeeds in reproducing while the other does not. Scriven (1959) invites us to imagine a case in which two identical twins are standing together in the forest. As it happens, one of them is struck by lightning, and the other is spared. The latter goes on to reproduce while the former leaves no offspring. Surely in this case there is no difference between the two organisms which accounts for their difference in reproductive success. Yet, on the traditional definition of ''fitness,'' the lucky twin is *far* fitter. Most undesirably, such a definition commits us to calling the intuitively less fit of two organisms the fitter, if it happens that this organism leaves the greater number of offspring of the two.[2]

Nor can these counterintuitive results be avoided by shifting the reference of fitness from individual organisms to groups. For, precisely as was the case with individuals, the intuitively less fit subgroup of a population may by chance come to predominate. For example, an earthquake or forest fire may destroy individuals irrespective of any traits they possess. In such a case, we do not wish to be committed to attributing the highest fitness values to whichever subgroup is left.

Since an organism's traits are obviously important in determining its fitness, it is tempting to suggest that fitness be defined entirely independently of survival and reproduction, as some function of traits. Hull (1974) hints at the desirability of such a definition. This suggestion derives *prima facie* support from the fact that given such a definition, explanations of differential offspring contribution which appeal to differences in fitness are noncircular. However, no one has seriously proposed such a definition, and it is easy to see why. The features of organisms which contribute to their survival and reproductive success are endlessly varied and context dependent. What do the fittest germ, the fittest geranium, and the fittest chimpanzee have in common? It cannot be any concretely characterized physical property, given that one and the same physical trait can be helpful in one environment and harmful in another. This is not to say that it is impossible that some as yet unsuspected (no doubt abstractly characterized) feature of organisms may be found which correlates with reproductive success. Rather, it is just to say that we need not, and should not, wait for the discovery of such a feature in order to give the definition of "fitness."

So far, we have seen that we cannot define fitness simply in terms of survival and reproductive success. But we cannot define fitness entirely independently of any reference to survival and reproduction, either. An ingenious alternative to either of these approaches has been offered by Williams (1970, 1973a). She suggests that we regard "fitness" as a primitive term of evolutionary theory, and that we therefore refuse to define it. As she points out, in the formal axiomatization of a theory, it is not possible that all terms be explicitly defined, on pain of circularity. However, the fact that we cannot formally define all the terms of a theory *within* the framework of the theory does not prevent us from stepping outside the theory and explaining the meaning of the term in a broader linguistic framework.[3] Such an explication need not amount to anything as restrictive as an operational definition or an explicit definition making the term eliminable without loss from the theory. Rather, such an explication should allow us to understand what sort of property fitness is, its relation to natural selection, and the role of references to fitness in evolutionary theorists' explanations. Thus, our criticism of Williams is not that she is wrong about fitness but that she does not go far enough. We believe that a more thorough explication is possible, through the *propensity* interpretation of fitness.[4]

Propensity Analysis of Fitness

Levins (1968) has remarked that "fitness enters population biology as a vague heuristic notion, rich in metaphor but poor in precision." No doubt this is accurate as a characterization of the unclarity surrounding the role of fitness in evolutionary theory, even among biologists who use the term. But such unclarity is quite compatible with the fact that fitness plays an essential explanatory role in evolutionary theory. It is to the task of increasing the precision of the concept of fitness as well as making explicit this explanatory role that we now turn.

We have already seen that fitness is somehow connected with success at survival and reproduction, although it cannot be defined in terms of actual survival and reproductive success. Why have evolutionary biologists continued to confuse fitness with actual descendant contribution? We believe that the confusion involves a misidentification of the *post facto* survival and reproductive success of an organism with the *ability* of an organism to survive and reproduce. We believe that "fitness" refers to the ability.

Actual offspring contribution, on the other hand is a sometimes reliable—sometimes unreliable—indicator of that ability. In the hypothetical cases above, actual descendant contribution is clearly an unreliable indicator of descendant contribution capability. The identical twins are equally *capable* of leaving offspring. And the camouflaged butterfly is more *capable* of leaving offspring than is the noncamouflaged butterfly.

Thus, we suggest that fitness be regarded as a complex *dispositional* property of organisms. Roughly speaking, the fitness of an organism is its *propensity* to survive and reproduce in a particularly specified environment and population. A great deal more will have to be added before the substance of this interpretation becomes clear. But before launching into details, let us note a few general features of this proposal.

First, if we take fitness to be a dispositional property of organisms, we can immediately see how references to fitness can be explanatory.[5] The fitness of an organism explains its success at survival and reproduction in a particular environment in the same way that the solubility of a substance explains the fact that it has dissolved in a particular liquid. When we say that an entity has a propensity (disposition, tendency, capability) to behave in a particular way, we mean that certain physical properties of the entity determine, or are causally relevant to, the particular behavior whenever the entity is subjected to appropriate "triggering conditions." For instance, the propensity of salt to dissolve in water (the "water solubility" of salt) consists in (i.e., "water solubility" *refers to*) its ionic crystalline character, which causes salt to dissolve whenever the appropriate triggering condition—immersion in water—is met. Likewise, the fitness of an organism consists in its having traits which condition its production of offspring in a given environment. For instance, the dark coloration of pepper moths in sooted, industrial areas of England effectively camouflages the moths from predators, enabling them to survive longer and leave more offspring. Thus, melanism is one of many physical properties which constitute the fitness, or reproductive propensity, of pepper moths in polluted areas (in the same sense that the ionic crystalline character of salt constitutes its propensity to dissolve in water).

The appropriate triggering conditions for the realization of offspring contribution dispositions include particular environmental conditions. We do not say that melanic moths are equally fit in polluted and unpolluted environments any more than we claim that salt is as soluble in water as it is in mercury or swiss cheese.[6]

In addition to the triggering conditions which cause a disposition to be manifested, we must, in explaining or predicting the manifestation of a disposition, consider whether any factors other than the relevant triggering conditions were present to interfere with the manifestation. When we say that salt has dissolved in water because it is soluble in water, we assume the absence of disturbing factors, such as the salt's having been coated in plastic before immersion. Likewise, when we explain an organism's (or type's) offspring contribution by referring to its degree of fitness, we assume, for instance, that environmental catastrophes (e.g., atomic holocausts, forest fires, etc.) and human intervention have not interfered with the manifestation offspring contribution dispositions. In general, we want to rule out the occurrence of any environmental conditions which separate successful from unsuccessful reproducers without regard to physical differences between them.

Now let us fill in some of the details of this proposal. First, we must clarify the view of propensities we are presupposing. In our view, propensities are dispositions of *individual objects*. It is each hungry rat which has a tendency or propensity to move in the maze in a certain way, not the class of hungry rats. Classes—abstract objects, in general—do not have dispositions, tendencies, or propensities in any orthodox sense of the term.[7] This aspect of propensities in general is also a feature of the (unexplicated) notion of fitness employed by biologists. Evolutionary biologists often speak of fitness as if it were a phenotypic trait—i.e., a property of individuals. For instance, Wallace (1963, p. 633) remarks, "That instances of overdominance exist, especially in relation to a trait as complex as fitness, is generally conceded."

However, evolutionary biologists also employ a notion of fitness which refers to *types* (e.g., Dobzhansky 1970, pp. 101–102). Fitness cannot be a propensity in this case, although as we will argue, it is a derivative of individual fitness propensities. Thus, we will introduce two definitions of "fitness": Fitness$_1$ of individual organisms and fitness$_2$ of types.

Fitness$_1$: Fitness of Individual Organisms

A paradigm case of a propensity is a subatomic particle's propensity to decay in a certain period of time. Whether a particle decays during some time interval is a qualitative, nonrepeatable property of that particle's event history. It might initially be thought that "propensity to reproduce" is also a qualitative nonrepeatable property of an organism: either it reproduces during its lifetime, or it does not. However, the property of organisms which is of interest to the evolutionary biologist is not the organism's propensity to reproduce or not to reproduce, but rather the *quantity* of offspring which the organism has the propensity to contribute. For the evolutionary biologist is interested in explaining proportions of types in populations, and from this point of view, an organism which leaves one offspring is much more similar to an organism which leaves no offspring than it is to an organism which leaves 100 offspring. Thus, when we speak of "reproductive propensity," this should be understood as a quantitative propensity like that of a lump of radioactive material (considered as an individual) to emit particles over time, rather than as a "yes-no" propensity, like that of an individual particle to decay or not decay during some time interval.

It may have struck the reader that given this quantitative understanding of "propensity to reproduce," there are many such propensities. There is an organism's propensity to leave zero offspring, its propensity to leave 1 offspring, 2 offspring, . . . , *n* offspring (during its lifetime). Determinists might claim that there is a unique number of offspring which an organism is determined to leave (i.e., with propensity 1) in a given environment. For nondeterminists, however, things are more complicated. Organisms may have propensities of different strengths to leave various numbers of offspring. The standard dispositions philosophers talk about are tendencies of objects to instantiate certain properties invariably under appropriate circumstances. But besides such "deterministic" dispositions, there are the tendencies of objects to produce one or another of a distribution of outcomes with predetermined frequency. As Coffa (1977) argues, it seems just as legitimate to suppose there are such nondeterministic, "probabilistic" causes as to posit deterministic dispositions.[8]

If we could assume that there were a unique number of offspring which any organism is determined to produce (i.e., which the organism has propensity 1 to produce), then the fitness$_1$ of an organism could be valued simply as the number of offspring which that organism is disposed to produce. But since it is quite possible that organisms may have a range or distribution of reproductive propensities, as was suggested above, we derive fitness$_1$ values taking these various propensities into consideration.

Unfortunately, we also cannot simply choose the number of offspring which an organism has the *highest* propensity to leave—that is, the mode of the distribution. For in the first place, an organism may not have a *high* propensity to leave any particular number of offspring. In the second place, there may not be one number of offspring which corresponds to the mode of the distribution. For example, an organism might have a .5 propensity to leave 10 offspring and a .5 propensity to leave 20 offspring. And finally, even if there is a number of offspring which an organism has a significantly higher propensity to leave than any other number of offspring, we must take into account the remainder of the distribution of reproductive propensities as well. For example, an organism with a .7 propensity to leave 5 offspring and a .3 propensity to leave 50 offspring is very different from an organism with a .7 propensity to leave 5 offspring and a .3 propensity to leave no offspring, even though each has the propensity to leave 5 offspring as its highest reproductive propensity.[9]

In lieu of these considerations, one might suggest that the fitness$_1$ of an organism be valued in terms of the entire distribution of its reproductive propensities. The simplest way to do this is just to assign distributions as values. For example, the fitness$_1$ of an imaginary organism *x* might be the following distribution.

number of offspring	1	2	3	4	5	6	7	8	9	10
propensity	.05	.05	.05	.2	.3	.2	.05	.05	.05	

However, our intuitions fail us in regard to the comparison of such distributions. How can we determine whether one organism is fitter than another, on the basis of their distributions alone? For instance, is *x* fitter or less fit than *y* and *z*, whose distributions (below) differ from *x*'s?

number of offspring	1	2	3	4	5	6	7	8	9	10
y					1.0					
z			.5		.3					.2

In order to avoid the uncertainties inherent in this method of valuation, and still take into account all an organisms' reproductive propensities, we suggest that fitness$_1$ values reflect an organism's *expected number* of offspring. The expected value of an event is the weighted sum of the values of its possible outcomes, where the appropriate weights are the probabilities of the various outcomes. As regards fitness$_1$, the event in question is an individual's total offspring contribution. The possible outcomes $0_1, 0_2, \ldots, 0_n$ are contributions of different numbers of offspring. Values $(1, 2, \ldots, n)$ of the outcomes correspond to the number of offspring left. And the weighting probability for each outcome 0_i is just the organism's propensity to contribute i offspring. Thus the imaginary organisms x, y, and z above all have the same expected number of offspring, or fitness value, of 5.

We propose, then, that "individual fitness" or "fitness$_1$" be defined as follows:

The *fitness*$_1$ of an organism x in environment E equals $n =_{df} n$ is the expected number of descendants which x will leave in E.[10]

It may have occurred to the reader that the fitness values assigned to organisms are not literally propensity values, since they do not range from 0 to 1. But this does not militate against our saying that the fitness of an organism is a complex of its various reproductive propensities. Consider for comparison another dispositional property of organisms: their intelligence. If everyone could agree that a particular intelligence test really measured intelligence, then an organism's intelligence could be defined as the expected score on this test. (We would not value intelligence as the score actually obtained in a particular taking of the test, for reasons precisely analogous to those which militate against definitions of fitness in terms of actual numbers of organisms left. Intelligence is a competence or capacity of organisms, rather than simply a measure concept.) Obviously, intelligence would not be valued as the strength of the propensity to obtain a *particular* score. Similarly, it is the expected number of offspring which determines an organism's fitness values, not the strength of the propensity to leave a particular number of offspring.

Fitness$_2$: Fitness of Types

Having defined fitness$_1$, we are in a position to define the fitness$_2$ of types. As will become apparent in what follows, it is the fitness of types which figures primarily in explanations of microevolutionary change.

Intuitively, the fitness of a type (genotype of phenotype) reflects the contribution of a particular gene or trait to the expected descendant contribution (i.e., the fitness$_1$)

of possessors of the gene or trait. Differences in the contributions of alternate genes or traits would be easy to detect in populations of individuals which were phenotypically identical except in regard to the trait or gene in question. In reality, though, individuals differ with regard to many traits, so that the contribution of one or another trait to fitness$_1$ is not so straightforward. In fact, the notion of any simple, absolute contribution is quite meaningless. For a trait acts in conjunction with many other traits in influencing the survival and reproductive success of its possessors. Thus, its contribution to different organisms will depend upon the different traits it is associated with in those organisms.

Yet, in order to explain the evolution and/or persistence of a gene or its phenotypic manifestation in a temporally extended population, we would like to show that possessors of the gene or trait were generally better able to survive and reproduce than possessors of alternate traits or genes. (By ''alternate genes'' we mean alternate alleles, or alternate genes at the same locus of the chromosome. ''Alternate traits'' are phenotypic manifestations of alternate genes.) In other words, we want to invoke the *average* fitness$_1$ of the members of each of the types under consideration. Let us refer to average fitness$_1$ as ''fitness$_2$.'' Given some information about the fitness$_2$ of each of a set of alternate types in a population, and given some information about the mechanisms of inheritance involved, we can predict and explain the evolutionary fate of the genes or traits which correspond to the alternate types. For instance, if we knew that possessors of a homozygous-based trait were able to contribute a higher average number of offspring than possessors of any of the alternate traits present in the population, we would have good grounds for predicting the eventual predominance of the trait in the population.

As the above discussion suggests, we actually invoke *relative* fitness$_2$ values in predictions and explanations of the evolutionary fate of genes and traits. That is, we need to know whether members of a particular type have a *higher* or *lower* average fitness$_1$ in order to predict the fate of the type. In order to capture this notion, and to accommodate biologists' extensive references to ''relative fitness'' or ''Darwinian fitness,'' we introduce ''relative fitness$_2$.'' Given a set of specified alternate types, there will be a type which is fittest in the fitness$_2$ sense (i.e., has highest average fitness$_1$, designated ''Max Fitness$_2$''). Using this notion of Max Fitness$_2$, we define relative fitness$_2$ as follows:

The relative fitness$_2$ of type X in $E =_{df}$ the fitness$_2$ of X in E/Max fitness$_2$ in E

The role of relative fitness$_2$ ascriptions in evolutionary explanations has been acknowledged (for instance by Williams's ''condition 3'' in her analysis of functional explanations [1976]). Yet very little attention has been paid to the establishment of these ascriptions. Perhaps we should say a few words about these claims. For it might be supposed that the *only* way in which fitness$_2$ ascriptions can be derived is through measurements of actual average offspring contributions of types. If this were the case, even though ''fitness$_2$'' is not *defined* in terms of such measures (so that explanations employing fitness$_2$ ascriptions to explain actual offspring contribution differences would not be formally circular), claims concerning the influence of fitness$_2$ differences upon offspring contribution could not be *tested*. This would obviously be disastrous for our analysis.

Evolutionary biologists frequently *derive* relative fitness claims from optimality models (e.g., Cody 1966); this is basically an engineering design problem. It involves determining, solely on the basis of design considerations, which of a set of specified alternate phenotypes maximizes expected descendant contribution. The solution to such a problem is only optimal relative to the other *specified* alternatives (there may be an unspecified, more optimal solution). Thus, optimality models provide some insight into the relative fitness of members of alternate types.

The theorems derived from optimality models can be confirmed by measurements of actual descendant contribution. Such measures can also be used to generate fitness$_2$ ascriptions. Given evidence that descendant contribution was affected primarily or solely by individual propensities for descendant contribution, we can infer that descendant contribution measurements are indicative of individual or type fitness.

Explaining Microevolutionary Phenomena

Having elaborated the notions of fitness$_1$ and relative fitness$_2$, we hope to show how these concepts function in explanations of evolutionary phenomena. Perhaps the clearest means of showing this is to work through an example of such an explanation. The example we are going to consider involves a change in the proportion of the two alleles at a single chromosomal locus, and a change in the frequency of genotypes associated with this locus, in a large population of organisms. In this population, at the locus in question, there are two alleles, A and a. A is fully dominant over a, so that AA and Aa individuals are phenotypically indistinguishable with respect to the trait determined by this locus. This trait is the ''natural gun'' trait. All individuals which are either homozygous (AA) or heterozygous (Aa) at this locus have a natural gun, whereas the unfortunate individuals of genotype aa have no gun. Let us suppose that for many generations this population has lived in peace in an environment E, in which no ammunition is available. (Were the terminology not in question, we would say that there had been no ''selective pressure'' for or against the natural gun trait.) However, at generation n, environment E changes to environment E', by the introduction of ammunition usable by the individuals with natural guns. At generation n, the proportion of A alleles is .5 and the proportion of a alleles is .5, with the genotypes distributed as follows:

AA: .25 Aa: .50 aa: .25.

What we want to explain is that in generation $n+1$, the new frequency of genotypes is as follows:

AA: .29 Aa: .57 aa: .14.

Let us suppose that the large size of this population makes such a change in frequency extremely improbable ($p = .001$) on the basis of chance.

We need two pieces of information concerning this population in order to explain the change in frequency. We need to know (1) the relative fitness_2 of the natural gun and non-natural gun types, and (2) whether any conditions obtain which would interfere with the actualization of the descendant contribution propensities which the relative fitness_2 valuations reflect. As was noted above, the fact that an organism does not survive and reproduce in an environment in which periodic cataclysms occur is no indication of its fitness (any more than the failure of salt to dissolve in water when coated with plastic would count against its solubility).

The latter qualification, stating that no factors other than fitness_2 differences were responsible for descendant contribution, corresponds to the "extremal clause," which, as Coffa (1977, p. 194) has made clear, is a component in the specification of most scientific laws. Such clauses state that no physical properties or events relevant to the occurrence of the outcome described in the law (other than those specified in the initial conditions) are present to interfere with that outcome. In stating scientific laws, the assumption is often tacitly made that no such disturbing factors are present. But as Coffa has pointed out, it is important to make this assumption explicit in an extremal clause. For, no scientific law can be falsified by an instance in which the event predicted by the law fails to occur, unless the extremal clause is satisfied. Thus, our ability to fill in the details of the extremal clause will determine our ability to distinguish between contexts which count as genuine falsifications of a law and contexts which do not. The fact that evolutionary theorists are fairly specific about the types of conditions which interfere with selection is an indication in favor of the testability of claims about fitness. As noted above, the influence of fitness upon offspring contribution is disturbed by any factors which separate successful from unsuccessful reproducers without regard to physical differences between them. In addition, certain other evolutionary factors such as mutation, migration, and departures from panmixia may disturb the systematic influence of fitness differences between types upon proportions of those types in subsequent generations.

Let us suppose that we do know the relative fitnesses_2 of the natural gun and non-natural gun types, and let us suppose the natural selection conditions are present (i.e., nothing is interfering with the manifestation of the fitness propensities). This information together with the relevant laws of inheritance will allow us to predict (and explain) the frequencies of types in generation $n+1$. We need not detail the principles of inheritance which allow this computation here (since they are available in any genetics text) other than to note that the Hardy–Weinberg Law allows us to compute the relative frequencies of types in a population, given information about the heritability of the types in question together with hypotheses about fitness_2 differences.

In light of these considerations, we construct the promised schema as follows:

1. In E', in generation n, the distribution of genotypes is:

AA: .25 Aa: .50 aa: .25.

2. $(x)(AAx \supset tx)$ & $(x)(Aax \supset tx)$ & $(x)(aax \supset -tx)$
3. In E', the relative fitness_2 of type t is 1.0.
4. In E', the relative fitness_2 of type not-t is 0.5.
5. For any three distinct genotypes X, Y, Z (generated from a single locus), if the proportions of X, Y, Z in generation n are P, Q, and R, respectively, and if the relative fitnesses_2 of genotypes X, Y, and Z are $F(Y)$ and $F(Z)$, respectively, then the proportion of X in generation $n+1$ is:

$P \cdot F(X)/P \cdot F(X) + Q \cdot F(Y) + R \cdot F(Z)$.

6. $EC(E)$.
7. Given the size of population P, the probability that the obtained frequencies were due to chance is less than .001.

In E' at generation $n+1$ the frequency of genotypes is:

AA: .29 Aa: .57 aa: .14.

This explanation is of the inductive-statistical variety, with the strength of the connection between explanans and explanadum determined, as indicated in premise (7), by the size of the population. Premise (1) is, obviously, a statement of the initial conditions. Premise (2) allows us to determine which genotypes determine each phenotype: all individuals with genotype AA or Aa have trait t, and all individuals of genotype aa lack trait t. Premises (3) and (4) indicate the relative fitness_2 of types t and not-t in environment E. Premise (5) is the above-mentioned consequence of the Hardy–Weinberg Law which allows computation of the expected frequencies in generation $n+1$, given information about reproductive rates at generation n, together with information about initial frequences of individuals of each genotype at generation n. Premise (6) affirms that the extremal clause (EC) was satisfied—that is, that the "natural selection conditions" were present for the environment (E) in question. Thus we can infer that propensities to contribute descendants will be reflected in actual reproductive rates. Each genotype receives the relative fitness_2 associated with the phenotype it determines, as indicated in premise (2). Thus by substitution of the values provided in premises (3) and (4) in formula (5) (i.e., $X = AA$, $F(X) = 1.0$, $P = .25$; $Y = Aa$, $F(Y) = 1.0$, $Q = .50, \ldots,$ eta.) we can obtain the values which appear in the explanandum.

To summarize, from knowledge of (1) initial frequencies of genotypes in generation n, (2) the relative fitness$_2$ of those genotypes, and (3) the fact that the extremal clause was satisfied, we can infer what the frequencies of genotypes will be in generation $n + 1$.

Of course, in this admittedly artificial example, it was presumed that the appropriate relative fitness$_2$ values were known. This suggests that we somehow investigated reproductive *capabilities*, and not just reproductive differences. We must emphasize, however, that actual reproductive differences may be regarded as measures of differences in reproductive capability, as long *as the measured differences are statistically significant*. This is the means of fitness determination in many, if not most, evolutionary investigations. But this must not mislead the reader into identifying fitness with actual reproductive contribution. For *statistically significant* differences would not be required to establish fitness differences in this case. Rather, statistically significant differences are required to establish that certain variables (fitness differences, in this case) are causally connected with other variables (in this case, differences in offspring contribution). Statistically significant differences are thus quite appropriate measures for fitness differences, given the propensity interpretation of fitness.

Having explained the role of statistical significance in measuring fitness differences, we can now consider a more realistic example of the role of fitness in population biology. Certainly one of greatest controversies in the history of population genetics concerns the differences in fitness of heterozygotes and homozygotes. The importance of the controversy lies in the fact that if heterozygotes are generally fitter than homozygotes, then breeding groups will retain a greater amount of genetic variation then if homozygotes were generally superior in fitness. And the amount of variation present in a population is of considerable importance to the evolutionary fate of the population. (For instance, greater variation provides some "flexibility" in the sense that a genetically variable population has more alternatives for adapting to changing environmental conditions.) Theodosius Dobzhansky, a principal protagonist in this controversy, maintained that heterozygotes at many loci were fitter than homozygotes at the same loci, and he and his collaborators gathered a good deal of statistically significant data to support this contention.

For instance, in one article, it was reported that members of the species *Drosophila pseudoobscura* which were heterozygous in regard to the structure of their third chromosome were more viable than the flies which were homozygous. Dobzhansky et al. correlated viability differences (note that *viability* differences are dispositional property differences) with fitness differences, and they performed a statistical analysis on their data, in order to conclude that "heterosis [heterozygote superiority in fitness] has . . . developed during the experiment, as indicated by the attainment of equilibrium and by a study of the viability of the flies derived from the cage. Both tests gave statistically significant results" (1951, p. 263). Again, statistical significance would be of no concern if fitness were identified straightforwardly with offspring contribution. Statistical significance is important, however, if fitness is identified with phenotypic properties causally connected with offspring contribution.

As these examples demonstrate, fitness ascriptions play not only a legitimate, but a crucial role in explanations of evolutionary change. While biologists have not been able to justify their usage of the concept of "fitness," their usage of that concept has nevertheless been consistent and appropriate. Philosophers have accused biologists of giving circular explanations of evolutionary phenomena because they have only taken into account the definitions of fitness biologists explicitly cite, and they have not looked for the interpretation implicit in biologists' usage.

A Propensity Analysis of Natural Selection

One consequence of our propensity interpretation of fitness is that the analysis also points to an improved definition of "natural selection." As was noted earlier, the concepts of fitness and natural selection are inextricably bound—so much so that misinterpretations of fitness are reflected in misinterpretations of natural selection.

Thus, according to one of the more popular interpretations of natural selection, that process occurs whenever two or more individuals leave different numbers of offspring, or whenever two or more types leave different average numbers of offspring. For example, Crow and Kimura (1970) stipulate that "selection occurs when one genotype leaves a different number of progeny than another" (p. 173). Insofar as it is correct to say that the *fittest* are *selected*, this definition of "selection" clearly reflects a definition of "fitness" in terms of actual descendant contribution.

But surely these definitions (see also Wallace 1963, p. 160; Wilson 1975, p. 489) do not adequately delimit the reference of "natural selection." For evolutionary biologists do not refer to just any case of differential offspring contribution as "natural selection." For instance, if predatory birds were to kill light- and dark-colored moths indiscriminately, and yet by chance killed more light than dark ones, we would not attribute the differential offspring contribution of light and dark moths to natural selection. But if the dark coloration acted as camouflage, enabling the dark moths to escape predation and leave more offspring, we would attribute the resulting differential offspring contribution to the action of natural selection. For only in the latter case are differences in offspring contribution due to differences in offspring contribution dispositions.

Thus, Kettlewell (1955, 1956) did not presume to have demonstrated the occurrence of natural selection simply by pointing out the dramatic increase in frequency of dark-colored pepper moths within industrial areas of England. In order to demonstrate that selection (vs. chance fluctuations, migration, etc.) had accounted for the change, Kettlewell had to provide evidence that the dark-colored moths were better able to survive and reproduce in the sooted forests of these regions. Nor did Cain and Sheppard (1950, 1954) and Ford (1964) consider differential contribution to be a sufficient demonstration of natural selection in their celebrated accounts of the influence of selection on geographical distribution. In order to support the hypothesis that natural selection had affected the geographic distribution of various color and banding-pattern traits of snails of the species *Cepaea nemoralis*, these men argued that the colors and band-patterns peculiar to an area were correlated with the background color and uniformity of that area. More precisely, yellow snails were predominant in green areas; red and brown snails were predominant in beechwoods (''with their red litter and numerous exposures of blackish soil'' [Ford 1964, p. 153]); and unbanded shells were predominant in more uniform environments. These traits effectively camouflaged their possessors from the sight of predators (Ford 1964, p. 155), thus *enabling* suitably marked snails to contribute more offspring than the unsuitably marked snails.

In each of these cases, selection is construed as involving more than just differential perpetuation. Rather, selection involves differential perpetuation caused by differential reproductive capabilities. So, just as we amended traditional definitions of ''fitness'' to take into account descendant contribution propensities, we must also amend traditional definitions of ''selection'' so as to emphasize the role of differential descendant contribution propensities. Selection, properly speaking, involves not just the differential contribution of descendants, but a differential contribution *caused* by differential propensities to contribute. On the basis of these considerations, let us define ''individual selection'' and ''type selection'' as follows:

Natural selection is occurring in population P in environment E with regard to organisms x, y, z (members of P) $=_{df}$ x, y, z differ in their descendant contribution dispositions in E, and these differences are manifested in E in P.

Natural selection is occurring in population P in environment E with regard to types X, Y, Z (included in P) $=_{df}$ members of X, Y, Z types differ in their average descendant contribution dispositions in E, and these differences are manifested in E in P.

We know from our previous analysis that when organisms leave numbers of offspring which reflect their reproductive propensities (i.e., when reproductive propensities are manifesting themselves) in a particular environment, this implies that no factors are interfering with the manifestation of these propensities (cf. our remarks on extremal clauses above). Put more positively, we have grounds for believing that, for example, no cataclysms, cases of human intervention, and so forth are occurring. Of course, the occurrence of natural selection is not precluded by the incidence of such factors. Fitter individuals might leave more offspring than less fit individuals (on account of their fitness differences), even though non-discriminating factors are operating to minimize the reproductive effects of fitness differences. In other words, the incidence of non-discriminating factors will not necessarily override the effects of fitness differences. Thus, we do not have to rule out the occurrence of non-selective factors in our definition of ''natural selection.'' But in explanations (such as our Hempelian schema above) of the precise evolutionary effects of selection, we must take these non-selective factors into account.

Conclusion

A science may well progress even though its practitioners are unable to account for aspects of its foundations in any illuminating way. We believe that this has been the case with evolutionary theory, but that the *propensity* analysis of fitness which we have described captures the implicit content in biologists' usage of the term. The propensity interpretation allows us to reconstruct explanations of microevolutionary phenomena in such a way that these explanations appear to be entirely respectable and noncircular. By their form, and by inspection of the premises and conclusion, such explanations appear to satisfy Hempelian adequacy requirements for explanations, and even appear to incorporate recent modifications of the Hempelian model for inductive explanations (Coffa 1974). We chose an example of microevolutionary change, since we wanted the least complicated instance possible in order to illuminate the form of explanations utilizing fitness ascriptions. We know of no reason to believe that a similar reconstruction could not be given for the case of macroevolutionary change.[11]

Notes

We wish to thank Professor Michael Ruse, for initially drawing our attention to the problems of the logical status of evolutionary theory, and for insightful criticisms of an early draft of this chapter. We are heavily indebted to Alberto Coffa, for providing us with explications both of propensities and of the nature of explanation, and for innumerable criticisms and suggestions. Ron Giere also suggested that the propensity interpretation was a little more complex than we originally suspected. However, we claim complete originality for our mistakes.

1. Bethell may have been misled by the fact that evolutionary biologists recognize mechanisms of evolutionary change other than fitness differences (e.g., drift). Nevertheless, there is no question that fitness differences have been and still are considered effective in producing evolutionary changes.

2. The counter-intuitiveness of the traditional definition is also suggested by the following hypothetical case. Imagine two butterflies of the same species, which are phenotypically identical except that one (C) has color markings which camouflage it from its species' chief predator, while the second (N) does not have such markings and is hence more conspicuous. If N nevertheless happens to leave more offspring than C, we are committed on the definition of fitness under consideration to conclude that (1) both butterflies had the same degree of fitness before reaching maturity (i.e., zero fitness) and (2) in the end, N is fitter, since it left more offspring than C.

3. Gary Hardegree suggested this to us in conversation.

4. As we recently learned, Mary Williams supports the propensity interpretation and has, independently, worked toward an application of this interpretation.

5. Where fitness is defined as a propensity we can also squeeze the empirical content out of the phrase "survival of the fittest" (i.e., the claim that the fittest survive), which has frequently been claimed to be tautological (e.g., by Bethell 1976, Popper 1974, and Smart 1963. Just as the claim that "the soluble (substance) is dissolving" is an empirical claim, so the claim that those which could gain predominance in a particular environment are in fact gaining predominance, is an empirical claim. In short, to claim that a dispositional property is manifesting itself is to make an empirical claim. Such a claim suggests that the conditions usually known to trigger the manifestation are present, and no factors are present to override this manifestation. It seems plausible to interpret "the survival of the fittest" as a loose way of claiming that the organisms which are leaving most offspring are also the most fit. That this is a plausible interpretation of Darwin's use of the phrase is also suggested by Darwin's concern (in *The Origin of Species*) to demonstrate that conditions favoring natural selection are widely in effect. But it should be emphasized that nothing hinges on providing such an interpretation for "the survival of the fittest." This catch-phrase is not an important feature of evolutionary theory, in spite of the controversy its alleged tautological status has generated.

6. As this discussion suggests, an organism's fitness is not only a function of the organism's traits, but also of characteristics of the organism's environment. Actually, this function may be even more complicated. For evolutionary biologists have also noted that the fitness of an individual may depend upon the characteristics of the population to which it belongs. For instance, there is evidence of "frequency dependent selection" in several species of Drosophila (Kojima 1971). This kind of selection is said to occur whenever the fitness of a type depends upon the frequency of the type. Some types appear to be fitter, and are selected for, when they are rare. Thus, fitness is relative to environmental and population characteristics. And consequently, the appropriate triggering conditions for the realization of descendant contribution dispositions include environmental and population structure conditions.

7. Given propensities apply to individual objects, (rather than chance set-ups or sequencies of trials) we also take them to be ontologically real—not merely epistemic properties. Our view is similar to Mellor's (for a good review of the views on propensities, cf. Kyburg 1974), but it most closely follows Coffa's analysis (1977, and his unpublished dissertation, *Foundations of Inductive Explanation*).

8. While an organism has a number of different propensities to leave *n* offspring, for different values of *n*, we do not have the additional complication that an organism has a number of different propensities to leave a particular number of offspring, *n*. An object has many different *relative probabilities* to manifest a given property, depending on the reference class in which it is placed. (In practice, choice of reference classes is dependent on our knowledge of the statistically relevant features of the situation.) But an object's *propensity* to manifest a certain property is a function of all of the causally relevant features of the situation, independent of our knowledge or ignorance of these factors. The totality of causally relevant features detemines the unique correct reference

class, and thus the unique strength of the propensity to manifest the property in question. (Thus it cannot be the case that an object has more than one propensity to manifest a particular property in a particular situation.)

9. It might initially be thought that these examples are highly artificial, since there are no such ''bimodal'' organisms. But organisms tend to have offspring in litters and swarms. For such organisms, their offspring contribution propensities will cluster around multiples of numbers typical of the litter or hatching size.

10. A note of clarification is in order concerning our definition of ''fitness$_1$.'' It is not clear whether ''expected descendant contribution'' refers to expected offspring contribution, or expected second-generation descendant contribution, or expected 100th generation descendant contribution. The problem can be illustrated as follows. One kind of individual may contribute a large number of offspring which are all very well adapted to the environment into which they are born, but cannot adapt to environmental changes. As a result, an individual of this type contributes a large number of offspring at times t, but due to an environmental change at $t + \Delta t$, these offspring in turn leave very few offspring, so that the original individual actually has very few second- or third-generation descendants. On the other hand, individuals of an alternate type may leave fewer offspring, yet these offspring may be very adaptable to environmental changes. Thus, although an individual of the latter type contributes a lower average number of offspring at time t, that individual may have a greater descendant contribution at $t + \Delta t$. Which individual is fitter? We suggest differentiating between long-term fitness and short-term fitness—or between first-generation fitness, second-generation fitness, $\ldots$, n-generation fitness. Thus, the latter type is fitter in the long term, while the former is fitter in the short term.

11. A great deal more needs to be done by way of clarifying the concepts of fitness and natural selection, given the many uses biologists make of these concepts. But we believe that the broad analyses we have given provide an adequate framework within which further distinctions and clarifications can be made. For example, within the categories of fitness$_1$ and relative fitness$_2$, distinctions can be drawn between short- and long-term fitness, by distinguishing between propensities to leave descendants in the short run (in the next few generations) *vs*. propensities to leave descendants in the long run (cf. note 10).

The propensity interpretation also lends itself to the much-discussed notion of ''frequency dependent fitness,'' wherein the fitness of a type differs according to the frequency of the type. Certain cases of mimicry have been explained via reference to frequency dependent fitness. For instance, it has been suggested that the mimetic resemblance of a prey species to a distasteful model may enhance the survival of the mimics so long as they are rare, because individual predators most readily learn to avoid the distasteful type (and hence the mimic) when the model is more common than the mimic. Surely the survival *ability* of the mimics, and not just their survival rates, are enhanced by the scarcity of their type.

The sociobiological notion of ''inclusive fitness'' also seems susceptible to a propensity analysis. Biologists have invoked this notion in order to explain the evolution of certain altruistic traits. The idea (very simply) is that some of the organisms benefiting from an altruistic action may be genetically related to the altruistic actor, and may therefore share the behavioral trait which led to the action (if the trait is genetically based). Thus, although an altruistic action may decrease the

fitness_1 of the actor, it may increase the fitness_2 of the altruistic trait. As a result, the trait may come to predominate within the population. "Inclusive fitness" values have been proposed as appropriate indicators of the evolutionary fate of altruistic traits. These values take into account not only the effect of altruistic actions upon the fitness of the actors, but also the probability that the action will benefit genetic relatives, and the extent of the benefit to relatives (cf. Hamilton 1964). Our colleague Greg Robischon is currently considering a propensity interpretation of inclusive fitness.

References

Bethell, T. (1976). "Darwin's Mistake." *Harper's Magazine*, 70–75.

Cain, A. J., and P. M. Sheppard (1950). "Selection in the Polymorphic Land Snail *Cepala Nemoralis*." vol. 4, *Heredity*: 275–294.

Cain, A. J., and P. M. Sheppard (1954). "Natural Selection in Cepaea." *Genetics* 39: 89–116.

Cody, M. (1966). "A General Theory of Clutch Size." *Evolution*, 20: 174–184.

Coffa, J. A. (1974). "Hempel's Ambiguity." *Synthese*, 28: 141–163.

———. (1977). "Probabilities: Reasonable or True?" *Philosophy of Science*, 43: 186–198.

Crow, J. F., and Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York: Harper and Row.

Dobzhansky, T. (1970). *Genetics of the Evolutionary Process*. New York: Columbia University Press.

Dobzhansky, T., and Levene, H. (1951). "Development of Heterosis Through Natural Selection in Experimental Populations of Drosophila Pseudoobscura." *American Naturalist* 85: 246–264.

Ford, E. B. (1964). *Ecological Genetics*. New York: Wiley.

Grant, V. (1977). *Organismic Evolution*. San Francisco: W. H. Freeman.

Hamilton, W. D. (1964). "The Genetical Evolution of Social Behavior. I." *Journal of Theoretical Biology*, 7: 1–16.

Hull, D. (1974). *Philosophy of Biological Theory*. Englewood Cliffs. New Jersey: Prentice-Hall.

Kettlewell, H. B. D. (1955). "Selection Experiments on Industrial Melanism in the Lepidoptera." *Heredity*, 9: 323–342.

———. (1956). "Further Selection Experiments on Industrial Melanism in the Lepidoptera." *Heredity*, 10: 287–301.

Kojima, K. (1971). "Is There a Constant Fitness Value for a Given Genotype?" *Evolution*, 25: 281–285.

Kyburg, H. (1974). "Propensities and Probabilities." *British Journal for the Philosophy of Science*, 25, no. 4: 358–374.

Lerner, I. M. (1958). *The Genetic Basis of Selection*. New York: Wiley.

Levins, R. (1970). "Fitness and Optimization." *Mathematical Topics in Population Genetics*. New York: Springer Verlag.

———. (1968). *Evolution in Changing Environments*. Princeton: Princeton University Press.

Manser, A. R. (1965). "The Concept of Evolution." *Philosophy*, 40: 18–34.

Mettler, L. E., and Gregg, T. G. (1969). *Population Genetics and Evolution*. Englewood Cliffs, New Jersey: Prentice-Hall.

Popper, K. (1974). "Intellectual Autobiography." *The Philosophy of Karl Popper* (Shilpp, ed.). LaSalle, Illinois: Open Court.

Ruse, M. (1969). "Confirmation and Falsification of Theories of Evolution." *Scientia*, 1–29.

———. (1973). *The Philosophy of Biology*. London: Hutchinson.

Scriven, M. (1959). "Explanation and Prediction in Evolutionary Theory." *Science*, 130: 477–482.

Smart, J. J. C. (1963). *Philosophy and Scientific Realism*. London: Routledge and Kegan Paul.

Thoday, J. M. (1953). "Components of Fitness." *Symposia of the Society for Experimental Biology*, 7: 96–113.

Waddington, C. H. (1968). "The Basic Ideas of Biology." *Towards a Theoretical Biology*, vol. 1. Chicago: Aldine.

Wallace, B. (1968). *Topics in Population Genetics*. New York: W. W. Norton.

———. (1963). "Further Data on the Over-dominance of Induced Mutations." *Genetics* 48: 633–651.

Williams, M. B. (1970). "Deducing the Consequences of Evolution: A Mathematical Model." *Journal of Theoretical Biology*, 29: 343–385.

———. (1973a). "The Logical Status of Natural Selection and Other Evolutionary Controversies: Resolution by Axiomatization." *Methodological Unity of Science* (Bunge, ed.). Dordrecht, Holland: Reidel.

———. (1973b). "Falsifiable Predictions of Evolutionary Theory." *Philosophy of Science*, 40: 518–537.

———. (1976). "The Logical Structure of Functional Explanations in Biology." *Proceedings of the Philosophy of Science Association 1976*: East Lansing: Philosophy of Science Association, 37–46.

Wilson, E. O. (1975). *Sociobiology*. Cambridge, Massachusetts: Harvard University Press.

Wright, S. (1955). "Classification of the Factors of Evolution." *Cold Spring Harbor Symposia on Quantitative Biology*, 20: 16–24.

2 The Two Faces of Fitness

Elliott Sober

The concept of fitness began its career in biology long before evolutionary theory was mathematized. Fitness was used to describe an organism's vigor, or the degree to which organisms "fit" into their environments. An organism's success in avoiding predators and in building a nest obviously contributes to its fitness and to the fitness of its offspring, but the peacock's gaudy tail seemed to be in an entirely different line of work. Fitness, as a term in ordinary language (as in "physical fitness") and in its original biological meaning, applied to the survival of an organism and its offspring, not to sheer reproductive output (Cronin 1991, Paul 1992). Darwin's separation of natural from sexual selection may sound odd from a modern perspective, but it made sense from this earlier point of view.

Biologists came to see that this limit on the concept of fitness is theoretically unjustified. Fitness is relevant to evolution because of the process of natural selection. Selection has an impact on the traits that determine how likely it is for an organism to survive from the egg stage to adulthood, but it equally has an impact on the traits that determine how successful an adult organism is likely to be in having offspring. Success concerns not just the robustness of offspring but their number. As a result, we now regard viability and fertility as two components of fitness. If p is the probability that an organism at the egg stage will reach adulthood, and e is the expected number of offspring that the adult organism will have, then the organism's *overall fitness* is the product pe, which is itself a mathematical expectation. Thus, a trait that enhances an organism's viability but renders it sterile has an overall fitness of zero. And a trait that slightly reduces viability, while dramatically augmenting fertility, may be very fit overall.

The expansion of the concept of fitness to encompass both viability and fertility resulted from the interaction of two roles that the concept of fitness plays in evolutionary theory. It describes the relationship of an organism to its environment. It also has a

From *Thinking About Evolution: Historical, Philosophical, and Political Perspectives*, ed. Rama S. Singh, Costas B. Krimbas, Diane B. Paul, and John Beatty, Volume 2 (Cambridge University Press, 2001), pp. 309–321. Reprinted by permission of Cambridge University Press.

mathematical representation that allows predictions and explanations to be formulated. *Fitness is both an ecological descriptor and a mathematical predictor*. The descriptive ecological content of the concept was widened to bring it into correspondence with the role that fitness increasingly played as a mathematical parameter in the theory of natural selection.

In this chapter I want to discuss several challenges that have arisen in connection with idea that fitness should be defined as expected number of offspring. Most of them are discussed in an interesting article by Beatty and Finsen (1989). Ten years earlier, they had championed a view they dubbed ''the propensity interpretation of fitness'' (Mills and Beatty 1979; see also Brandon 1978). In the more recent article, they ''turn critics.'' Should fitness be defined in terms of a one-generation time frame—why focus on expected number of *offspring* rather than *grand*offspring, or more distant descendants still? And is the concept of mathematical expectation the right one to use? The details of my answers to these questions differ in some respects from those suggested by Beatty and Finsen, but my bottom line will be the same—expected number of offspring is not always the right way to define fitness.

In what follows, I will talk about an *organism*'s fitness even though evolutionary theory shows scant interest in individual organisms but prefers to talk about the fitness values of *traits* (Sober 1984). Charlie the Tuna is not a particularly interesting object of study, but tuna dorsal fins are. Still, for the theory of natural selection to apply to the concrete lives of individual organisms, it is essential that the fitness values assigned to traits have implications concerning the reproductive prospects of the individuals that have those traits. How are trait fitnesses and individual fitnesses connected? Because individuals that share one trait may differ with respect to others, it would be unreasonable to demand that individuals that share a trait have identical fitness values. Rather, the customary connection is that the fitness value of a trait is the average of the fitness values of the individuals that have the trait. For this reason, my talk in what follows about the fitness of organisms will be a harmless stylistic convenience.

To begin, let us remind ourselves of what the idea of a mathematical expectation means. An organism's expected number of offspring is not necessarily the number of offspring one expects the organism to have. For example, suppose an organism has the following probabilities of having different numbers of offspring:

number (i) of offspring	0	1	2	3
probability (p_i) of having exactly i offspring	0.5	0.25	0.125	0.125

The expected number of offspring is $\sum ip_i = 0(0.5) + 1(0.25) + 2(0.125) + 3(0.125) = 0.875$, but we do not expect the organism to have precisely 7/8ths of an offspring. Rather, ''expectation'' means *mathematical* expectation, a technical term; the expected value is, roughly, the (arithmetic) average number that the individual would have if it got to live its life again and again in identical circumstances. This is less weird than it sounds; a fair coin has 3.5 as the expected number of times it will land heads if it is tossed 7 times.

In this example, the expected number of offspring will not exactly predict an individual's reproductive output, but it will probably come pretty close. However, there are cases in which the expected value provides a very misleading picture as to what one should expect. Lewontin and Cohen (1969) develop this idea in connection with models of population growth. Suppose, to use one of their examples, that each year a population has a probability of 0.9 of having a growth rate of 1.1 and a probability of 0.1 of having a growth rate of 0.3. The expected (arithmetic mean) growth rate per year is $(0.9)(1.1) + (0.1)(0.3) = 1.02$; thus, the expected size of the population increases by 2% per year. At the end of a long stretch of time, the population's expected size will be much larger than its initial size. However, the fact of the matter is that the population is virtually certain to go extinct in the long run. This can be seen by computing the geometric mean growth rate. The geometric mean of n numbers is the nth root of their product; because $[(1.1)^9(0.3)]^{1/10}$ is less than unity, we expect the population to go extinct. To see what is going on here, imagine a very large number of populations that each obey the specified pattern of growth. If we follow this ensemble for, say, 1000 years, what we will find is that almost all of the populations will go extinct, but a very small number will become huge; averaging over these end results, we will obtain the result that, on average, populations grow by 2% a year. Lewontin and Cohen point out that this anomaly is characteristic of multiplicative processes.

A simpler and more extreme example that illustrates the same point is a population that begins with a census size of 10 individuals and each year has a 0.5 chance of tripling in size and a 0.5 chance of going extinct. After 3 years, the probability is 7/8 that the population has gone extinct, but there is a probability of 1/8 that the population has achieved a census size of $(3)(3)(3)10 = 270$. The expected size of the population is $(7/8)(0) + (1/8)(270) = 33.75$. This expected size can be computed by taking the expected yearly growth rate of $(0.5)(3) + (0.5)0 = 1.5$ and raising it to the third power; $(1.5)(1.5)(1.5)10 = 33.75$. In expectation, the population increases by 50% per year, but you should expect the population to go extinct.

Probabilists will see in this phenomenon an analogue of the St. Petersburg paradox (Jeffrey 1983). Suppose you are offered a wager in which you toss a coin repeatedly until tails appears, at which point the game is over. You will receive 2^n dollars, where n is the number of tosses it takes for tails to appear. If the coin is fair, the expected payoff of the wager is

$$(1/2)\$2 + (1/4)\$4 + (1/8)\$8 + \cdots$$

The expected value of this wager is infinite, but very few people would spend more than, say, \$10 to buy into it. If rationality means maximizing expected utility, then people seem to be irrational—they allegedly should be prepared to pay a zillion dollars for such a golden opportunity. Regardless of whether this normative point is correct, I suspect that people may be focusing on what will *probably* happen, not on what the average payoff is over all possible outcomes, no matter how improbable. Notice that the probability is only 1/8 that the game will last more than three rounds. What we expect to be paid in this game deviates enormously from the expected payoff.

For both ecologists and gamblers, the same advice is relevant: Caveat emptor! If you want to make predictions about the outcome of a probabilistic process, think carefully before you settle on expected value as the quantity you will compute.

2.1 The Long-Term and the Short-Term

The definition of fitness as expected number of offspring has a one-generation time scale. Why think of fitness in this way rather than as having a longer time horizon? Consider figure 2.1 adapted from Beatty and Finsen (1989). Trait A produces more offspring than trait B (in expectation) before time t^*; however, after t^*, A produces fewer offspring than B, and in fact A eventually produces zero offspring. The puzzle is that A seems to be fitter than B in the short term, whereas B seems to be fitter than A in the long term. Which of these descriptions is correct?

The issue of whether fitness should be defined as a short-term or a long-term quantity will be familiar to biologists from the work of Thoday (1953, 1958), who argued that fitness should be defined as the probability of leaving descendants in the very long run; he suggests 10^8 years as an appropriate time scale. Thoday (1958, p. 317) says that a long-term measure is needed to obtain a definition of evolutionary progress. This reason for requiring a long-term concept will not appeal to those who think that progress is not a scientific concept at all (see, for example, discussion in Nitecki 1988

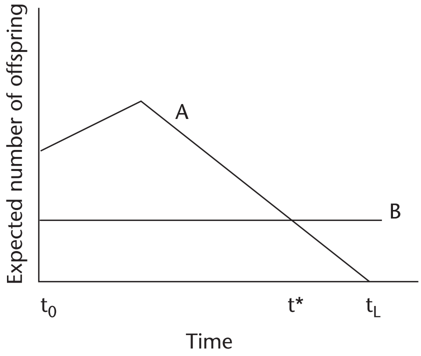


Figure 2.1

Trait A is fitter than Trait B initially, but later on the reverse is true. This means that B has a higher long-term fitness than A.

and Sober 1994). Thoday's argument also has the drawback that it repeatedly adverts to the good of the species without recognizing that this may conflict with what is good for individual organisms.

Setting aside Thoday's reason for wanting a log-term concept of fitness, does this concept make sense? Brandon (1990, pp. 24–5) criticizes Thoday's approach and the similar approach of Cooper (1984) on the grounds that selection "proceeds through generational time" and "has no foresight." I think both these criticisms miss the mark. Long-term probabilities imply foresight no more than short-term probabilities do. And the fact that selection occurs one generation at a time does not mean that it is wrong to define a quantity that describes a trait's long-term expected fate. Brandon also faults Thoday's proposal for failing to be operational. How are we to estimate the probability that a present organism or species will have descendants in the distant future? The point is well taken when the inference is *prospective*; in this case, the short-term is more knowable than the long-term future. However, when we make *retrospective* inferences, the situation reverses. An inferred phylogeny may reveal that a derived character displaced an ancestral character in one or more lineages. This information may provide evidence for the claim that the derived trait had the higher long-term fitness. In contrast, the one-generation fitnesses that obtained 60 million years ago may be quite beyond our ken.

Rather than rejecting a long-term concept of fitness and defending a short-term measure, I suggest that there is frequently no need to choose. In the accompanying figure, the y values for A and B at a given time tell us which trait had the higher short-term fitness at that time. The long-term fitness of a trait—its fitness, say, from t_0 to t^* or from t_0 to t_L—is a statistic that summarizes the relevant short-term values. There is no paradox in the fact that A has the higher short-term fitness whereas B has the higher long-term fitness. The same pattern can be found in two babies. The first has the higher probability of reaching age 20, whereas the second has the higher probability of surviving to age 60. The probability of a baby's reaching age 60 is a product—*Pr* (surviving to age 20 | you are a baby) *Pr* (surviving to age 60 | you have survived to age 20) $= (s_1)(s_2)$. The first baby may have a higher value on s_1 than the second, whereas the second has a higher value on s_2 than the first; overall, the first baby's product may be lower than that of the second. Long-term fitness is a coherent concept that may be useful in the context of certain problems; however, its coherence and desirability do not undermine the concept of short-term fitness.

2.2 When a One-Generation Time Frame Is Inadequate

The concept of short-term fitness discussed so far has a one-generation time frame—an organism at the egg stage has a probability p of reaching reproductive age and, once it is an adult, it has e as its expected number of offspring—the product pe is its overall fitness. However, a one-generation time frame will not always be satisfactory for the concept of short-term fitness. Fisher's (1930) model of sex ratio shows why (Sober 1984). If, in expectation, one female has 5 sons and 5 daughters whereas another produces 10 daughters and 0 sons, how can their different sex-ratio strategies make a difference in their fitnesses? Fisher saw that the answer is invisible if we think one generation ahead but falls into place if we consider two. The sex ratio exhibited by a female's progeny influences how many *grand*offspring she will have.

Other examples may be constructed of the same type. Parental care is a familiar biological phenomenon, but let us consider its extension—care of grandoffspring. If A individuals care for their grandoffspring, but B individuals do not, it may turn out that A individuals are fitter. However, the advantage of A over B surfaces only if we consider the expected numbers of grandoffspring that survive to adulthood. This example may be more of a logical possibility than a biological reality; still, it and sex ratio illustrate the same point. In principle, there is no a priori limit on the size of the time frame over which the concept of fitness may have to be stretched. If what an organism does in its lifetime affects the life prospects of organisms in succeeding generations, the concept of fitness may have to encompass those far-reaching effects.

2.3 Stochastic Variation in Offspring Number

Let us leave the question of short-term versus long-term behind and turn now to the question of whether fitness should be defined as a mathematical expectation. This is not an adequate definition when there is stochastic variation in viability or fertility. Dempster (1955), Haldane and Jayakar (1963), and Gillespie (1973, 1974, 1977) consider stochastic variation among generations; Gillespie (1974, 1977) addresses the issue of within-generation variation. These cases turn out to have different mathematical consequences for how fitness should be defined. However, in both of them, selection favors traits that have lower variances. In what follows, I will not attempt to reproduce the arguments these authors give for drawing this conclusion. Rather, I will describe two simple examples that exhibit the relevant qualitative features.

Let us begin with the case of stochastic variation among generations. Suppose a population begins with two A individuals and two B's. A individuals always have two offspring, whereas the B individuals in a given generation all have one offspring or all have three, with equal probability. Notice that the expected (arithmetic average) offspring number is the same for both traits—2. However, we will see that the expected frequency of B declines in the next generation.

Assume that these individuals reproduce asexually and then die and that offspring always resemble their parents. Given the numbers just described, there will be four A individuals in the next generation and either two B individuals or six, with equal probability. Although the two traits begin with the same population frequency and have the same expected number of offspring, their expected frequencies in the next generation differ:

Expected frequency of A $= (1/2)(4/6 + 4/10) = 0.535$

Expected frequency of C $= (1/2)(2/6 + 6/10) = 0.465$

The trait with the lower variance can be expected to increase in frequency. The appropriate measure for fitness in this case is the geometric mean of offspring number averaged over time; this is the same as the expected log of the number of offspring. Trait B has the lower geometric mean because $[(3)(1)]^{1/2} = 1.7 < [(2)(2)]^{1/2} = 2$. The geometric mean is approximately the arithmetic expected number minus $\sigma^2/2$.

Let us now consider the case of within-generation variance in offspring number. Gillespie (1974) describes the example of a bird whose nest has a probability of escaping predators of about 0.1. Should this bird put all its eggs in one nest or establish separate nests? If the bird lays 10 eggs in just one nest, it has a probability of 0.9 of having 0 offspring and a probability of 0.1 of having 10. Alternatively, if the bird creates 2 nests containing 5 eggs each, it has a probability of $(0.9)^2$ of having 0 offspring, a probability of $2(0.9)(0.1)$ of having 5, and a probability of $(0.1)^2$ of having 10. The expected value is the same in both cases—1.0 offspring—but the strategy of putting all eggs in one nest has the higher variance in outcomes. This example illustrates the idea of within-generation variance because two individuals in the same generation who follow the same strategy may have different numbers of offspring.

Does the process of natural selection vindicate the maxim that there is a disadvantage in putting all one's eggs in one basket? The answer is yes. To see why, let us examine a population that begins with two A individuals and two C's. A individuals always have two offspring, whereas each C individual has a 50% chance of having 1 offspring and a 50% chance of having 3. Here C individuals in the same generation may vary in fitness, but the expected value in one generation is the same as in any other. In the next generation, there will be four A individuals. There are four equiprobable arrangements of fitnesses for the two C individuals, and thus there are four equiprobable answers to the question of how many C individuals there will be in the next generation—two, four, four, and six. The expected number of C individuals in the next generation is four, but the expected frequencies of the two traits change:

Expected frequency of A $= (1/4)(4/6 + 4/8 + 4/8 + 4/10) = 0.52$

Expected frequency of C $= (1/4)(2/6 + 4/8 + 4/8 + 6/10) = 0.48$

Once again, the trait with the lower variance can be expected to increase in frequency.

In this example, the population grows from four individuals in the first generation to somewhere between 6 and 10 individuals in the second. Suppose we require that population size remain constant; after the four parents reproduce, random sampling reduces the offspring generation to four individuals. When this occurs, the trait with the higher variance has the higher probability of going extinct.

Gillespie (1974, 1977) constructed a model to describe the effect of within-generation variance. A trait's variance (σ^2) influences what happens only when population size (N) is finite; in the infinite limit, variance plays no role. On the basis of this model, Gillespie says that a trait's fitness is approximately its arithmetic mean number of offspring minus the quantity σ^2/N. Notice that this correction factor will be smaller than the one required for between-generation variance if $N > 2$.

Why, in the case of within-generation variance, does the number of individuals (N) in the whole *population* appear in the expression that describes the fitness of a single *trait*, which may be one of many traits represented in the population? In our example, why does the fitness of C depend on the total number of C and A individuals? And why does the effect of selecting for lower variance decline as population size increases? The reasons can be glimpsed in the simple calculation just described. To figure out the expected frequency of C, we summed over the four possible configurations that the population has in the next generation. There is a considerable difference among these four possibilities—trait C's absolute frequency is either 2/6, 4/8, 4/8, or 6/10. In contrast, if there were 2 C parents but 100 A's, there still would be four fractions to consider, but their values would be 2/202, 4/204, 4/204, and 6/206; these differ among themselves much less than the four that pertain to the case of 2 A's and 2 C's. The same diminution occurs if we increase the number of C parents; there would then be a larger number of possible configurations of the next generation to consider, and these would differ among themselves less than the four described initially. In the limit, if the population were infinitely large, there would be no difference, on average, among the different possible future configurations.

The presence of N in the definition of fitness for the case of within-generation variance suggests that the selection process under discussion is density dependent. Indeed, Gillespie (1974, p. 602) says that the population he is describing is "density-regulated," for a fixed population size is maintained. However, we need to recognize two differences between the case he is describing and the more standard notion of density dependence that is used, for example, to describe the effects of crowding. In the case of crowding, the size of the population has a causal impact on an organism's expected number of offspring. However, the point of Gillespie's analysis of within-generation variance is to show that fitness should not be defined as expected number of offspring. In addition, the case he is describing does not require that the size of the population have any causal influence on the reproductive behavior of individuals. The two A's and two C's in my example might be four cows standing in the four corners of a large pasture; the two A's have two calves each, whereas each of the C's flips a coin to decide whether she will have one calf or three. The cows are causally isolated from each other, but the fitnesses of the two strategies reflect population size.

In the two examples just presented, within-generation variance and between-generation variance have been understood in such a way that the former entails the latter, but not conversely. Because each C individual in each generation tosses a coin to determine whether she will have one offspring or three, it is possible for the mean offspring number produced by C parents in one generation to differ from the mean produced by the C parents in another. However, B parents in the same generation always have the same number of offspring. What this means is that B is a strategy that produces a purely between-generation variance, whereas C is a strategy that produces both within- and between-generation variance.

In both of the examples I have described, the argument that fitness must reflect variance as well as the (arithmetic) mean number of offspring depends on the assumption that fitnesses should predict *frequencies* of traits. If, instead, one merely demanded that the fitness of a trait should allow one to compute the expected *number* of individuals that will have the trait in the future, given the number of individuals that have the trait initially, the argument would not go through. The expected number of individuals in some future generation is computed by using the arithmetic mean number of offspring. When the population begins with two B individuals or with two C individuals, the expected number of B or C individuals in the next generation is four. The value that generates this next-generation prediction is two—the arithmetic mean of one and three. Note that the variance in offspring number and the size of the whole population (N) are irrelevant to this calculation.

That fitness is influenced by variance may seem paradoxical at first, but it makes sense in the light of a simple mathematical consideration. If traits X and Y are exclusive and exhaustive, then the number of X and Y individuals in a given generation determines the frequencies with which the two types occur at that time; however, it is not true that the *expected* number of X and Y individuals determines their *expected* frequencies. The reason is that frequency is a quotient:

frequency of X individuals =

(number of X individuals)/(total number of individuals).

The important point is that the expected value of a quotient is not identical with the quotient of expected values:

E(frequency of X individuals) $\neq$

E(number of X individuals)/E(total number of individuals).

This is why a general definition of fitness cannot equate fitness with expected offspring number. The fitness values of traits, along with the number of individuals initially possessing each trait, are supposed to entail the expected frequencies of the traits one or more generations in the future (if selection is the only force influencing evolutionary change). Expected number of offspring determines the value of the quotient on the right, but the expected frequency is left open.

Notice that this point about the definition of fitness differs from the one that Lewontin and Cohen (1969) made concerning population growth. Their point was to warn against using the *expected number* of individuals as a predictor. The present idea is that if one wants to predict the *expected frequencies* of traits, something beyond the expected number of individuals having the different traits must be taken into account.

2.4 Conclusion

Evolutionists are often interested in long-term trends rather than in short-term events. However, this fact about the interests of *theorists* does not mean that the *theory* enshrines an autonomous concept called "long-term fitness." The long-term is a function of what happens in successive short terms. This metaphysical principle is alive and well in evolutionary theory. However, traits like sex ratio show that the short term sometimes has to be longer than a single generation.

The example of sex ratio aside, we may begin thinking about the fitness of a trait by considering a total probability distribution, which specifies an individual's probability of having $0, 1, 2, 3 \ldots$ offspring. The expected value is a summary statistic of this distribution. Although this statistic sometimes is sufficient to predict expected frequencies, it is not always a sufficient predictor; when there is stochastic variation in offspring number, the variance is relevant as well.

Are the mean and variance together sufficient to define the concept of fitness? Beatty and Finsen (1989) point out that the skew of the distribution is sometimes relevant. In principle, fitness may depend on all the details of the probability distribution. However, Gillespie's analysis of within-generation variance leads to a more radical conclusion. When there is stochastic variation within generations, Gillespie says that the fitness of a trait is approximately the mean offspring number minus σ^2/N. Notice that the correction factor adverts to N, the population size; this is a piece of information not contained in the probability distribution associated with the trait. It is surprising that population size exerts a general and positive effect on fitness.

The results of Dempster, Haldane and Jayakar, and Gillespie show how the mathematical development of a theoretical concept can lead to a reconceptualization of its empirical meaning. In Newtonian mechanics, an object's mass does not depend on its velocity or on the speed of light; in relativity theory, this classical concept is replaced with relativistic mass, which is the classical mass divided by $(1 - v^2/c^2)^{1/2}$. As an object's velocity approaches zero, its relativistic mass approaches the classical value. In similar fashion, the corrected definition of fitness approaches the "classical" definition as σ^2 approaches zero. People reacted to Einstein's reconceptualization of mass by saying that it is strange and unintuitive, but the enhanced predictive power of relativity theory meant that these intuitions had to be re-educated. A definition of fitness that reflects the expected number of offspring, the variance in offspring number, and the population size yields more accurate predictions of expected population frequencies than the classical concept, and so it is preferable for the same reason.

It is sometimes said that relativity theory would not be needed if all objects moved slowly. After all, the correction factor $(1 - v^2/c^2)^{1/2}$ makes only a trivial difference when $v \ll c$. The claim is correct when the issue is prediction, but science has goals beyond that of making accurate predictions. There is the goal of understanding nature—of grasping what reality is like. Here we want to know which laws are true, and relativity theory has value here, whether or not we need to use that theory to make reasonable predictions. A similar point may apply to the corrected definition of fitness; perhaps evolving traits rarely differ significantly in their values of σ^2; if so, the corrected definitions will not be very useful when the goal is to predict new trait frequencies. This is an empirical question whose answer depends not just on how traits differ with respect to their variances but on the population size; after all, even modest differences in fitness can be important in large populations. But quite apart from the goal of making predictions, there is the goal of understanding nature—we want to understand what fitness is. In this theoretical context, the corrected definition of fitness is interesting.

What is the upshot of this discussion for the ''propensity interpretation of fitness?'' This interpretation has both a nonmathematical and a mathematical component. The nonmathematical idea is that an organism's fitness is its propensity to survive and be reproductively successful. Propensities are probabilistic dispositions. An organism's fitness is like a coin's probability of landing heads when tossed. Just as a coin's probability of landing heads depends on how it is tossed, so an organism's fitness depends on the environment in which it lives. And just as a coin's probability may fail to coincide exactly with the actual frequency of heads in a run of tosses, so an organism's fitness need not coincide exactly with the actual number of offspring it produces.

These ideas about fitness are not threatened by the foregoing discussion. However, the propensity interpretation also has its mathematical side, and this is standardly expressed by saying that fitness is a mathematical expectation (see, for example, Brandon 1978, Mills and Beatty 1979, Sober 1984). As we have seen, this characterization is not adequate in general, although it is correct in special circumstances. But perhaps all we need do is modify the mathematical characterization of fitness while retaining the idea that fitness is a propensity (Brandon 1990, p. 20).

This modest modification seems unobjectionable when there is between-generation variation in fitness; after all, if an organism's expected (= arithmetic mean) number of offspring reflects a ''propensity'' that it has, so too does its geometric mean averaged over time. However, when there is within-generation variation, the propensity interpretation is more problematic. The problem is the role of population size (N) in the definition. To say that a coin is fair—that $p = 1/2$, where p is the coin's probability of landing heads when tossed—is to describe a dispositional property that it has. However, suppose I define a new quantity, which is the coin's probability of landing heads minus σ^2/N, where N is the number of coins in some population that happens to contain the coin of interest. This new quantity $(p - \sigma^2/N)$ does not describe a property (just) of the coin. The coin is described by p and by σ^2, but N adverts to a property that is quite extrinsic to the coin.

Is it really tenable to say that p describes a propensity that the coin has but that $(p - \sigma^2/N)$ does not? After all, the coin's value for p reflects a fact about how the coin is tossed just as much as it reflects a fact about the coin's internal composition. Perhaps the propensity is more appropriately attributed to the entire coin-tossing device. However, $(p - \sigma^2/N)$ brings in a feature of the environment—N—that has no causal impact whatever on the coin's behavior when it is tossed. It is for this reason that we should decline to say that $(p - \sigma^2/N)$ represents a propensity of the coin.

I conclude that an organism's fitness is not a propensity that it has—at least not when fitness must reflect the existence of within-generation variance in offspring number. In this context, fitness becomes a more "holistic" quantity; it reflects properties of the organism's relation to its environment that affect how many offspring the organism has; but fitness also reflects a property of the containing population—namely, its census size—that may have no effect on the organism's reproductive behavior. Of course, the old idea that fitness is a mathematical expectation was consistent with the possibility that this expectation might be influenced by various properties of the population; frequency-dependent and density-dependent fitnesses are nothing new. What is new is that the *definition* of fitness, not just the factors that sometimes affect an individual's expected number of offspring, includes reference to census size.

2.5 Acknowledgments

I am very much in Dick Lewontin's debt. I spent my first sabbatical (1980–81) in his laboratory at the Museum of Comparative Zoology at Harvard. I had written one or two pieces on philosophy of biology by then, but I was very much a rookie in the subject. Dick was enormously generous with his time—we talked endlessly—and I came away convinced that evolutionary biology was fertile ground for philosophical reflection. While in his laboratory, I worked on the units of selection problem and on the use of a parsimony criterion in phylogenetic inference. I still have not been able to stop thinking and writing about these topics. Thanks to Dick, 1980–81 was the most intellectually stimulating year of my life.

Dick is a "natural philosopher." I do not mean this in the old-fashioned sense that he is a *scientist* (though of course he is that) but in the sense that he is a *natural at doing philosophy*. It was a striking experience during that year to find that Dick, a scientist,

was interested in the philosophical problems I was thinking about and that he was prepared to consider the possibility that they might be relevant to scientific questions. I came to the laboratory with the rather "theoretical" conviction that there should be common ground between science and philosophy, but the experience I had in the laboratory made me see that this could be true, not just in theory, but in practice.

During that year, I attended Dick's courses in biostatistics and population genetics; I gradually started to see how deeply the concept of probability figures in evolutionary biology. The present chapter, I think, is on a subject that is up Dick's alley. It is a pleasure to contribute this chapter to a volume that honors him.

My thanks to Martin Barrett, John Beatty, James Crow, Carter Denniston, Branden Fitelson, John Gillespie, David Lorvick, Steve Orzack, and to Dick as well for useful discussion of earlier drafts of this chapter.

References

Beatty, J., and Finsen, S. (1989). Rethinking the propensity interpretation—a peek inside Pandora's box. In *What the Philosophy of Biology Is*, ed. M. Ruse, pp. 17–30. Dordrecht, The Netherlands: Kluwer Publishers.

Brandon, R. (1978). Adaptation and evolutionary theory. *Studies in the History and Philosophy of Science* 9: 181–206.

Brandon, R. (1990). *Adaptation and Environment*. Princeton, NJ: Princeton University Press.

Cooper, W. (1984). Expected time to extinction and the concept of fundamental fitness. *Journal of Theoretical Biology* 107: 603–29.

Cronin, H. (1991). *The Ant and the Peacock*. Cambridge, UK: Cambridge University Press.

Dempster, E. R. (1955). Maintenance of genetic heterogeneity. *Cold Spring Harbor Symposium on Quantitative Biology* 20: 25–32.

Fisher, R. (1930). *The Genetical Theory of Natural Selection*. New York: Dover, 2d edition, 1957.

Gillespie, J. (1973). Natural selection with varying selection coefficients—a haploid model. *Genetical Research* 21: 115–20.

Gillespie, J. (1974). Natural selection for within-generation variance in offspring. *Genetics* 76: 601–6.

Gillespie, J. (1977). Natural selection for variances in offspring numbers—a new evolutionary principle. *American Naturalist* 111: 1010–14.

Haldane, J. B. S., and Jayakar, S. D. (1963). Polymorphism due to selection of varying direction. *Journal of Genetics* 58: 237–42.

Jeffrey, R. (1983). *The Logic of Decision*. Chicago: University of Chicago Press. 2d edition.

Lewontin, R., and Cohen, D. (1969). On population growth in a randomly varying environment. *Proceedings of the National Academy of Sciences* 62: 1056–60.

Mills, S., and Beatty, J. (1979). The propensity interpretation of fitness. *Philosophy of Science* 46: 263–86.

Nitecki, M. (ed.) (1988). *Evolutionary Progress*. Chicago: University of Chicago Press.

Paul, D. (1992). Fitness—historical perspectives. In *Keywords in Evolutionary Biology*, ed. E. Keller and E. Lloyd, pp. 112–14. Cambridge, MA: Harvard University Press.

Sober, E. (1984). *The Nature of Selection*. Cambridge: MIT Press. 2d edition, Chicago: University of Chicago Press, 1994.

Sober, E. (1994). Progress and direction in evolution. In *Progressive Evolution*?, ed. J. Campbell. Boston: Jones and Bartlett.

Thoday, J. (1953). Components of fitness. *Symposium of the Society for Experimental Biology* 7: 96–113.

Thoday, J. (1958). Natural selection and biological process. In *A Century of Darwin*, ed. S. Barnett, pp. 313–33. London: Heinemann.

II Units of Selection

3 Excerpts from *Adaptation and Natural Selection*

George C. Williams

I hope that this book will help to purge biology of what I regard as unnecessary distractions that impede the progress of evolutionary theory and the development of a disciplined science for analyzing adaptation. It opposes certain of the recently advocated qualifications and additions to the theory of natural selection, such as genetic assimilation, group selection and cumulative progress in adaptive evolution. It advocates a ground rule that should reduce future distractions and at the same time facilitate the recognition of really justified modifications of the theory. The ground rule—or perhaps *doctrine* would be a better term—is that adaptation is a special and onerous concept that should be used only where it is really necessary. When it must be recognized, it should be attributed to no higher a level of organization than is demanded by the evidence. In explaining adaptation, one should assume the adequacy of the simplest form of natural selection, that of alternative alleles in Mendelian populations, unless the evidence clearly shows that this theory does not suffice. . . .

Benefits to groups can arise as statistical summations of the effects of individual adaptations. When a deer successfully escapes from a bear by running away, we can attribute its success to a long ancestral period of selection for fleetness. Its fleetness is responsible for its having a *low probability* of death from bear attack. The same factor repeated again and again in the herd means not only that it is a herd of fleet deer, but also that it is a fleet herd. The group therefore has a *low rate* of mortality from bear attack. When every individual in the herd flees from a bear, the result is effective protection of the herd.

As a very general rule, with some important exceptions, the fitness of a group will be high as a result of this sort of summation of the adaptations of its members. On the other hand, such simple summations obviously cannot produce collective fitness as

George C. Williams, *Adaptation and Natural Selection*, 4–5, 16–19, 22–25, 92–101, 108–124, 208–212. © 1966 Princeton University Press. 1994 renewed PUP. Reprinted by permission of Princeton University Press.

high as could be achieved by an adaptive organization of the group itself. We might imagine that mortality rates from predation by bears on a herd of deer would be still lower if each individual, instead of merely running for its life when it saw a bear, would play a special role in an organized program of bear avoidance. There might be individuals with especially well-developed senses that could serve as sentinels. Especially fleet individuals could lure bears away from the rest, and so on. Such individual specialization in a collective function would justify recognizing the herd as an adaptively organized entity. Unlike individual fleetness, such group-related adaptation would require something more than the natural selection of alternative alleles as an explanation.

It may also happen that the incidental effects of individual activities, of no functional significance in themselves, can have important statistical consequences, sometimes harmful, sometimes beneficial. The depletion of browse is a harmful effect of the feeding activities of each member of a dense population of deer. If browse depletion were beneficial, I suspect that someone, sooner or later, would have spoken of the feeding behavior of deer as a mechanism for depleting browse. A statement of this sort should not be based merely on the evidence that the statistical effect of eating is beneficial; it should be based on an examination of the causal mechanisms to determine whether they cannot be adequately explained as individual adaptations for individual nourishment.

The feeding activities of earthworms would be a better example, because here the incidental statistical effects *are* beneficial, from the standpoint of the population and even of the ecological community as a whole. As the earthworm feeds, it improves the physical and chemical properties of the soil through which it moves. The contribution from each individual is negligible, but the collective contribution, cumulative over decades and centuries, gradually improves the soil as a medium for worm burrows and for the plant growth on which the earthworm's feeding ultimately depends. Should we therefore call the causal activities of the earthworm a soil-improvement mechanism? Apparently Allee (1940) believed that some such designation is warranted by the fact that soil improvement is indeed a result of the earthworm's activities. However, it we were to examine the digestive system and feeding behavior of an earthworm, I assume that we would find it adequately explained on the assumption of design for individual nutrition. The additional assumption of design for soil improvement would explain nothing that is not also explainable as a nutritional adaptation. It would be a violation of parsimony to assume both explanations when one suffices. Only if one denied that some benefits can arise by chance instead of by design, would there be a reason for postulating an adaptation behind every benefit.

On the other hand, suppose we did find some features of the feeding activities of earthworms that were inexplicable as trophic adaptations but were exactly what we should expect of a system designed for soil improvement. We would then be forced to recognize the system as a soil-modification mechanism, a conclusion that implies a quite different level of adaptive organization from that implied by the nutritional function. As a digestive system, the gut of a worm plays a role in the adaptive organization of that worm and nothing else, but as a soil-modification system it would play a role in the adaptive organization of the whole community. This, as I will argue at length in later chapters, is a reason for rejecting soil-improvement as a purpose of the worm's activities if it is possible to do so. Various levels of adaptive organization, from the subcellular to the biospheric, might conceivably be recognized, but the principle of parsimony demands that we recognize adaptation at the level necessitated by the facts and no higher.

It is my position that adaptation need almost never be recognized at any level above that of a pair of parents and associated offspring. As I hope to show in the later chapters, this conclusion seldom has to rest on appeals to parsimony alone, but is usually supported by specific evidence.

The most important function of this book is to echo a plea made many years ago by E. S. Russell (1945) that biologists must develop an effective set of principles for dealing with the general phenomenon of biological adaptation. This matter is considered mainly in the final chapter.

The essence of the genetical theory of natural selection is a statistical bias in the relative rates of survival of alternatives (genes, individuals, etc.). The effectiveness of such bias in producing adaptation is contingent on the maintenance of certain quantitative relationships among the operative factors. One necessary condition is that the selected entity must have a high degree of permanence and a low rate of endogenous change, relative to the degree of bias (differences in selection coefficients). Permanence implies reproduction with a potential geometric increase.

Acceptance of this theory necessitates the immediate rejection of the importance of certain kinds of selection. The natural selection of phenotypes cannot in itself produce cumulative change, because phenotypes are extremely temporary manifestations. They are the result of an interaction between genotype and environment that produces what we recognize as an individual. Such an individual consists of genotypic information and information recorded since conception. Socrates consisted of the genes his parents gave him, the experiences they and this environment later provided, and a growth and development mediated by numerous meals. For all I know, he may have been very successful in the evolutionary sense of leaving numerous offspring. His phenotype, nevertheless, was utterly destroyed by the hemlock and has never since been duplicated. If the hemlock had not killed him, something else soon would have. So however natural selection may have been acting on Greek phenotypes in the fourth century B.C. it did not of itself produce any cumulative effect.

The same argument also holds for genotypes. With Socrates' death, not only did his phenotype disappear, but also his genotype. Only in species that can maintain unlimited clonal reproduction is it theoretically possible for the selection of genotypes to be an important evolutionary factor. This possibility is not likely to be realized very often, because only rarely would individual clones persist for the immensities of time that are important in evolution. The loss of Socrates' genotype is not assuaged by any consideration of how prolifically he may have reproduced. Socrates' genes may be with us yet, but not his genotype, because meiosis and recombination destroy genotypes as surely as death.

It is only the meiotically dissociated fragments of the genotype that are transmitted in sexual reproduction, and these fragments are further fragmented by meiosis in the next generation. If there is an ultimate indivisible fragment it is, by definition, "the gene" that is treated in the abstract discussions of population genetics. Various kinds of suppression of recombination may cause a major chromosomal segment or even a whole chromosome to be transmitted entire for many generations in certain lines of descent. In such cases the segment or chromosome behaves in a way that approximates the population genetics of a single gene. In this book I use the term *gene* to mean "that which segregates and recombines with appreciable frequency." Such genes are potentially immortal, in the sense of there being no physiological limit to their survival, because of their potentially reproducing fast enough to compensate for their destruction by external agents. They also have a high degree of qualitative stability. Estimates of mutation rates range from about 10^{-4} to 10^{-10} per generation. The rates of selection of alternative alleles can be much higher. Selection among the progeny of individuals heterozygous for recessive lethals would eliminate half the lethal genes in one generation. Aside from lethal and markedly deleterious genes in experimental populations, there is abundant evidence (e.g., Fisher and Ford 1947; Ford 1956; Clarke, Dickson, and Sheppard 1963) for selection coefficients in nature that exceed mutation rates by one to many multiples of ten. There can be no doubt that the selective accumulation of genes can be effective. In evolutionary theory, a gene could be defined as any hereditary information for which there is a favorable or unfavorable selection bias equal to several or many times its rate of endogenous change. The prevalence of such stable entities in the heredity of populations is a measure of the importance of natural selection.

Natural selection would produce or maintain adaptation as a matter of definition. Whatever gene is favorably selected is better adapted than its unfavored alternatives. This is the reliable outcome of such selection, the prevalence of well-adapted genes. The selection of such genes of course is mediated by the phenotype, and to be favorably selected, a gene must augment phenotypic reproductive success as the arithmetic mean effect of its activity in the population in which it is selected. . . .

This [work] is a rejoinder to those who have questioned the adequacy of the traditional model of natural selection to explain evolutionary adaptation. The topics considered in the preceding chapters relate mainly to the adequacy of this model in the realms of physiological, ecological, and developmental mechanisms, matters of primary concern to individual organisms. At the individual level the adequacy of the selection of alternative alleles has been challenged to only a limited degree. Many more doubts on the importance of such selection have been voiced in relation to the phenomenon of interactions among individuals. Many biologists have implied, and a moderate number have explicitly maintained, that groups of interacting individuals may be adaptively organized in such a way that individual interests are compromised by a functional subordination to group interests.

It is universally conceded by those who have seriously concerned themselves with this problem (e.g., Allee *et al.* 1949; Haldane 1932; Lewontin 1958, 1962; Slobodkin 1954; Wynne-Edwards 1962; Wright 1945) that such group-related adaptations must be attributed to the natural selection of alternative *groups* of individuals and that the natural selection of alternative alleles within populations will be opposed to this development. I am in entire agreement with the reasoning behind this conclusion. Only by a theory of between-group selection could we achieve a scientific explanation of group-related adaptations. However, I would question one of the premises on which the reasoning is based. Chapters 5 to 8 [of *Adaptation and Natural Selection*] will be primarily a defense of the thesis that group-related adaptations do not, in fact, exist. A *group* in this discussion should be understood to mean something other than a family and to be composed of individuals that need not be closely related.

The present chapter examines the logical structure of the theory of selection between groups, but first I wish to consider an apparent exception to the rule that the natural selection of individuals cannot produce group-related adaptations. This exception may be found in animals that live in stable social groups and have the intelligence and other mental qualities necessary to form a system of personal friendships and animosities that transcend the limits of family relationship. Human society would be impossible without the ability of each of us to know, individually, a variety of neighbors. We learn that Mr. X is a noble gentleman and that Mr. Y is a scoundrel. A moment of reflection should convince anyone that these relationships may have much to do with evolutionary success. Primitive man lived in a world in which stable interactions of personalities were very much a part of his ecological environment. He had to adjust to this set of ecological factors as well as to any other. If he was socially acceptable, some of his neighbors might bring food to himself and his family when he was temporarily incapacitated by disease or injury. In time of dearth, a stronger neighbor might rob our primitive man of food, but the neighbor would be more likely to rob a detestable primitive Mr. Y and his troublesome family. Conversely, when a poor Mr. X is sick our primitive man will, if he can, provide for him. Mr. X's warm heart will know the emotion of gratitude and, since he recognizes his benefactor and remembers the help provided, will probably reciprocate some day. A number of people, including Darwin

(1896, Chap. 5), have recognized the importance of this factor in human evolution. Darwin speaks of it as the "lowly motive" of helping others in the hope of future repayment. I see no reason why a conscious motive need be involved. It is necessary that help provided to others be occasionally reciprocated if it is to be favored by natural selection. It is not necessary that either the giver or the receiver be aware of this.

Simply stated, an individual who maximizes his friendships and minimizes his antagonisms will have an evolutionary advantage, and selection should favor those characters that promote the optimization of personal relationships. I imagine that this evolutionary factor has increased man's capacity for altruism and compassion and has tempered his ethically less acceptable heritage of sexual and predatory aggressiveness. There is theoretically no limit to the extent and complexity of group-related behavior that this factor could produce, and the immediate goal of such behavior would always be the well-being of some other individual, often genetically unrelated. Ultimately, however, this would not be an adaptation for group benefit. It would be developed by the differential survival of individuals and would be designed for the perpetuation of the genes of the individual providing the benefit to another. It would involve only such immediate self-sacrifice for which the probability of later repayment would be sufficient justification. The natural selection of alternative alleles can foster the production of individuals willing to sacrifice their lives for their offspring, but never for mere friends.

The prerequisites for the operation of this evolutionary factor are such as to confine it to a minor faction of the Earth's biota. Many animals form dominance hierarchies, but these are not sufficient to produce an evolutionary advantage in mutual aid. A consistent interaction pattern between hens in a barnyard is adequately explained without postulating emotional bonds between individuals. One hen reacts to another on the basis of the social releasers that are displayed, and if individual recognition is operative, it merely adjusts the behavior towards another individual according to the immediate results of past interactions. There is no reason to believe that a hen can harbor grudges against or feel friendship toward another hen. Certainly the repayment of favors would be out of the question.

A competition for social goodwill cannot fail to have been a factor in human evolution, and I would expect that it would operate in many of the other primates. Altman (1962) described the formation of semipermanent coalitions between individuals within bands of wild rhesus monkeys and cited similar examples from other primates. Members of such coalitions helped each other in conflicts and indulged in other kinds of mutual aid. Surely an individual that had a better than average ability to form such coalitions would have an evolutionary advantage over its competitors. Perhaps this evolutionary factor might operate in the evolution of porpoises. This seems to be the most likely explanation for the very solicitous behavior that they sometimes show toward each other (Slijper 1962, pp. 193–197). I would be reluctant, however, to recognize this factor in any group but the mammalia, and I would imagine it to be confined to a minority of this group. For the overwhelming mass of the Earth's biota, friendship and hate are not parts of the ecological environment, and the only way for socially beneficial self-sacrifice to evolve is through the biased survival and extinction of populations, not by selective gene substitution within populations.

To minimize recurrent semantic difficulties, I will formally distinguish two kinds of natural selection. The natural selection of alternative alleles in a Mendelian population will henceforth be called *genic selection*. The natural selection of more inclusive entities will be called *group selection*, a term introduced by Wynne-Edwards (1962). *Intrademic* and *interdemic*, and other terms with the same prefixed, have been used to make the same distinction. It has been my experience, however, that the repeated use in the same discussion of "inter" and "intra" for specifically contrasted concepts is a certain cause of confusion, unless a reader exerts an inconvenient amount of attention to spelling, or a speaker indulges in highly theatrical pronunciation.

The definitions of other useful terms, and the conceptual relations between the various creative evolutionary factors and the production of adaptation are indicated in figure 3.1. Genic selection should be assumed to imply the current conception of natural selection often termed *neo-Darwinian*. An *organic adaptation* would be a mechanism designed to promote the success of an individual organism, as measured by the extent to which it contributes genes to later generations of the population of which it is a member. It has the individual's *inclusive fitness* (Hamilton 1964) as its goal. Biotic evolution is any change in a biota. It can be brought about by an evolutionary change in one or more of the constituent populations, or merely by a change in their relative numbers. A *biotic adaptation* is a mechanism designed to promote the success of a biota,

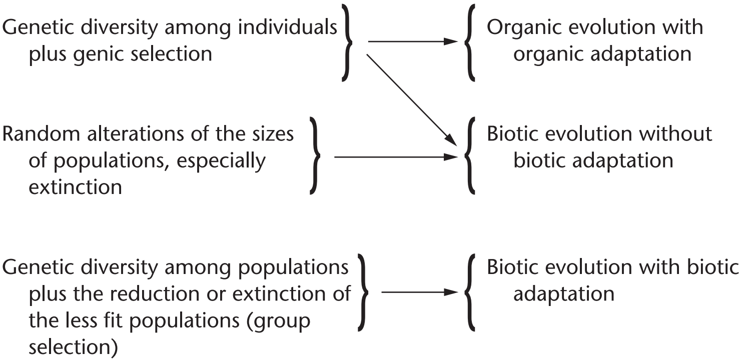


Figure 3.1
Summary comparison of organic and biotic evolution, and of organic and biotic adaptation.

as measured by the lapse of time to extinction. The biota considered would have to be restricted in scope so as to allow comparison with other biotas. It could be a single biome, or community, or taxonomic group, or, most often, a single population. A change in the fish-fauna of a lake would be considered biotic evolution. It could come about through some change in the characters of one or more of the constituent populations or through a change in the relative numbers of the populations. Either would result in a changed fish-fauna, and such a change would be biotic evolution. A biotic adaptation could be a mechanism for the survival of such a group as the fish-fauna of a lake, or of any included population, or of a whole species that lives in that lake and elsewhere.

I believe that it is useful to make a formal distinction between biotic and organic evolution, and that certain fallacies can be avoided by keeping the distinction in mind. It should be clear that, in general, the fossil record can be a direct source of information on organic evolution only when changes in single populations can be followed through a continuous sequence of strata. Ordinarily the record tells us only that the biota at time t' was different from that at time t and that it must have changed from one state to the other during the interval. An unfortunate tendency is to forget this and to assume that the biotic change must be ascribed to appropriate organic change. The horse-fauna of the Eocene, for instance, was composed of smaller animals than that of the Pliocene. From this observation, it is tempting to conclude that, at least most of the time and on the average, a larger than mean size was an advantage to an individual horse in its reproductive competition with the rest of its population. So the component populations of the Tertiary horse-fauna are presumed to have been evolving larger size most of the time and on the average. It is conceivable, however, that precisely the opposite is true. It may be that at any given moment during the Tertiary, most of the horse populations were evolving a smaller size. To account for the trend towards larger size it is merely necessary to make the additional assumption that group selection favored such a tendency. Thus, while only a minority of the populations may have been evolving a larger size, it could have been this minority that gave rise to most of the populations of a million years later. Figure 3.2 shows how the same observations on the fossil record can be rationalized on two entirely different bases. The unwarranted assumption of organic evolution as an explanation for biotic evolution dates at least from Darwin. In *The Origin of Species* he dealt with a problem that he termed "advance in organization." He interpreted the fossil record as indicating that the biota has evolved progressively "higher" forms from the Cambrian to Recent, clearly a change in the biota. His explanation, however, is put largely in terms of the advantage that an individual might have over his neighbors by virtue of a larger brain, greater histological complexity, etc. Darwin's reasoning here is analogous to that of someone who would expect that if the organic evolution of horses proceeded toward larger size during the Tertiary, most equine mutations during this interval must have

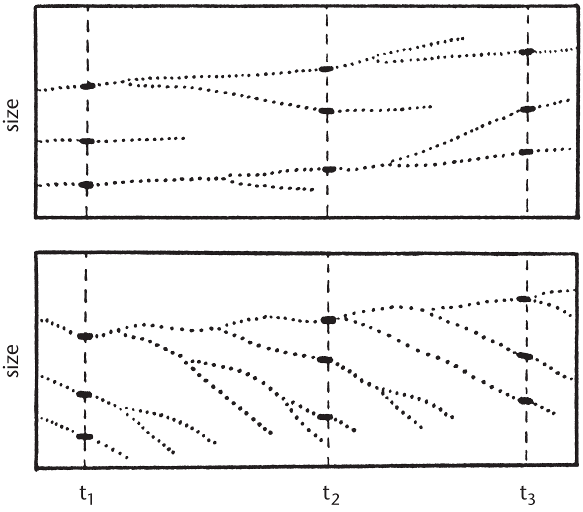


Figure 3.2

Alternative ways of interpreting the same observations of the fossil record. Average sizes in hypothetical horse species at three different times are indicated by boldface marks on the vertical timescale at times t_1, t_2, and t_3. Upper and lower diagrams show the same observations. In the upper, hypothetical phylogenies explain the observations as the result of the organic evolution of increased size and of occasional chance extinction. In the lower, hypothetical phylogenies indicate the organic evolution mainly of decreased size, but with effective counteraction by group selection so that the biota evolves a larger average size.

caused larger size in the affected individuals. I suspect that most biologists would tend toward the opposite view, and expect that random changes in the germ plasm would be more likely to curtail growth than to augment it. Organic evolution would normally run counter to the direction of mutation pressure. There is a formally similar relation between organic evolution and group selection. Organic evolution provides genetically different populations, the raw material on which group selection acts. There is no necessity for supposing that the two forces would normally be in precisely the same direction. It is conceivable that at any given moment since the Cambrian, the majority of organisms were evolving along lines that Darwin would consider retrogression, degeneration, or narrow specialization, and that only a minority were progressing. If the continued survival of populations were sufficiently biased in favor of this minority, however, the biota as a whole might show "progress" from one geologic period to the next. I expect that the fossil record is actually of little use in evaluating the relative potency of genic and group selection.

In another respect the analogy between mutation and organic evolution as sources of diversity may be misleading. Mutations occur at random and are usually destructive of any adaptation, whereas organic evolution is largely concerned with the production or at least the maintenance of organic adaptation. Any biota will show a system of adaptations. If there is no group selection, i.e., if extinction is purely by chance, the adaptations shown will be a random sample of those produced by genic selection. If group selection does operate, even weakly, the adaptations shown will be a biased sample of those produced by genic selection. Even with such bias in the kinds of adaptations actually represented, we would still recognize genic selection as the process that actually produced them. We could say that the adaptations were produced by group selection only if it was so strong that it constantly curtailed organic evolution in all but certain favored directions and was thus able, by its own influence, to accumulate the functional details of complex adaptations. This distinction between the production of a biota with a certain set of organic adaptations and the production of the adaptations of a biota will be emphasized again in a number of contexts.

. . . It is essential, before proceeding further with the discussion, that the reader firmly grasp the general meaning of biotic adaptation. He must be able to make a conceptual distinction between a population of adapted insects and an adapted population of insects. The fact that an insect population survives through a succession of generations is not evidence for the existence of biotic adaptation. The survival of the population may be merely an incidental consequence of the organic adaptations by which each insect attempts to survive and reproduce itself. The survival of the population depends on these individual efforts. To determine whether this survival is the proper function or merely an incidental by-product of the individual effort must be decided by a critical examination of the reproductive processes. We must decide: Do these processes show an effective design for maximizing the number of descendants of the individual, or do they show an effective design for maximizing the number, rate of growth, or numerical stability of the population or larger system? Any feature of the system that promotes group survival and cannot be explained as an organic adaptation can be called a biotic adaptation. If the population has such adaptations it can be called an adapted population. If it does not, if its continued survival is merely incidental to the operation of organic adaptations, it is merely a population of adapted insects.

Like the theory of genic selection, the theory of group selection is logically a tautology and there can be no sane doubt about the reality of the process. Rational criticism must center on the importance of the process and on its adequacy in explaining the phenomena attributed to it. An important tenet of evolutionary theory is that natural selection can produce significant cumulative change only if selection coefficients are high relative to the rates of change of the selected entity. Since genic selection coefficients are high relative to mutation rates, it is logically possible for the natural selection of alternative alleles to have important cumulative effects. It was pointed out [above] that there can be no effective selection of somata. They have limited life spans and (often) zero biotic potential. The same considerations apply to populations of somata. I also pointed out that genotypes have limited lives and fail to reproduce themselves (they are destroyed by meiosis and recombination), except where clonal reproduction is possible. This is equally true of populations of genotypes. All of the genotypes of fruit-fly populations now living will have ceased to exist in a few weeks. Within a population, only the gene is stable enough to be effectively selected. Likewise in selection among populations, only populations of genes (gene pools) seem to qualify with respect to the necessary stability. Even gene pools will not always qualify. If populations are evolving rapidly and have a low rate of extinction and replacement, the rate of endogenous change might be too great for group selection to have any cumulative effect. This argument precisely parallels that which indicates that mutation rates must be low relative to selection coefficients for genic selection to be effective.

If a group of adequately stable populations is available, group selection can theoretically produce biotic adaptations, for the same reason that genic selection can produce organic adaptations. Consider again the evolution of size among Tertiary horses. Suppose that at one time there was a genus of two species, one that averaged 100 kilograms when full grown and another that averaged 150 kilograms. Assume that genic selection in both species favored a smaller size so that a million years later the larger of the two averaged only 130 kilograms and the smaller had become extinct, but had lost 20 kilograms before it did so. In this case we could say that the genus evolved an increased size, even though both of the included species evolved a decreased size. If the extinction of the smaller species is not just a chance event but is attributable to its smaller size, we might refer to large size as a biotic adaptation of a simple sort. However, it is the origin of complex adaptations, for which the concept of functional design would be applicable, that is the important consideration.

If alternative gene pools are not themselves stable, it is still conceivable that group selection could operate among more or less constant rates of change. A system of relatively stable rates of change in the gene frequencies of a population might be called an evolutionary trajectory. It could be described as a vector in n-dimensional space, with n being the number of relevant gene frequencies. In a given sequence of a few generations a gene pool may be undergoing certain kinds of change at a certain rate. This is only one of an infinite number of other evolutionary trajectories that might conceivably be followed. Some trajectories may be more likely to lead to extinction than others, and group selection will then operate by allowing different kinds of evolutionary change to continue for different average lengths of time. There is paleontological evidence that certain kinds of evolutionary change may continue for appreciable lengths of time on a geological scale. Some of the supposed examples disappear as the evidence accumulates and shows that actual courses of evolution are more complex than they may have seemed at first. Other examples are apparently real and are attributed by Simpson (1944, 1953) to continuous genic selection in certain directions, a process he terms ''ortho-selection.''

Wright (1945) proposed that group selection would be especially effective in a species that was divided up into many small populations that were almost but not quite isolate from each other. Most of the evolutionary change in such a species would be in accordance with genic selection coefficients, but the populations are supposed to be small enough so that genes would occasionally be fixed by drift in spite of adverse selection within a population. Some of the genes so fixed might benefit the population as a whole even though they were of competitive disadvantage within the population. A group so favored would increase in size (regarded as a benefit in Wright's discussion) and send out an augmented number of emigrants to neighboring populations. These migrants would partly or wholly counteract the adverse selection of the gene in neighboring populations and give them repeated opportunity for the chance fixation of the gene. The oft-repeated operation of this process eventually would produce complex adaptations of group benefit, but of competitive disadvantage to an individual. According to this theory, selection not only can act on preexisting variation, but also can help to produce the variation on which it acts, by repeatedly introducing the favored gene into different populations.

Wright formally derived this model in a review of a book by G. G. Simpson. Later, Simpson (1953, pp. 123, 164–165) briefly criticized Wright's theory by pointing out that it leaves too much to a rather improbable concatenation of the population parameters of size, number, degree of isolation, and the balance of genic and group selection coefficients. The populations have to be small enough for genetic drift to be important, but not so small that they are in danger of extinction, and they have to be big enough for certain gene substitutions to be more important than chance factors in determining size and rate of emigration. The unaugmented rates of immigration must be too small to reestablish the biotically undesirable gene after it is lost by drift. The populations must be numerous enough for the postulated process to work at a variety of loci, and each of the populations must be within the necessary size range. Lastly, the balance of these various factors must persist long enough for an appreciable amount of evolutionary change to take place. At the moment, I can see no hope of achieving any reliable estimate of how frequently the necessary conditions have been realized, but surely the frequency of such combinations of circumstances must be relatively low and the combinations quite temporary when they do occur. Simpson also expressed doubts on the reality of the biotic adaptations that Wright's theory was proposed to explain.

A number of writers have since postulated a role for the selection of alternative populations within a species in the production of various supposed ''altruistic'' adaptations. Most of these references, however, have completely ignored the problem that Wright took such pains to resolve. They have ignored the problem of how whole populations can acquire the necessary genes in high frequency in the first place. Unless some do and some do not, there is no set of alternatives for group selection to act upon. Wright was certainly aware, as some later workers apparently were not, that even a minute selective disadvantage to a gene in a population of moderate size can cause an almost deterministic reduction of the gene to a negligible frequency. This is why he explicitly limited the application of his model to those species that are subdivided into many small local populations with only occasional migrants between them. Others have postulated such group selection as an evolutionary factor in species that manifestly do not have the requisite population structures. Wynne-Edwards (1962), for example, postulated the origin of biotic adaptations of individual disadvantage, by selection among populations of smelts, in which even a single spawning aggregation may consist of tens of thousands of individuals. He envisioned the same process for marine invertebrates that may exist as breeding adults by the million per square mile and have larval stages that may be dispersed many miles from their points of origin.

A possible escape from the necessity of relying on drift in small populations to fix the genes that might contribute to biotic adaptation, is to assume that such genes are not uniformly disadvantageous in competitive individual relationships. If such a gene were, for some reason, individually advantageous in one out of ten populations, group selection could work by making the descendants of that population the sole representatives of the species a million years later. However, this process also loses plausibility on close examination. Low rates of endogenous change relative to selection coefficients are a necessary precondition for any effective selection. The necessary stability is the general rule for genes. While gene pools or evolutionary trajectories can persist little altered through a long period of extinction and replacement of populations, there is no indication that this is the general rule. Hence the effectiveness of group selection is open to question at the axiomatic level for almost any group of organisms. The possibility of effective group selection can be dismissed for any species that consists, as many do, of a single population. Similarly the group selection of alternative species cannot direct the evolution of a monotypic genus, and so on.

Even in groups in which all of the necessary conditions for group selection might be demonstrated, there is no assurance that these conditions will continue to prevail. Just as the evolution of even the simplest organic adaptation requires the operation of selection at many loci for many generations, so also would the production of biotic adaptation require the selective substitution of many groups. This is a major theoretical difficulty. Consider how rapid is the turnover of generations in even the slowest breeding organisms, compared to the rate at which populations replace each other. The genesis of biotic adaptation must for this reason be orders of magnitude slower than that of organic adaptation. Genic selection may take the form of the replacement of one allele by another at the rate of 0.01 per generation, to choose an unusually high figure.

Would the same force of group selection mean that a certain population would be 0.01 larger, or be growing 0.01 faster, or be 0.01 less likely to become extinct in a certain number of generations, or have a 0.01 greater emigration rate than another population? No matter which meaning we assign, it is clear that what would be a powerful selective force at the genic level would be trivial at the group level. For group selection to be as strong as genic selection, its selection coefficients would have to be much greater to compensate for the low rate of extinction and replacement of populations.

The rapid turnover of generations is one of the crucial factors that makes genic selection such a powerful force. Another is the large absolute number of individuals in even relatively small populations, and this brings us to another major difficulty in group selection, especially at the species level. A species of a hundred different populations, sufficiently isolated to develop appreciable genetic differences, would be exceptional in more groups of organisms. Such a complexly subdivided group, however, might be in the same position with respect to a bias of 0.01 in the extinction and replacement of groups, as a population of fifty diploid individuals with genic selection coefficients that differ by 0.01. In the population of fifty we would recognize genetic drift, a chance factor, as much more important than selection as an evolutionary force. Numbers of populations in a species, or of taxa in higher categories, are usually so small that chance would be much more important in determining group survival than would even relatively marked genetic differences among the groups. By analogy with the conclusions of population genetics, group selection would be an important creative force only where there were at least some hundreds of populations in the group under consideration.

Obviously the comments above are not intended to be a logically adequate evaluation of group selection. Analogies with the conclusions on genic selection are only analogies, not rigorously reasoned connections. I would suggest, however, that they provide a reasonable basis for skepticism about the effectiveness of this evolutionary force. The opposite tendency is frequently evident. A biologist may note that, logically and empirically, the evolutionary process is capable of producing adaptations of great complexity. He then assumes that these adaptations must include not only the organic but also the biotic, usually discussed in such terms as “for the good of the species.” A good example is provided by Montagu (1952), who summarized the modern theory of natural selection and in so doing presented an essentially accurate picture of selective gene substitution by the differential reproductive survival of individuals. Then in the same work he states, “We begin to understand then, that evolution itself is a process which favors cooperating rather than disoperating groups and that ‘fitness’ is a function of the group as a whole rather than separate individuals.” This kind of evolution and fitness is attributed to the previously described natural selection of individuals. Such an extrapolation from conclusions based on analyses of the possibilities of selective gene substitutions *in* populations to the production of biotic adaptations *of* populations is entirely unjustified. Lewontin (1961) has pointed out that population genetics as it is known today relates to genetic processes in populations, not of populations.

Lewontin (1962; Lewontin and Dunn 1960) has produced what seems to me to be the only convincing evidence for the operation of group selection. There is a series of alleles symbolized by *t* in house-mouse populations that produces a marked distortion of the segregation ratio of sperm. As much as 95 percent of the sperm of a heterozygous male may bear such a gene, and only 5 percent bear the wild-type allele. This marked selective advantage is opposed by other adverse effects in the homozygotes, either an embryonic lethality or male sterility. Such characters as lethality, sterility, and measurable segregation ratios furnish an excellent opportunity for calculating the effect of selection as a function of gene frequency in hypothetical populations. Such calculations, based on a deterministic model of selection, indicate that these alleles should have certain equilibrium frequencies in the populations in which they occur. Studies of wild populations, however, consistently give frequencies below the calculated values. Lewontin concludes that the deficiency must be ascribed to some force in opposition to genic selection, and that group selection is the likely force. He showed that by substituting a stochastic model of natural selection, so as to allow for a certain rate of fixation of one or another allele in family groups and small local populations, he could account for the observed low frequencies of the *t*-alleles.

It should be emphasized that this example relates to genes characterized by lethality or sterility and extremely marked segregation distortions. Selection of such genes is of the maximum possible intensity. Important changes in frequency can occur in a very few generations as a result of genic selection, and no long-term isolation is necessary. Populations so altered would then be subject to unusually intense group selection. A population in which a segregation distorter reaches a high frequency will rapidly become extinct. A small population that has such a gene in low frequency can lose it by drift and thereafter replace those that have died out. Only one locus is involved. One cannot argue form this example that group selection would be effective in producing a complex adaptation involving closely adjusted gene frequencies at a large number of loci. Group selection in this example cannot maintain very low frequencies of the biotically deleterious gene in a population because even a single heterozygous male immigrant can rapidly "poison" the gene pool. The most important question about the selection of these genes is why they should produce such extreme effects. The segregation distortion makes the genes extremely difficult to keep at low frequency by either genic or group selection. Why has there not been an effective selection of modifiers that would reduce this distortion? Why also has there not been effective selection for modifiers that would abolish the lethality and sterility. The *t*-alleles certainly must constitute an important part of the genetic environment of every other gene in the population. One would certainly expect the other genes to become adapted to their presence.

Segregation distortion is something of a novelty in natural populations. I would be inclined to attribute the low frequency of such effects to the adjustment of each gene to its genetic environment. When distorter genes appear they would be expected to replace their alleles unless they produced, like the *t*-alleles, drastic reductions in fitness at some stage of development. When such deleterious effects are mild, the population would probably survive and would gradually incorporate modifiers that would reduce the deleterious effects. In other words, the other genes would adjust to their new genetic environment. It is entirely possible, however, that populations and perhaps entire species could be rendered extinct by the introduction of such genes as the *t*-alleles of mice. Such an event would illustrate the production, by genic selection, of characters that are highly unfavorable to the survival of the species. The gene in question would produce a high phenotypic fitness in the gamete stage. It might have a low effect on some other stage. The selection coefficient would be determined by the mean of these two effects relative to those of alternative alleles, regardless of the effect on population survival. I wonder if anyone has thought of controlling the mouse population of an area by flooding it with *t*-carriers.

I am entirely willing to concede that the kinds of adaptations evolved by a population, for instance segregation distortion, might influence its chance for continued survival. I question only the effectiveness of this extinction-bias in the production and maintenance of any adaptive mechanisms worthy of the name. This is not the same as denying that extinction can be an important factor in biotic evolution. The conclusion is inescapable that extinction has been extremely important in producing the Earth's biota as we know it today. Probably only on the order of a dozen Devonian vertebrates have left any Recent descendants. If it had happened that some of these dozen had not survived, I am sure that the composition of today's biota would be profoundly different.

Another example of the importance of extinction can be taken from human evolution. The modern races and various extinct hominids derive from a lineage that diverged from the other Anthropoidea a million or perhaps several million years ago. There must have been a stage in which man's ancestors were congeneric with, but specifically distinct from, the ancestors of the modern anthropoid apes. At this time there were probably several and perhaps many other species in this genus. All but about four, however, became extinct. One that happened to survive produced the gibbons, another the orang, another the gorilla and chimpanzee, and another produced the hominids. These were only four (or perhaps three or five) of an unknown number of contemporary Pliocene alternatives. Suppose that the number had been one less, with man's ancestor being assigned to the group that became extinct! We have no idea how many narrow escapes from extinction man's lineage may have experienced. There would have been nothing extraordinary about his extinction; on the contrary, this is the statistically most likely development. The extinction of this lineage would, however, have provided the world today with a strikingly different biota. This one ape, which must have had a somewhat greater than average tendency toward bipedal locomotion and, according to recent views, a tendency towards predatory pack behavior, was transferred by evolution from an ordinary animal, with an ordinary existence, to a cultural chain reaction. The production and maintenance of such tributary adaptations as an enlarged brain, manual dexterity, the arched foot, etc. was brought about by the gradual shifting of gene frequencies at each genetic locus in response to changes in the genetic, somatic, and ecological environments. It was this process that fashioned a man from a beast. The fashioning was not accomplished by the survival of one animal type and the extinction of others.

I would concede that such matters of extinction and survival are extremely important in biotic evolution. Of the systems of adaptations produced by organic evolution during any given million years, only a small proportion will still be present several million years later. The surviving lines will be a somewhat biased sample of those actually produced by genic selection, biased in favor of one type of adaptive organization over another, but survival will always be largely a matter of historical accident. It may be that some people would not even recognize such chance extinction as important in biotic evolution. Ecologic determinists might attribute more of a role to the niche factor; man occupies an ecologic niche, and if one ancestral ape had failed to fill it, another would have. This sort of thinking probably has some validity, but surely historical contingency must also be an important factor in evolution. The Earth itself is a unique historical phenomenon, and many unique geological and biological events must have had a profound effect on the nature of the world's biota.

There is another example that should be considered, because it has been used to illustrate a contrary point of view. The extinction of the dinosaurs may have been a necessary precondition to the production of such mammalian types as elephants and bears. This extinction, however, was not the creative force that designed the locomotor and trophic specializations of these mammals. That force can be recognized in genic selection in the mammalian populations. There are analogies in human affairs. In World War II there was a rubber shortage due to the curtailment of imports of natural rubber. Scientists and engineers were thereby stimulated to develop suitable substitutes, and today we have a host of their inventions, some of which are superior to natural rubber for many uses. Necessity may have been the mother of invention, but she was not the inventor. I would liken the curtailment of imports, surely not a creative process, to the extinction of the dinosaurs, and the efforts of the scientists and engineers, which certainly were creative, to the selection of alternative alleles within the mammalian populations. In this attitude I ally myself with Simpson (1944) and against Wright (1945), who argued that the extinction of the dinosaurs, since it may have aided the adaptive radiation of the mammals, should be regarded as a creative process.

Group selection is the only conceivable force that could produce biotic adaptation. It was necessary, therefore, in this discussion of biotic adaptation to examine the nature of group selection and to attempt some preliminary evaluation of its power. The issue, however, cannot be resolved on the basis of hypothetical examples and appeals to intuitive judgments as to what seems likely or unlikely. A direct assessment of the importance of group selection would have to be based on an accurate knowledge of rates of genetic change, due to different causes, within populations; rates of proliferation and extinction of populations and larger groups; relative and absolute rates of migration and interbreeding; relative and absolute values of the coefficients of genic and group selection; etc. We would need such information for a large and unbiased sample of present and past taxa. Obviously this ideal will not be met, and some indirect method of evaluation will be necessary. The only method that I can conceive of as being reliable is an examination of the adaptations of animals and plants to determine the nature of the goals for which they are designed. The details of the strategy being employed will furnish indications of the purpose of its employment. I can conceive of only two ultimate purposes as being indicated, genic survival and group survival. All other kinds of survival, such as that of individual somata, will be of the nature of tactics employed in the grand strategy, and such tactics will be employed only when they do, in fact, contribute to the realization of a more general goal.

The basic issue then is whether organisms, by and large, are using strategies for genic survival alone, or for both genic and group survival. If both, then which seems to be the predominant consideration? If there are many adaptations of obvious group benefit which cannot be explained on the basis of genic selection, it must be conceded that group selection has been operative and important. If there are no such adaptations, we must conclude that group selection has not been important, and that only genic selection—natural selection in its most austere form—need be recognized as the creative force in evolution. We must always bear in mind that group selection and biotic adaptation are more onerous principles than genic selection and organic adaptation. They should only be invoked when the simpler explanation is clearly inadequate. Our search must be specifically directed at finding adaptations that promote group survival but are clearly neutral or detrimental to individual reproductive survival in within-group competition. The criteria for the recognition of these biotic adaptations are essentially the same as those for organic adaptations. The system in question should produce group benefit in an economical and efficient way and involve enough potentially independent elements that mere chance will not suffice as an explanation for the beneficial effect.

The examples considered above all related to interactions between individuals, and the important consideration was to find a parsimonious explanation of why one individual would expend its own resources or endanger itself in an attempt to aid another. There remain a number of examples of individuals' acting, at their own expense, in a manner that benefits their conspecific neighbors in general, not specific individuals. Such activity can take place only when the animals occur in unrelated groups larger than two. The important initial problem is why animals should exist in groups of several to many individuals.

It is my belief that two basic misconceptions have seriously hampered progress in the study of animals in groups. The first misconception is the assumption that when one demonstrates that a certain biological process produces a certain benefit, one has demonstrated *the* function, or at least *a* function of the process. This is a serious error. The demonstration of a benefit is neither necessary nor sufficient in the demonstration of function, although it may sometimes provide insight not otherwise obtainable. It is both necessary and sufficient to show that the process is designed to serve the function. A relevant example is provided by Allee (1931). He observed that a certain marine flatworm, normally found in aggregated groups, can be killed by placement in a hypotonic solution. The harmfulness of such a solution is reduced when large numbers of worms, not just one or a few, are exposed to it. The effect is caused by the liberation of an unknown substance from the worms, especially dead ones, into the water. The substance is not osmotically important in itself, but somehow protects the worms against hypotonicity. Allee saw great significance in this observation, and assumed that he had demonstrated that a beneficial chemical conditioning of the environment is a function of aggregation in these worms. The fallacy of such a conclusion should be especially clear when it relates to very artificial situations like placing large numbers of worms in a small volume of brackish water. The kind of evidence that would be acceptable would be the demonstration that social cohesion increased as the water became hypotonic or underwent some other chemically harmful change; that specific integumentary secretory machinery was activated by the deleterious change; that the substance secreted not only provided protection against hypotonicity, but was an extraordinarily effective substance for this protection. One or two more links in such a chain of circumstances would provide the necessary evidence of functional design and leave no doubt that protection from hypotonicity was a function of aggregation, and not merely an effect.

The second misconception is the assumption that to explain the functional aspects of groups, one must look for group functions. An analogy with human behavior will illustrate the nature of this fallacy. Suppose a visitor from Mars, unseen, observed the social behavior of a mob of panic-stricken people rushing from a burning theatre. If he was burdened with the misconception in question he would assume that the mob must show some sort of an adaptive organization for the benefit of the group as a whole. If he was sufficiently blinded by this assumption he might even miss the obvious conclusion that the observed behavior could result in total survival below what would have resulted from a wide variety of other conceivable types of behavior. He would be impressed by the fact that the group showed a rapid "response" to the stimulus of fire. It went rapidly from a widely dispersed distribution to the formation of dense aggregations that very effectively sealed off the exits.

Someone more conversant with human nature, however, would find the explanation not in a functioning of the group, but in the functioning of individuals. An individual finds himself in a theatre in which a dangerous fire has suddenly broken out. If he is sitting near an exit he may run for it immediately. If he is a bit farther away he sees others running for the exits and, knowing human nature, realizes that if he is to get out at all he must get out quickly; so he likewise runs for the door, and in so doing, intensifies the stimulus that will cause others to behave in the same way. This behavior is clearly adaptive from the standpoint of individual genetic survival, and the behavior of the mob is easily understood as the statistical summation of individual adaptation.

This is an extreme example of damage caused by the social consequences of adaptive behavior, but undoubtedly such effects do occur, and they may be fairly common in some species. There are numerous reports, at least at the anecdotal level, of the mass destruction of large ungulates when individuals in the van of a herd are pushed off cliffs by the press from the rear. Less spectacular examples of harm deriving from social grouping are probably of greater significance. I would imagine the most important damage from social behavior to be the spread of communicable disease.

The statistical summation of adaptive individual reactions, which I believe to underlie all group action, need not be harmful. On the contrary, it may often be beneficial, perhaps more often than not. An example of such a benefit would be the retention of warmth by close groups of mammals or birds in cold weather, but there is no more reason to assume that a herd is designed for the retention of warmth than to assume that it is designed for transmitting diseases. The huddling behavior of a mouse in cold weather is designed to minimize its own heat loss, not that of the group. In seeking warmth from its neighbors it contributes heat to the group and thereby makes the collective warmth a stronger stimulus in evoking the same response from other individuals. The panic-stricken man in the theatre contributed to the panic stimulus in a similar fashion. Both man and mouse probably aid in the spread of disease. Thus the demonstration of effects, good or bad, proves nothing. To prove adaptation one must demonstrate a functional design.

References

Allee, W. C. 1931. *Animal Aggregations: A Study in General Sociology*. Chicago: University of Chicago Press.

———. 1940. "Concerning the origin of sociality in animals." *Scientia* 1940: 154–160.

Allee, W. C., Alfred E. Emerson, Orlando Park, Thomas Park, and Karl P. Schmidt. 1949. *Principles of Animal Ecology*. Philadelphia: W. B. Saunders.

Altman, Stuart A. 1962. "A field study of the sociobiology of rhesus monkeys, *Macaca mulatta*." *Ann. N. Y. Acad. Sci.* 102: 338–435.

Clarke, C. A., C. G. G. Dickson, P. M. Sheppard. 1963. "Larval color pattern in *Papilio demodocus*." *Evolution* 17: 130–137.

Darwin, Charles R. 1896. *The Descent of Man and Selection in Relation to Sex*. New York: D. Appleton.

Fisher, R. A., E. B. Ford. 1947. "The spread of a gene in natural conditions in a colony of the moth, *Panaxia dominula* (L)." *Heredity* 1: 143–174.

Ford, E. B. 1956. "Rapid evolution and the conditions which make it possible." *Cold Spring Harbor Symp. Quant. Biol.* 20: 230–238.

Haldane, J. B. S. 1932. *The Causes of Evolution*. London: Longmans.

Hamilton, W. D. 1964. "The genetical evolution of social behaviour, I." *J. Theoret. Biol.* 7: 1–16.

Lewontin, R. C. 1958. "The adaptations of populations to varying environments." *Cold Spring Harbor Symp. Quant. Biol.* 22: 395–408.

———. 1961. "Evolution and the theory of games." *J. Theoret. Biol.* 1: 382–403.

———. 1962. "Interdeme selection controlling a polymorphism in the house mouse." *Am. Naturalist* 96: 65–78.

Lewontin, R. C., and L. C. Dunn. 1960. "The evolutionary dynamics of a polymorphism in the house mouse." *Genetics* 45: 705–722.

Montagu, M. F. Ashley. 1952. *Darwin, Competition and Cooperation*. New York: Henry Schuman.

Russell, E. S. 1945. *The Directiveness of Organic Activities*. Cambridge: Cambridge University Press.

Simpson, George Gaylord. 1944. *Tempo and Mode in Evolution*. New York: Columbia University Press.

———. 1953. *The Major Features of Evolution*. New York: Columbia University Press.

Slijper, E. J. 1962. *Whales*. Trans. A. J. Pomerans, New York: Basic Books.

Slobodkin, L. Basil. 1954. "Population dynamics of *Daphnia obtusa* Kurz." *Ecol. Monog.* 24: 69–88.

Williams, George C. 1966. *Adaptation and Natural Selection*. Princeton: Princeton University Press.

Wright, Sewall. 1945. "*Tempo and mode in evolution*: A critical review." *Ecology* 26: 415–419.

Wynne-Edwards, V. C. 1962. *Animal Dispersion in Relation to Social Behaviour*. Edinburgh and London: Oliver and Boyd.

4 Levels of Selection: An Alternative to Individualism in Biology and the Human Sciences

David Sloan Wilson

Biology and many branches of the human sciences are dominated by an individualistic tradition that treat groups and communities as collections of organisms without themselves having the properties implicit in the word "organism." In biology, the individualistic tradition achieves generality only by defining self-interest as "anything that evolves by natural selection." A more meaningful definition of self-interest shows that natural selection operates on a hierarchy of units from genetic elements to multispecies communities, and that a unit becomes organismic to the degree that natural selection operates at the level of that unit. I review levels-of-selection theory in biology and sketch a parallel argument for the human sciences.

Introduction

The related concepts of adaptation, function, intention and purpose are central to both biology and the human sciences. Natural selection endows species with the functional design required to survive and reproduce in their environments. Humans organize their behavior to achieve various proximate goals in their everyday lives.

Biology and the human sciences also share a controversy over the units that can be said to have the properties of adaptation, function, intention, and purpose. Almost everyone would grant these properties to individuals, but some biologists also speak of social groups and multi-species communities as if they were single purposeful organisms. Similarly, some psychologists, anthropologists, and sociologists speak of culture and society as superorganisms in which individuals are mere cells.

In recent decades the hierarchical view of functional organization has fallen on hard times. Larger entities are regarded as mere collections of organisms, without themselves having the properties of organisms. In biology the reductionistic trend has proceeded so far that even individuals are sometimes treated as upper units of the hierarchy, mere collections of "selfish" genes (Dawkins 1976, 1982). The human sciences are more heterogeneous, but many of its branches appear to be dominated by the individualistic view.

From *Social Networks* 11 (1989): 257–272. Elsevier Science Publishers B. V., North Holland.

Despite its widespread acceptance, the case for individualism as a general prediction that emerges from evolutionary theory, or as a general principle to explain human behavior, actually is very frail. In this chapter I will describe why functional organization in nature is necessarily hierarchical and then will attempt to sketch a parallel argument for the human sciences.

The Evolution of Altruism

In biology, the debate over units of adaptation has centered on the evolution of seemingly altruistic behaviors that benefit others at the expense of the self. Consider a population of N individuals. Two types exist, A and S, in proportions p and $(1-p)$, respectively. Each A-type expresses a behavior toward a single recipient, chosen at random from the population. As a result, the recipient has an additional number b of offspring while the altruist has c fewer offspring. The average number of offspring, W, can then be calculated for each type.

$$W_A = X - c + b(Np-1)/(N-1), \qquad W_S = X + bNp/(N-1) \tag{1}$$

X is the number of offspring in the absence of altruistic behaviors, and is the same for both types. In addition to the cost of being an altruist, each A-type can serve as a recipient to the $(Np-1)$ other altruists who are distributing their benefits among $(N-1)$ individuals in the group. Selfish S-types have no cost of altruism and can serve as recipients to all Np altruists in the group. S-types have more offspring than A-types whenever $W_S > W_A$, which reduces to the inequality.

$$b/(N-1) > -c. \tag{2}$$

This inequality always holds, because b, c, and N are positive numbers and N is greater than 1. Thus, selfish types always have more offspring than altruistic types. To the degree that the behaviors are heritable, selfish types will be found at a greater frequency in the next generation.

A numerical example is shown in table 4.1, in which $N = 100$, $p = 0.5$, $X = 10$, $b = 5$, and $c = 1$. Thus, the altruist bestows an additional 5 offspring on the recipient at a cost of 1 offspring to itself. The average altruist has 11.47 offspring, while the average selfish type has 12.53 offspring. Assume that the types reproduce asexually, such that the offspring exactly resemble the parents. The proportion of altruists among the progeny is then $p' = 0.478$, a decline from the parental value of $p = 0.5$. Since populations cannot grow to infinity, we also assume that mortality occurs equally among the A- and S-types, returning the population to a density of $N = 100$. At this point we expect approximately 52 selfish and 48 altruistic types. If this procedure is iterated many times, representing natural selection acting over many generations, the A-types continue to decline in frequency and ultimately become extinct.

Table 4.1
Evolution in a single population

$N = 100,\ p = 0.5, X = 10, b = 5, c = 1$
$W_A = X - c + b(Np - 1)/(N - 1) = 10 - 1 + 49(5)/99 = 11.47$
$W_S = X + bNp/(N - 1) = 10 + 50(5)/99 = 12.53$
$N' = N(pW_A + (1 - p)W_S) = 100(0.5(11.47) + 0.5(12.53)) = 1200$
$p' = NpW_A/N' = 100(0.5)(11.47)/1200 = 0.478$

Note: The altruistic type declines from a frequency of $p = 0.5$ before selection to a frequency of $p' = 0.478$ after selection.

This is the paradox that makes altruism such a fascinating subject for evolutionary biologists. As humans we would like to think that altruism can evolve, as biologists we see animal behaviors that appear altruistic in nature, yet almost by definition it appears that natural selection will act against them. This is the sense in which evolution appears to be an inherently selfish theory.

The paradox, however, can be resolved by a simple alteration of the model. Table 4.2 differs from table 4.1 in only two respects: (1) we now have two groups instead of one; and (2) the groups have different proportions of altruistic and selfish types. Looking at each group separately, we reach the same conclusion as for table 4.1; selfish types have more offspring than altruistic types. Adding the individuals from both groups together, however, we get the opposite answer: altruistic types have more offspring than selfish types.[1]

What has happened to produce this interesting (and for many people counterintuitive) result? First, there must be more than one group; there must be a *population of groups*. Second, the groups cannot all have the same proportion of altruistic types, for then the results would not differ from a single group. The groups must *vary* in the proportion of altruistic types. Third, there must be a direct relationship between the proportion of altruists and the total number of offspring produced by the group; groups of altruists must be *more fit* than groups without altruists. These are the necessary conditions for the evolution of altruism in the elaborated model. To be sufficient, the differential fitness of groups—the force favoring the altruists—must be great enough to counter the differential fitness of individuals within groups—the force favoring the selfish types.

Readers familiar with evolutionary theory immediately will recognize a similarity between the above conditions and Darwin's original theory of natural selection, which requires a *population of individuals*, that *vary* in their genetic composition, with some variants *more fit* than others. Thus, natural selection can operate simultaneously at more than one level. Individual selection promotes the fitness of individuals relative to others in the same group. Group selection promotes the fitness of groups, relative to

Table 4.2

Evolution in two groups that differ in the proportion of the altruistic type

Group 1	Group 2
$N_1 = 100,\ p_1 = 0.2$	$N_2 = 100,\ p_2 = 0.8$
$W_A = 10 - 1 + 19(5)/99 = 9.96$	$W_A = 10 - 1 + 79(5)/99 = 12.99$
$W_S = 10 + 20(5)/99 = 11.01$	$W_S = 10 + 80(5)/99 = 14.04$
$N_1' = 1080$	$n_2' = 1320$
$p_1' = 0.184$	$p_2' = 0.787$
Global population	
$N = 200, P = 0.5$	
$N' = N_1' + N_2' = 2400$	
$P' = (N_1'p_1' + N_2'p_2')/(N_1' + N_2') = 0.516$	

Note: Values for X, b, c and the functions for W_A and W_S are provided in Table 1. The altruistic type declines in frequency within each group (compare p_1' with p_1 and p_2' with p_2) but increases in frequency when both groups are considered together (compare P' with P). This is because group 2, with the most altruists, is more productive than group 1 (compare N_2' with N_1').

other groups in the global population. These levels of selection are not always in conflict. A single behavior can benefit both the individual performing it and others in the group. Altruistic behaviors by definition are costly to self and beneficial to others, however, and so are favored by group selection and disfavored by individual selection.

This simple numerical example shows that the process of natural selection does not inevitably evolve selfish behaviors. A notion of *group-interest* must be added to the notion of *self-interest*, to the extent that group selection is important in nature.

Valid Individualism and Cheap Individualism

Let us now consider the individualistic claim that "virtually all adaptations evolve by individual selection." If by individual selection we mean within-group selection, we are saying that *A*-types virtually never evolve in nature, that we should observe only *S*-types. This is a meaningful statement because it identifies a set of traits that conceivably could evolve, but does not, because between-group selection is invariably weak compared to within-group selection. Let us call this *valid individualism*.

There is, however, another way to calculate fitness in the two-group model that leads to another definition of individual selection. Instead of separately considering evolution within groups and the differential fitness of groups, we can directly average the fitness of *A*- and *S*-types across all groups. Thus, the 2 *A*-types in groups one have 9.96 offspring and the 8 *A*-types in group two have 12.99 offspring, for an average fitness of $0.2(9.96) + 0.8(12.99) = 12.38$. The 8 *S*-types in group one have 11.01 offspring and the 2 *S*-types in group two have 14.04 offspring, for an average fitness of $0.8(11.01) + 0.2(14.04) = 11.62$. The average *A*-type individual is more fit then the average *S*-type individual, which is merely another way of saying that it evolves.

Let us now return to the individualistic claim that "virtually all adaptations evolve by individual selection." If by individual selection we mean the fitness of individuals averaged across all groups, we have said nothing at all. Since this definition includes both within- and between-group selection, it makes "individual selection" synonymous with "whatever evolves," including either *S*-types or A-types. It does not identify any set of traits that conceivably could evolve but does not. Let us therefore call it *cheap individualism*.

Cheap individualism is so meaningless that no one would explicitly endorse it. Even the most ardent individualists, such as G. C. Williams (1966, 1985), R. Dawkins (1976, 1982), and J. Maynard Smith (1987), believe that there is something outside individual selection called group selection that in principle can evolve altruistic traits. Nevertheless, the history of individual selection from 1960 to the present has been a slow slide from valid individualism to cheap individualism. Before documenting this claim it is necessary to review three reasons why the slide could occur unnoticed.

First, group-structured population models such as the one described above can be applied to an enormous range of biological phenomena. The single groups can be isolated demes that persist for many generations, groups of parasites interacting within single hosts, clusters of caterpillars interacting on a single leaf, or coalitions of baboons that behaviorally segregate within a larger troop. The groups can be communities whose members are separate species, social units whose members are conspecifics, or even single organisms whose "members" are genes of cell lineages (Crow 1979; Cosmides and Tooby 1981; Buss 1987). Historically, however, the first group selection models focused on a particular conception of isolated demes that persist for many generations. Thus, it has been possible for biologists studying other kinds of groups to assume that they are not invoking group selection, when in fact their models are miniature versions of traditional group selection models.

Second, many biologists today regard group selection as a heretical concept that was discarded twenty years ago and consider their own work to be entirely within the grand tradition of "individual selection." Gould (1982:xv) remembers "the hooting dismissal of Wynne-Edwards and group selection in any form during the late 1960's and most of the 1970's," and even today graduate students tell me how difficult it is for them to think about group selection in a positive light after being taught in their courses that it "just doesn't happen." The vast majority of authors who claim that such-and-such evolves by individual selection do not even include an explicit model of group selection to serve as a possible alternative. Individual selection truly has become the modern synonym for "everything that evolves in my model," and group selection is mentioned only as a bogey man in the introduction or the conclusion of the paper.

Third, averaging the fitness of individual types across groups is a useful, intuitively reasonable procedure that correctly predicts the outcome of natural selection. Biologists commonly average the fitness of types across a range of physical environments, and it seems reasonable to average across social environments in the same way. I emphasize that there is nothing wrong with this procedure—it merely cannot be used to define individual selection because it leaves nothing outside of it.

Now I must document my claim that individualism in biology achieves generality only by averaging the fitness of individuals across groups.

Three Examples of Cheap Individualism in Biology

The Evolution of Avirulence in Parasites and Diseases

Disease organisms provide an excellent real-world example of a group-structured population similar to the model outlined above. Each infected host comprises an isolated group of disease organisms, which compete with other groups to infect new hosts. Natural selection within single hosts is expected to favor strains with high growth rates. Excessively high growth rates tend to kill the host, however, driving the entire group of disease organisms extinct (assuming that transmission requires the host to be alive). Avirulent strains therefore can be envisioned as "altruists" that increase the survival of entire groups, but which nevertheless decline in frequency within every group containing more virulent strains. Lewontin (1970) was the first to recognize that avirulence evolves by between-group selection, and the process has been well documented in a *myxoma* virus that was introduced into Australia to control the European rabbit (Fenner and Ratcliffe 1965). Nevertheless, consider the following account in the first edition of Futuyma's (1979 : 455) textbook *Evolutionary Biology*:

> In many interactions the exploiter cannot evolve to be avirulent; it profits a fox nothing to spare the hare. But if the fitness of an individual parasite or its offspring is lowered by the death of its host, avirulence is advantageous. The *myxoma* virus, introduced into Australia to control European rabbits, at first caused immense mortality. But within a few years mortality levels were lower, both because the rabbits had evolved resistance and because the virus had evolved to be less lethal. . . . Because the virus is transmitted by mosquitoes that feed only on living rabbits, *virulent virus genotypes are less likely to spread than benign genotypes* [italics mine]. Avirulence evolves not to assure a stable future supply of hosts, but to benefit individual parasites.

Thus, by the simple procedure of comparing the fitness of virulent and avirulent types across all hosts (see italicized portion of text), rather than within single hosts, the evolution of avirulence can be made to appear an individualistic process. Futuyma, incidently, is sympathetic to the concept of group selection and properly attributes avirulence to between-group selection in the second edition of his textbook (1986:496–497). This example of cheap individualism therefore is inadvertent, and shows how easily selection at multiple levels can be represented as occurring entirely at the lowest level.

Inclusive Fitness Theory

Within the individualistic tradition in biology, natural selection is widely thought to maximize a property called inclusive fitness, which is the sum of an individual's effects on the fitness of others multiplied by the probability that the others will share the genes causing the behavior. As Hamilton (1963:354–355) originally put it:

> Despite the principle of "survival of the fittest" the ultimate criterion which determines whether G [an altruistic allele] will spread is not whether the behavior is to the benefit of the behavior but whether it is to the benefit of the gene G; and this will be the case if the average net result of the behavior is to add to the gene-pool a handful of genes containing G in higher concentration than does the gene-pool itself. With altruism this will happen only if the affected individual is a relative of the altruist, therefore having an increased chance of carrying the gene, and if the advantage conferred is large enough compared to the personal disadvantage to offset the regression, or "dilution," of the altruist's genotype in the relative in question.

In this formulation, individuals evolve to maximize the fitness of "their genes" relative to other genes in the population, regardless of whether "their genes" are located in children, siblings, cousins, parents, and so on. Aid-giving toward relatives therefore ceases to appear altruistic, and becomes part of an individual's "selfish" strategy to maximize its inclusive fitness. Even sterility and death can be inclusive fitness maximizing if the positive effects on relatives are sufficiently great.

Let us pursue this idea by considering an Aa female who mates with an aa male and produces a clutch of ten offspring, five of whom are Aa and the other five aa. The dominant allele A codes for an altruistic behavior that is expressed only toward siblings. The sibling group therefore is equally divided between altruists and nonaltruists, and the fitness of the two genotypes from equation (1) is

$$W_{\mathrm{Aa}} = X - c + b(4/9), \qquad W_{\mathrm{aa}} = X + b(5/9).$$

The selfish aa genotype is inevitably most fit, which merely reiterates the general conclusion obtained [previously] for evolution in all single groups. The fact that the group in this case consists of full siblings is irrelevant to the conclusion. To see how altruism expressed toward siblings evolves, we must consider a large number of family groups, initiated by all combinations of parental genotypes −AA × AA, AA × Aa, Aa × Aa, AA × aa, Aa × aa, aa × aa. Within-group selection favors the selfish a-allele in all groups containing both altruistic and selfish genotypes. The fitness of entire sibling groups, however, is directly proportional to the frequency of altruistic A-alleles in the group.

Thus, Hamilton's conclusions cannot be reached without combining within-group selection and between-group selection into a single measure of ''inclusive fitness.''

The idea that aid-giving toward relatives is a form of ''true'' altruism that requires between-group selection has been reached by many authors (reviewed in Wilson 1983). Nevertheless, evolutionists within the individualistic tradition continue to use inclusive fitness theory as their guiding light to explain the evolution of ''apparently'' altruistic behaviors, ''without invoking group selection.'' This is cheap individualism.

Diploid Population Genetics and Evolutionary Game Theory

My final example involves a comparison between two seemingly different bodies of theory in evolutionary biology. Diploid population genetics models begin with a population of gametic types (A, a) which combine into pairs to form diploid genotypes (AA, Aa, aa). Selection usually is assumed to occur in the diploid stage, after which the genotypes dissociate back into gametes and the process is reiterated. The most common way for selection to occur in these models is for some genotypes to survive and reproduce better than others, the standard process of between-individual selection. In addition, however, it is possible for some alleles to survive and reproduce better than others *within single individuals*. For example, the rules of meiosis usually cause the two chromosome sets to be equally represented in the gametes. Some alleles manage to break the rules of meiosis, however, biasing their own transmission into the sperm and eggs of heterozygotes. The differential fitness of alleles within heterozygotes is termed meiotic drive, and can cause the evolution of genes that have neutral or even deleterious effects on the fitness of individuals (Crow 1979; Cosmides and Tooby 1981). In short, diploid population genetics models are explicitly hierarchical by recognizing the existence of both between- and within-individual selection.

Evolutionary game theory (also called ESS theory for ''evolutionarily stable strategy'') begins with a population of individual types (A, a) that combine into groups of size N for purposes of interaction. Selection occurs during the grouped stage, after which the groups dissociate back into individuals and the process is reiterated. Usually $N = 2$, which yields three types of groups (AA, Aa, aa). ESS theory was borrowed directly from economic game theory (Maynard Smith and Price 1973; Maynard Smith 1982) but the two are not identical. In particular, economic game theory assumes that the players are rational actors trying to maximize their (absolute) payoff, while ESS theory assumes that natural selection will favor the strategy that delivers the highest payoff relative to other competing strategies in the population.

It should be obvious that the population structure of genes combining into individuals in a diploid model is identical to the population structure of individuals combining into groups of $N = 2$ in an ESS model. Similarly, natural selection in an ESS model can happen in two ways: groups can outperform other groups or individuals can outperform other individuals within groups. In the familiar hawk-dove model, for example, dove-dove groups (in which resources are equitably shared) are more fit than hawk-hawk groups (in which resources are contested), while hawks are more fit than doves within hawk-dove groups. To be consistent with population genetics models we should say that hawks are favored by within-group selection and doves by between-group selection. ESS theorists, however, average the fitness of individual types across groups and call everything that evolves the product of "individual selection." The term "between-group selection" is never used, and Maynard Smith actually borrowed game theory from economics as an alternative to group selection (Maynard Smith and Price 1973; Maynard Smith 1982). As Dawkins (1980 : 360) puts it: "There is a common misconception that cooperation within a group at a given level of organization must come about through selection between groups. . . . ESS theory provides a more parsimonious alternative." This one passage provides all the elements of cheap individualism: the fitness of individuals is averaged across groups, everything that evolves is called the product of individual selection, and something else is called group selection, outside the model and completely unspecified, except to say that it need not be invoked.

These three examples show that, despite its widespread acceptance, individualism in biology is on very thin ice. Self-interest defined as "whatever evolves" is meaningless, and yet when self-interest is defined more meaningfully as "within-group selection" it cannot claim to explain everything that evolves in nature. We must therefore accept a hierarchical view of evolution in which the properties of functional organization implicit in the word "organism" need not be restricted to individuals. The differential fitness of genetic elements within individuals ushers us into a bizarre world in which the genetic elements are the purposeful organisms and individuals are mere collections of quarreling genes, the way we usually think of groups. The differential fitness of individuals within groups ushers us into a familiar world in which groups are mere collections of purposeful individuals. The differential fitness of groups ushers us into another bizarre world (for individualists) in which the groups are the organisms whose properties are caused by individuals acting in a coordinated fashion, the way we usually think of genes and the organs they code for. See Wilson and Sober (1989) for a more detailed review of levels-of-selection theory in biology.

A Parallel Argument for the Human Sciences

If human behavior is measured against the dual standard of effects on self and effects on others, it appears to show the full range of potential. Individuals have sacrificed their lives for the benefit of others, and they have sacrificed the lives of others for their own trivial gain. Viewed at the society level, some human groups are so well coordinated that they invite comparison to single organisms, while others show all the disorganization of a bar-room brawl.

Humans also are frequently embedded in a complex network of interactions in which single expressions of a behavior affect the actor and a relatively small number of associates. Put another way, human populations are subdivided into clusters of associates similar to the local populations of the evolutionary models outlined above. It seems possible that a theory of human behavior in social networks could be developed that parallels levels-of-selection theory in biology, leading to a similar hierarchical view of functional organization in human affairs.

As with any theory of human behavior, the first step is to specify the rules that cause people to choose among alternative behaviors, which serve as the analog of natural selection in an evolutionary model. Following Axelrod and others (Axelrod and Hamilton 1981; Brown et al. 1982; Pollock 1988), assume that humans adopt behaviors that maximize a given utility, relative to competing behaviors in the population. The utility might be pleasure (to a psychologist), annual income (to an economist), or genetic fitness (to a sociobiologist). The details of the utility are relatively unimportant because the hallmark of a hierarchical model is not the nature of the utility but the way it is partitioned into within- and between-group components. Consider, for example, a behavior that decreases the utility of self and increases the utility of others. If others include the entire population, then the utility of those expressing the behavior will be lower than those that do not, and the behavior will be rejected precisely as it is selected against in the one-group evolutionary model. Now assume that the human population is subdivided into a mosaic of associates in which the expression of behavior is nonrandom; some groups of associates behave primarily one way, other groups the other way. The utility of the behaviors now depends on the frame of comparison. The behavior fares poorly in all groups in which the alternative behavior is expressed, but may still deliver the highest utility when averaged across all groups, exactly as in the multigroup evolutionary model. Adoption of the behavior therefore depends on two factors, the effect on self and others and the interaction structure within which the behavior is embedded.

Theories of behavior in the human sciences frequently consider both factors but combine them into an overarching definition of self-interest as ‘‘utility-maximizing behavior’’—i.e., all behaviors adopted by rational humans! This is cheap individualism, that achieves generality only by definitional fiat. Levels-of-selection theory keeps the factors separate, defining behaviors as self-interested when they increase relative utility within single groups, and group-interested when they increase the average utility of groups, relative to other groups. This provides a framework in which rational (utility maximizing) humans need not be self-interested by definition.

As for the situation in biology, many human behaviors that are catagorized as selfish by cheap individualism emerge as ‘‘groupish’’ in a levels-of-selection model.[2] The concept of morality, for example, involves rules of conduct that promote the common good. This implies a category of immoral behaviors—frequently termed "selfish" in everyday language—that benefit individuals at the expense of the common good. Since moral behaviors are vulnerable to exploitation, they succeed only if they can be segregated from the expression of immoral behaviors. This is nicely illustrated by the following passage from a seventeenth-century Hutterite document (English translation in Ehrenpreis 1978:67):

> The bond of love is kept pure and intact by the correction of the Holy Spirit. People who are burdened with vices that spread and corrupt can have no part in it. This harmonious fellowship excludes any who are not part of the unanimous spirit. . . . If a man hardens himself in rebellion, the extreme step of separation is unavoidable. Otherwise the whole community would be dragged into his sin and become party to it. . . . The Apostle Paul therefore says "Drive out the wicked person from among you."

The maintenance of behaviorally pure groups allowed the Hutterites to practice such extreme altruism that their communities are best regarded as the human equivalent of a bee colony (a metaphor that they themselves used to describe themselves). More generally, human societies everywhere possess mechanisms for segregating behaviors, allowing less extreme forms of morally acceptable behavior to be successful. The distinction between moral and immoral behavior, and the mechanisms whereby both can be advantageous, correspond nicely to "groupish" and "selfish" behaviors in a levels-of-selection model. In contrast, cheap individualism is placed in the awkward situation of defining both moral and immoral behavior as brands of self-interest.

Many authors have expressed the idea that higher entities such as biological communities and human societies can be organisms in their own right. Unfortunately, the idea usually is stated as a poetic metaphor or as an axiom that is not subject to disproof. Levels-of-selection theory shows that single-species groups and multispecies communities can become functionally organized by the exact same process of between-unit selection that causes the groups of genes known as individuals to become functionally organized. For the first time, the hierarchical view in biology now enjoys a solid mechanistic foundation. Perhaps this foundation also will be useful within the human sciences to show how people sometimes coalesce into society-level organisms.

Notes

This research was funded from a J. S. Guggenheim fellowship. I thank G. Pollock, R. Boyd, P. Richerson, and virtually dozens of other people for helpful conversations.

1. Adding the contents of both groups is justified biologically only if the occupants of the groups physically mix during a dispersal stage or compete for the colonization of new groups. See Wilson (1977, 1980, 1983) for a more detailed discussion of the nature of groups in levels-of-selection models.

2. Both cheap individualism and levels-of-selection models define their terms on the basis of utilities, which do not translate easily into psychological definitions of altruism and selfishness based on internal motivation. In outlining his economic theory of human behavior, Becker (1976:7) states that it does not matter how people actually feel or think about what they do as long as the end result of their behavior is utility maximizing. In the same way, behaviors categorized as group interested in a levels-of-selection model do not imply that the actor is internally motivated to help others. This does not mean that psychological definitions of altruism are irrelevant, but only that their relationship with definitions based on utility are complex. I hope to explore the complexities in a future paper.

References

Axelrod, R., and W. D. Hamilton. 1981. "The evolution of cooperation." *Science* 211: 1390–1396.

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

Brown, J. S., M. J. Sanderson, and R. E. Michod. 1982. "Evolution of social behavior by reciprocation." *Journal of Theoretical Biology* 99: 319–339.

Buss, L. W. 1987. *The Evolution of Individuality.* Princeton: Princeton University Press.

Comides, L. M., and J. Tooby. 1981. "Cytoplasmic inheritance and intragenomic conflict." *Journal of Theoretical Biology* 89: 83–129.

Crow, J. F. 1979. "Genes that violate Mendel's rules." *Scientific American* 240: 104–113.

Dawkins, R. 1976. *The Selfish Gene.* Oxford: Oxford University Press.

———. 1980. "Good strategy or evolutionary stable strategy?" In G. W. Barlow and J. Silverberg (eds.), *Sociobiology: Beyond Nature/Nurture?* pp. 331–367. Boulder. CO: Westview Press.

———. 1982. *The Extended Phenotype.* Oxford: Oxford University Press.

Ehrenpreis, A. 1978. *Brotherly Community: The Highest Command of Love.* Rifton, NY: Plough Publishing Co.

Fenner, F., and F. N. Ratcliffe. 1965. *Myxomatosis.* London: Cambridge University Press.

Futuyma, D. J. 1979. *Evolutionary Biology* (first edn.). Sunderland, MA: Sinauer Press.

———. 1986. *Evolutionary Biology* (second edition). Sunderland, MA: Sinauer Press.

Gould, S. J. 1982. *The Uses of Heresy: An Introduction to Richard Goldschmidt's The Material Basis of Evolution*, pp. xiii–xlii. New haven, CT: Yale University Press.

Hamilton, W. D. 1963. "The evolution of altruistic behavior." *American Naturalist* 97: 354–356.

Lewontin, R. C. 1970. "The units of selection." *Annual Review of Ecology and Systematics* 1: 1–18.

Maynard Smith, J. 1982. *Evolution and the Theory of Games.* Cambridge: Cambridge University Press.

———. 1987. “How to model evolution.” In J. Dupre (ed.), *The Latest on the Best: Essays on Evolution and Optimality*, pp. 117–131. Cambridge, MA: MIT Press.

———, and G. R. Price. 1973. “The logic of animal conflict.” *Nature* 246: 15–18.

Pollock, G. B. 1988. “Population structure, spite, and the iterated prisoner’s dilemma.” *American Journal of Physical Anthropology* 77: 459–469.

Williams, G. C. 1966. *Adaptation and Natural Selection*. Princeton: Princeton University Press.

———. 1985. “A defense of reductionism in evolutionary biology.” In R. Dawkins and M. Ridley (eds.), *Oxford Surveys in Evolutionary Biology*. Vol. 2, pp. 1–27. Oxford: Oxford University Press.

Wilson, D. S. 1977. “Structured demes and the evolution of group-advantageous traits.” *American Naturalist* 111: 157–185.

———. 1980. *The Natural Selection of Populations and Communities*. Menlo Park, CA: Benjamin-Cummings.

———. 1983. “The groups selection controversy: history and current status.” *Annual Review of Ecology and Systematics* 14: 159–187.

———, and E. Sober. 1989. “Reviving the superorganism.” *Journal of Theoretical Biology* 136: 337–356.

III Adaptationism

5 The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme

Stephen Jay Gould and Richard C. Lewontin

An adaptationist program has dominated evolutionary thought in England and the United States during the past forty years. It is based on faith in the power of natural selection as an optimizing agent. It proceeds by breaking an organism into unitary "traits" and proposing an adaptive story for each considered separately. Trade-offs among competing selective demands exert the only brake upon perfection; nonoptimality is thereby rendered as a result of adaptation as well. We criticize this approach and attempt to reassert a competing notion (long popular in continental Europe) that organisms must be analyzed as integrated wholes, with *Baupläne* so constrained by phyletic heritage, pathways of development, and general architecture that the constraints themselves become more interesting and more important in delimiting pathways of change than the selective force that may mediate change when it occurs. We fault the adaptationist program for its failure to distinguish current utility from reasons for origin (male tyrannosaurs may have used their diminutive front legs to titillate female partners, but this will not explain why they got so small); for its unwillingness to consider alternatives to adaptive stories; for its reliance upon plausibility alone as a criterion for accepting speculative tales; and for its failure to consider adequately such competing themes as random fixation of alleles, production of nonadaptive structures by developmental correlation with selected features (allometry, pleiotropy, material compensation, mechanically forced correlation), the separability of adaptation and selection, multiple adaptive peaks, and current utility as an epiphenomenon of nonadaptive structures. We support Darwin's own pluralistic approach to identifying the agents of evolutionary change.

Introduction

The great central dome of St. Mark's Cathedral in Venice presents in its mosaic design a detailed iconography expressing the mainstays of Christian faith. Three circles of figures radiate out from a central image of Christ: angels, disciples, and virtues. Each circles is divided into quadrants, even though the dome itself is radially symmetrical in structure. Each quadrant meets one of the four spandrels in the arches below the dome. Spandrels—the tapering triangular spaces formed by the intersection of two rounded arches at right angles (figure 5.1)—are necessary architectural by-products of

From *Proc. R. Soc. Lond.* B 205, 581–598 (1979). Reprinted by permission of The Royal Society.

Figure 5.1
One of the four spandrels of St. Mark's; seated evangelist above, personification of river below.

mounting a dome on rounded arches. Each spandrel contains a design admirably fitted into its tapering space. An evangelist sits in the upper part flanked by the heavenly cities. Below, a man representing one of the four biblical rivers (Tigris, Euphrates, Indus, and Nile) pours water from a pitcher in the narrowing space below his feet.

The design is so elaborate, harmonious, and purposeful that we are tempted to view it as the starting point of any analysis, as the cause in some sense of the surrounding architecture. But this would invert the proper path of analysis. The system begins with an architectural constraint: the necessary four spandrels and their tapering triangular form. They provide a space in which the mosaicists worked; they set the quadripartite symmetry of the dome above.

Figure 5.2
The ceiling of King's College Chapel.

Such architectural constraints abound, and we find them easy to understand because we do not impose our biological biases upon them. Every fan-vaulted ceiling must have a series of open spaces along the midline of the vault, where the sides of the fans intersect between the pillars (figure 5.2). Since the spaces must exist, they are often used for ingenious ornamental effect. In King's College Chapel in Cambridge, for example, the spaces contain bosses alternately embellished with the Tudor rose and portcullis. In a sense, this design represents an "adaptation," but the architectural constraint is clearly primary. The spaces arise as a necessary by-product of fan vaulting; their appropriate use is a secondary effect. Anyone who tried to argue that the structure exists because the alternation of rose and portcullis makes so much sense in a Tudor chapel would be inviting the same ridicule that Voltaire heaped on Dr. Pangloss: "Things cannot be other than they are. . . . Everything is made for the best purpose. Our noses were made to carry spectacles, so we have spectacles. Legs were clearly intended for breeches, and we wear them." Yet evolutionary biologists, in their tendency to focus exclusively on immediate adaptation to local conditions, do tend to ignore architectural constraints and perform just such an inversion of explanation.

As a closer example, recently featured in some important biological literature on adaptation, anthropologist Michael Harner has proposed (1977) that Aztec human sacrifice arose as a solution to chronic shortage of meat (limbs of victims were often consumed, but only by people of high status). E. O. Wilson (1978) has used this explanation as a primary illustration of an adaptive, genetic predisposition for carnivory in humans. Harner and Wilson ask us to view an elaborate social system and a complex set of explicit justifications involving myth, symbol, and tradition as mere epiphenomena generated by the Aztecs as an unconscious rationalization masking the "real" reason for it all: need for protein. But Sahlins (1978) has argued that human sacrifice represented just one part of an elaborate cultural fabric that, in its entirety, not only represented the material expression of Aztec cosmology, but also performed such utilitarian functions as the maintenance of social ranks and systems of tribute among cities.

We strongly suspect that Aztec cannibalism was an "adaptation" much like evangelists and rivers in spandrels, or ornamented bosses in ceiling spaces: a secondary epiphnomenon representing a fruitful use of available parts, not a cause of the entire system. To put it crudely: a system developed for other reasons generated an increasing number of fresh bodies; use might as well be made of them. Why invert the whole system in such a curious fashion and view an entire culture as the epiphenomenon of an unusual way to beef up the meat supply? Spandrels do not exist to house the evangelists. Moreover, as Sahlins argues, it is not even clear that human sacrifice was an adaptation at all. Human cultural practices can be orthogenetic and drive toward extinction in ways that Darwinian processes, based on genetic selection, cannot. Since each new monarch had to outdo his predecessor in even more elaborate and copious sacrifice, the practice was beginning to stretch resources to the breaking point. It would not have been the first time that a human culture did itself in. And, finally, many experts doubt Harner's premise in the first place (Ortiz de Montellano 1978). They argue that other sources of protein were not in short supply, and that a practice awarding meat only to privileged people who had enough anyway, and who used bodies so inefficiently (only the limbs were consumed, and partially at that), represents a mighty poor way to run a butchery.

We deliberately chose nonbiological examples in a sequence running from remote to more familiar: architecture to anthropology. We did this because the primacy of architectural constraint and the epiphenomenal nature of adaptation are not obscured by our biological prejudices in these examples. But we trust that the message for biologists will not go unheeded: if these had been biological systems, would we not, by force of habit, have regarded the epiphenomenal adaptation as primary and tried to build the whole structural system from it?

The Adaptationist Program

We wish to question a deeply engrained habit of thinking among students of evolution. We call it the adaptationist program, or the Panglossian paradigm. It is rooted in a notion popularized by A. R. Wallace and A. Weismann (but not, as we shall see, by Darwin) toward the end of the nineteenth century: the near omnipotence of natural selection in forging organic design and fashioning the best among possible worlds. This program regards natural selection as so powerful and the constraints upon it so few that direct production of adaptation through its operation becomes the primary cause of nearly all organic form, function, and behavior, Constraints upon the pervasive power of natural selection are recognized, of course (phyletic inertia primarily among them, although immediate architectural constraints, as discussed in the last section, are rarely acknowledged). But they are usually dismissed as unimportant or else, and more frustratingly, simply acknowledged and then not taken to heart and invoked.

Studies under the adaptationist program generally proceed in two steps:

1. An organism is atomized into "traits" and these traits are explained as structures optimally designed by natural selection for their functions. For lack of space, we must omit an extended discussion of the vital issue, "What is a trait?" Some evolutionists may regard this as a trivial, or merely a semantic problem, It is not. Organisms are integrated entities, not collections of discrete objects. Evolutionists have often been led astray by inappropriate atomization, as D'Arcy Thompson (1942) loved to point out. Our favorite example involves the human chin (Gould 1977, pp. 381–382; Lewontin 1978). If we regard the chin as a "thing," rather than as a product of interaction between two growth fields (alveolar and mandibular), then we are led to an interpretation of its origin (recapitulatory) exactly opposite to the one now generally favored (neotenic).

2. After the failure of part-by-part optimization, interaction is acknowledged via the dictum that an organism cannot optimize each part without imposing expenses on others. The notion of "trade-off" is introduced, and organisms are interpreted as best compromises among competing demands. Thus interaction among parts is retained completely within the adaptationist program. Any suboptimality of a part is explained as its contribution to the best possible design for the whole. The notion that suboptimality might represent anything other than the immediate work of natural selection is usually not entertained. As Dr. Pangloss said in explaining to Candide why he suffered from venereal disease: "It is indispensable in this best of worlds. For if Columbus, when visiting the West Indies, had not caught this disease, which poisons the source of generation, which frequently even hinders generation, and is clearly opposed to the great end of Nature, we should have neither chocolate nor cochineal." The adaptationist program is truly Panglossian. Our world may not be good in an abstract sense, but it is the very best we could have. Each trait plays its part and must be as it is.

At this point, some evolutionists will protest that we are caricaturing their view of adaptation. After all, do they not admit genetic drift, allometry, and a variety of reasons for nonadaptive evolution? They do, to be sure, but we make a different point. In natural history, all possible things happen sometimes; you generally do not support your favored phenomenon by declaring rivals impossible in theory. Rather, you acknowledge the rival but circumscribe its domain of action so narrowly that it cannot have any importance in the affairs of nature. Then, you often congratulate yourself for being such an undogmatic and ecumenical chap. We maintain that alternatives to selection for best overall design have generally been relegated to unimportance by this mode of argument. Have we not all heard the catechism about genetic drift: it can only be important in populations so small that they are likely to become extinct before playing any sustained evolutionary role (but see Lande 1976).

The admission of alternatives in principle does not imply their serious consideration in daily practice. We all say that not everything is adaptive; yet, faced with an organism, we tend to break it into parts and tell adaptive stories as if trade-offs among competing, well-designed parts were the only constraint upon perfection for each trait. It is an old habit. As Romanes complained about A. R. Wallace in 1900: "Mr. Wallace does not expressly maintain the abstract impossibility of laws and causes other than those of utility and natural selection.... Nevertheless, as he nowhere recognizes any other law or cause ... he practically concludes that, on inductive or empirical grounds, there is *no* such other law or cause to be entertained."

The adaptationist program can be traced through common styles of argument. We illustrate just a few; we trust they will be recognized by all:

1. If one adaptive argument fails, try another. Zig-zag commissures of clams and brachiopods, once widely regarded as devices for strengthening the shell, become sieves for restricting particles above a given size (Rudwick 1964). A suite of external structures (horns, antlers, tusks), once viewed as weapons against predators, become symbols of intraspecific competition among males (Davitashvili 1961). The Eskimo face, once depicted as "cold engineered" (Coon et al. 1950), becomes an adaptation to generate and withstand large masticatory forces (Shea 1977). We do not attack these newer interpretations; they may all be right. We do wonder, though, whether the failure of one adaptive explanation should always simply inspire a search for another of the same general form, rather than a consideration of alternatives to the proposition that each part is "for" some specific purpose.

2. If one adaptive argument fails, assume that another must exist; a weaker version of the first argument. Costa and Bisol (1978), for example, hoped to find a correlation between genetic polymorphism and stability of environment in the deep sea, but they failed. They conclude (1978, pp. 132, 133): "The degree of genetic polymorphism found would seem to indicate absence of correlation with the particular environmental

factors which characterize the sampled area. The results suggest that the adaptive strategies of organisms belonging to different phyla are different."

3. In the absence of a good adaptive argument in the first place, attribute failure to imperfect understanding of where an organism lives and what it does. This is again an old argument. Consider Wallace on why all details of color and form in land snails must be adaptive, even if different animals seem to inhabit the same environment (1899, p. 148): "The exact proportions of the various species of plants, the numbers of each kind of insect or of bird, the peculiarities of more or less exposure to sunshine or to wind at certain critical epochs, and other slight differences which to us are absolutely immaterial and unrecognizable, may be of the highest significance to these humble creatures, and be quite sufficient to require some slight adjustments of size, form, or color, which natural selection will bring about."

4. Emphasize immediate utility and exclude other attributes of form. Fully half the explanatory information accompanying the full-scale Fiberglass *Tyrannosaurus* at Boston's Museum of Science reads: "Front legs a puzzle: how *Tyrannosaurus* used its tiny front legs is a scientific puzzle; they were too short even to reach the mount. They may have been used to help the animal rise from a lying position." (We purposely choose an example based on public impact of science to show how widely habits of the adaptationist program extend. We are not using glass beasts as straw men; similar arguments and relative emphases, framed in different words, appear regularly in the professional literature.) We don't doubt that *Tyrannosaurus* used its diminutive front legs for something. If they had arisen *de novo*, we would encourage the search for some immediate adaptive reason. But they are, after all, the reduced product of conventionally functional homologues in ancestors (longer limbs of allosaurs, for example). As such, we do not need an explicitly adaptive explanation for the reduction itself. It is likely to be a developmental correlate of allometric fields for relative increase in head and hindlimb size. This nonadaptive hypothesis can be tested by conventional allometric methods (Gould 1974, in general; Lande 1978, on limb reduction) and seems to us both more interesting and fruitful than untestable speculations based on secondary utility in the best of possible worlds. One must not confuse the fact that a structure is used in some way (consider again the spandrels, ceiling spaces, and Aztec bodies) with the primary evolutionary reason for its existence and conformation.

Telling Stories

All this is a manifestation of the rightness of things, since if there is a volcano at Lisbon it could not be anywhere else. For it is impossible for things not to be where they are, because everything is for the best.

—Dr. Pangloss on the great Lisbon earthquake of 1755, in which up to 50,000 people lost their lives

We would not object so strenuously to the adaptationist program if its invocation, in any particular case, could lead in principle to its rejection for want of evidence. We might still view it as restrictive and object to its status as an argument of first choice. But if it could be dismissed after failing some explicit test, then alternatives would get their chance. Unfortunately, a common procedure among evolutionists does not allow such definable rejection for two reasons. First, the rejection of one adaptive story usually leads to its replacement by another, rather than to a suspicion that a different kind of explanation might be required. Since the range of adaptive stories is as wide as our minds are fertile, new stories can always be postulated. And if a story is not immediately available, one can always plead temporary ignorance and trust that it will be forthcoming, as did Costa and Bisol (1978), cited above. Second, the criteria for acceptance of a story are so loose that many pass without proper confirmation. Often, evolutionists use *consistency* with natural selection as the sole criterion and consider their work done when they concoct a plausible story. But plausible stories can always be told. The key to historical research lies in devising criteria to identify proper explanations among the substantial set of plausible pathways to any modern result.

We have, for example (Gould 1978) criticized Barash's (1976) work on aggression in mountain bluebirds for this reason. Barash mounted a stuffed male near the nests of two pairs of bluebirds while the male was out foraging. He did this at the same nests on three occasions at ten-day intervals: the first before eggs were laid, the last two afterward. He then counted aggressive approaches of the returning male toward both the model and the female. At time one, aggression was high toward the model and lower toward females but substantial in both nests. Aggression toward the model declined steadily for times two and three and plummeted to near zero toward females. Barash reasoned that this made evolutionary sense, since males would be more sensitive to intruders before eggs were laid than afterward (when they can have some confidence that their genes are inside). Having devised this plausible story, he considered his work as completed (1976, pp. 1099, 1100):

> The results are consistent with the expectations of evolutionary theory. Thus aggression toward an intruding male (the model) would clearly be especially advantageous early in the breeding season, when territories and nests are normally defended. . . . The initial aggressive response to the mated female is also adaptive in that, given a situation suggesting a high probability of adultery (i.e., the presence of the model near the female) and assuming that replacement females are available, obtaining a new mate would enhance the fitness of males. . . . The decline in male-female aggressiveness during incubation and fledgling stages could be attributed to the impossibility of being cuckolded after the eggs have been laid. . . . The results are consistent with an evolutionary interpretation.

They are indeed consistent, but what about an obvious alternative, dismissed without test by Barash? Male returns at times two and three, approaches the model, tests it a bit, recognizes it as the same phoney he saw before, and doesn't bother his female.

Why not at least perform the obvious test for this alternative to a conventional adaptive story: expose a male to the model for the *first* time after the eggs are laid?

After we criticized Barash's work, Morton et al. (1978) repeated it, with some variations (including the introduction of a female model), in the closely related eastern bluebird *Sialia sialis*. "We hoped to confirm," they wrote, that Barash's conclusions represent "a widespread evolutionary reality, at least within the genus *Sialia*. Unfortunately, we were unable to do so." They found no "anticuckoldry" behavior at all: males never approached their females aggressively after testing the model at any nesting stage. Instead, females often approached the male model and, in any case, attacked female models more than males attacked male models. "This violent response resulted in the near destruction of the female model after presentations and its complete demise on the third, as a female flew off with the model's head early in the experiment to lose it for us in the brush" (1978, p. 969). Yet, instead of calling Barash's selected story into question, they merely devise one of their own to render both results in the adaptationist mode. Perhaps, they conjecture, replacement females are scarce in their species and abundant in Barash's. Since Barash's males can replace a potentially "unfaithful" female, they can afford to be choosy and possessive. Eastern bluebird males are stuck with uncommon mates and had best be respectful. They conclude: "If we did not support Barash's suggestion that male bluebirds show anticuckoldry adaptations, we suggest that both studies still had 'results that are consistent with the expectations of evolutionary theory' (Barash 1976, p. 1099), as we presume any careful study would." But what good is a theory that cannot fail in careful study (since by "evolutionary theory," they clearly mean the action of natural selection applied to particular cases, rather than the fact of transmutation itself)?

The Master's Voice Reexamined

Since Darwin has attained sainthood (if not divinity) among evolutionary biologists, and since all sides invoke God's allegiance, Darwin has often been depicted as a radical selectionist at heart who invoked other mechanisms only in retreat, and only as a result of his age's own lamented ignorance about the mechanisms of heredity. This view is false. Although Darwin regarded selection as the most important of evolutionary mechanisms (as do we), no argument from opponents angered him more than the common attempt to caricature and trivialize his theory by stating that it relied exclusively upon natural selection. In the last edition of the *Origin*, he wrote (1872, p. 395):

> As my conclusions have lately been much misrepresented, and it has been stated that I attribute the modification of species exclusively to natural selection, I may be permitted to remark that in the first edition of this work, and subsequently, I placed in a most conspicuous position—namely at the close of the Introduction—the following words: "I am convinced that natural selection has been the main, but not the exclusive means of modification." This has been of no avail. Great is the power of steady misinterpretation.

Romanes, whose once famous essay (1900) on Darwin's pluralism versus the panselectionism of Wallace and Weismann deserves a resurrection, noted of this passage (1900, p. 5): "In the whole range of Darwin's writings there cannot be found a passage so strongly worded as this: it presents the only note of bitterness in all the thousands of pages which he has published." Apparently, Romanes did not know the letter Darwin wrote to *Nature* in 1880, in which he castigated Sir Wyville Thomson for caricaturing his theory as panselectionist (1880, p. 32):

> I am sorry to find that Sir Wyville Thomson does not understand the principle of natural selection. . . . If he had done so, he could not have written the following sentence in the Introduction to the Voyage of the Challenger: "The character of the abyssal fauna refuses to give the least support to the theory which refers the evolution of species to extreme variation guided only by natural selection." This is a standard of criticism not uncommonly reached by theologians and metaphysicians when they write on scientific subjects, but is something new as coming from a naturalist. . . . Can Sir Wyville Thomson name any one who has said that the evolution of species depends only on natural selection? As far as concerns myself, I believe that no one has brought forward so many observations on the effects of the use and disuse of parts, as I have done in my "Variation of Animals and Plants under Domestication"; and these observations were made for this special object. I have likewise there adduced a considerable body of facts, showing the direct action of external conditions on organisms.

We do not now regard all of Darwin's subsidiary mechanisms as significant or even valid, though many, including direct modification and correlation of growth, are very important. But we should cherish his consistent attitude of pluralism in attempting to explain Nature's complexity.

A Partial Typology of Alternatives to the Adaptationist Program

In Darwin's pluralistic spirit, we present an incomplete hierarchy of alternatives to immediate adaptation for the explanation of form, function, and behavior.

1. No adaptation and no selection at all. At present, population geneticists are sharply divided on the question of how much genetic polymorphism within populations and how much of the genetic differences between species is, in fact, the result of natural selection as opposed to purely random factors. Populations are finite in size, and the isolated populations that form the first step in the speciation process are often founded by a very small number of individuals. As a result of this restriction in population size, frequencies of alleles change by *genetic drift*, a kind of random genetic sampling error. The stochastic process of change in gene frequency by random genetic drift, including the very strong sampling process that goes on when a new isolated population is formed from a few immigrants, has several important consequences. First, populations and species will become genetically differentiated, and even fixed for different alleles at a locus in the complete absence of any selective force at all.

Second, alleles can become fixed in a population *in spite of natural selection*. Even if an allele is favored by natural selection, some proportion of populations, depending upon the product of population size N and selection intensity s, will become homozygous for the less fit allele because of genetic drift. If Ns is large, this random fixation for unfavorable alleles is a rare phenomenon, but if selection coefficients are on the order of the reciprocal of population size ($Ns = 1$) or smaller, fixation for deleterious alleles is common. If many genes are involved in influencing a metric character like shape, metabolism, or behavior, then the intensity of selection on each locus will be small and Ns per locus may be small. As a result, many of the loci may be fixed for non-optimal alleles.

Third, new mutations have a small chance of being incorporated into a population, even when selectively favored. Genetic drift causes the immediate loss of most new mutations after their introduction. With a selection intensity s, a new favorable mutation has a probability of only $2s$ of ever being incorporated. Thus one cannot claim that, eventually, a new mutation of just the right sort for some adaptive argument will occur and spread. "Eventually" becomes a very long time if only one in 1,000 or one in 10,000 of the "right" mutations that do occur ever get incorporated in a population.

2. No adaptation and no selection on the part at issue; form of the part is a correlated consequence of selection directed elsewhere. Under this important category, Darwin ranked his "mysterious" laws of the "correlaton of growth." Today, we speak of pleiotropy, allometry, "material compensation" (Rensch 1959, pp. 179–187) and mechanically forced correlations in D'Arcy Thompson's sense (1942; Gould 1971). Here we come face to face with organisms as integrated wholes, fundamentally not decomposable into independent and separately optimized parts.

Although allometric patterns are as subject to selection as static morphology itself (Gould 1966), some regularities in relative growth are probably not under immediate adaptive control. For example, we do not doubt that the famous 0.66 interspecific allometry of brain size in all major vertebrate groups represents a selected "design criterion," though its significance remains elusive (Jerison 1973). It is too repeatable across too wide a taxonomic range to represent much else than a series of creatures similarly well designed for their different sizes. But another common allometry, the 0.2 to 0.4 intraspecific scaling among homeothermic adults differing in body size, or among races within a species, probably does not require a selectionist story, though many, including one of us, have tried to provide one (Gould 1974). R. Lande (personal communication) has used the experiments of Falconer (1973) to show that selection upon *body size alone* yields a brain-body slope across generations of 0.35 in mice.

More compelling examples abound in the literature on selection for altering the timing of maturation (Gould 1977). At least three times in the evolution of arthropods (mites, flies, and beetles), the same complex adaptation has evolved, apparently for rapid turnover of generations in strongly *r*-selected feeders on superabundant but ephemeral fungal resources: females reproduce as larvae and grow the next generation within their bodies. Offspring eat their mother from inside and emerge from her hollow shell, only to be devoured a few days later by their own progeny. It would be foolish to seek adaptive significance in paedomorphic morphology per se; it is primarily a by-product of selection for rapid cycling of generations. In more interesting cases, selection for small size (as in animals of the interstitial fauna) or rapid maturation (dwarf males of many crustaceans) has occurred by progenesis (Gould 1977, pp. 324–336), and descendant adults contain a mixture of ancestral juvenile and adult features. Many biologists have been tempted to find primary adaptive meaning for the mixture, but it probably arises as a byproduct of truncated maturation, leaving some features "behind" in the larval state, while allowing others, more strongly correlated with sexual maturation, to retain the adult configuration of ancestors.

3. The decoupling of selection and adaptation.

(i) Selection without adaptation. Lewontin (1979) has presented the following hypothetical example: "A mutation which doubles the fecundity of individuals will sweep through a population rapidly. If there has been no change in efficiency of resource utilization, the individuals will leave no more offspring than before, but simply lay twice as many eggs, the excess dying because of resource limitation. In what sense are the individuals or the population as a whole better adapted than before? Indeed, if a predator on immature stages is led to switch to the species now that immatures are more plentiful, the population size may actually decrease as a consequence, yet natural selection at all times will favour individuals with higher fecundity."

(ii) Adaptation without selection. Many sedentary marine organisms, sponges and corals in particular, are well adapted to the flow régimes in which they live. A wide spectrum of "good design" may be purely phenotypic in origin, largely induced by the current itself. (We may be sure of this in numerous cases, when genetically identical individuals of a colony assume different shapes in different microhabitats.) Larger patterns of geographic variation are often adaptive and purely phenotypic as well. Sweeney and Vannote (1978), for example, showed that many hemimetabolous aquatic insects reach smaller adult size with reduced fecundity when they grow at temperatures above and below their optima. Coherent, climatically correlated patterns in geographic distribution for these insects—so often taken as a priori signs of genetic adaptation—may simply reflect this phenotypic plasticity.

"Adaptation"—the good fit of organisms to their environment—can occur at three hierarchical levels with different causes. It is unfortunate that our language has focused on the common result and called all three phenomena "adaptation": the differences in process have been obscured, and evolutionists have often been misled to extend the Darwinian mode to the other two levels as well. First, we have what physiologists call "adaptation": the phenotypic plasticity that permits organisms to mold their form to prevailing circumstances during ontogeny. Human ''adaptations'' to high altitude fall into this category (while others, like resistance of sickling heterozygotes to malaria, are genetic, and Darwinian). Physiological adaptations are not heritable, though the capacity to develop them presumably is. Second, we have a ''heritable'' form of non-Darwinian adaptation in humans (and, in rudimentary ways, in a few other advanced social species): cultural adaptation (with heritability imposed by learning). Much confused thinking in human sociobiology arises from a failure to distinguish this mode from Darwinian adaptation based on genetic variation. Finally, we have adaptation arising from the conventional Darwinian mechanism of selection upon genetic variation. The mere existence of a good fit between organism and environment is insufficient for inferring the action of natural selection.

4. Adaptation and selection but no selective basis for differences among adaptations. Species of related organisms, or subpopulations within a species, often develop different adaptations as solutions to the same problem. When ''multiple adaptive peaks'' are occupied, we usually have no basis for asserting that one solution is better than another. The solution followed in any spot is a result of history; the first steps went in one direction, though others would have led to adequate prosperity as well. Every naturalist has his favorite illustration. In the West Indian land snail *Cerion*, for example, populations living on rocky and windy coasts almost always develop white, thick, and relatively squat shells for conventional adaptive reasons. We can identify at least two different developmental pathways to whiteness from the mottling of early whorls in all *Cerion*, two paths of thickened shells and three styles of allometry leading to squat shells. All twelve combinations can be identified in Bahamian populations, but would it be fruitful to ask why—in the sense of optimal design rather than historical contingency—*Cerion* from eastern Long Island evolved one solution, and *Cerion* from Acklins Island another?

5. Adaptation and selection, but the adaptation is a secondary utilization of parts present for reasons of architecture, development, or history. We have already discussed this neglected subject in the first section on spandrels, spaces, and cannibalism. If blushing turns out to be an adaptation affected by sexual selection in humans, it will not help us to understand why blood is red. The immediate utility of an organic structure often says nothing at all about the reason for its being.

Another, and Unfairly Maligned, Approach to Evolution

In continental Europe, evolutionists have never been much attracted to the Anglo-American penchant for atomizing organisms into parts and trying to explain each as a direct adaptation. Their general alternative exists in both a strong and a weak form. In the strong form, as advocated by such major theorists as Schindewolf (1950), Remane (1971), and Grassé (1977), natural selection under the adaptationist program can explain superficial modifications of the *Bauplan* that fit structure to environment: why moles are blind, giraffes have long necks, and ducks webbed feet, for example. But the important steps of evolution, the construction of the *Bauplan* itself and the transition between *Baupläne*, must involve some other unknown, and perhaps ''internal,'' mechanism. We believe that English biologists have been right in rejecting this strong form as close to an appeal to mysticism.

But the argument has a weaker—and paradoxically powerful—form that has not been appreciated, but deserves to be. It also acknowledges conventional selection for superficial modifications of the *Bauplan*. It also denies that the adaptationist program (atomization plus optimizing selection on parts) can do much to explain *Baupläne* and the transitions between them. But it does not therefore resort to a fundamentally unknown process. It holds instead that the basic body plans of organisms are so integrated and so replete with constraints upon adaptation (categories 2 and 5 of our typology) that conventional styles of selective arguments can explain little of interest about them. It does not deny that change, when it occurs, may be mediated by natural selection, but it holds that constraints restrict possible paths and modes of change so strongly that the constraints themselves become much the most interesting aspect of evolution.

Rupert Riedl, the Austrian zoologist who has tried to develop this thesis for English audiences (1977 and 1975, translated into English by R. Jeffries in 1978) writes:

> The living world happens to be crowded by universal patterns of organization which, most obviously, find no direct explanation through environmental conditions or adaptive radiation, but exist primarily through universal requirements which can only be expected under the systems conditions of complex organization itself. . . . This is not self-evident, for the whole of the huge and profound thought collected in the field of morphology, from Goethe to Remane, has virtually been cut off from modern biology. It is not taught in most American universities. Even the teachers who could teach it have disappeared.

Constraints upon evolutionary change may be ordered into at least two categories. All evolutionists are familiar with *phyletic* constraints, as embodied in Gregory's classic distinction (1936) between habitus and heritage. We acknowledge a kind of phyletic inertia in recognizing, for example, that humans are not optimally designed for upright posture because so much of our *Bauplan* evolved for quadrupedal life. We also invoke phyletic constraint in explaining why no molluscs fly in air and no insects are as large as elephants.

Developmental constraints, a subcategory of phyletic restrictions, may hold the most powerful rein of all over possible evolutionary pathways. In complex organisms, early stages of ontogeny are remarkably refractory to evolutionary change, presumably because the differentiation of organ systems and their integration into a functioning body is such a delicate process so easily derailed by early errors with accumulating effects. Von Baer's fundamental embryological laws (1828) represent little more than a recognition that early stages are both highly conservative and strongly restrictive of later development. Haeckel's biology law, the primary subject of late nineteenth-century evolutionary biology, rested upon a misreading of the same data (Gould 1977). If development occurs in integrated packages and cannot be pulled apart piece by piece in evolution, then the adaptationist program cannot explain the alteration of developmental programs underlying nearly all changes of *Bauplan*.

The German palaeontologist A. Seilacher, whose work deserves far more attention than it has received, has emphasized what he calls "*bautechnischer*, or *architectural*, constraints" (Seilacher 1970). These arise not from former adaptations retained in a new ecological setting (phyletic constraints as usually understood), but as architectural restrictions that never were adaptations but rather were the necessary consequences of materials and designs selected to build basic *Baupläne*. We devoted the first section of this chapter to nonbiological examples in this category. Spandrels must exist once a blueprint specifies that a dome shall rest on rounded arches. Architectural constraints can exert a far-ranging influence upon organisms as well. The subject is full of potential insight because it has rarely been acknowledged at all.

In a fascinating example, Seilacher (1972) has shown that the divaricate form of architecture (figure 5.3) occurs again and again in all groups of molluscs, and in brachiopods as well. This basic form expresses itself in a wide variety of structures: raised ornamental lines (not growth lines because they do not conform to the mantle margin at any time), patterns of coloration, internal structures in the mineralization of calcite and incised grooves. He does not know what generates this pattern and feels that traditional and nearly exclusive focus on the adaptive value of each manifestation has diverted attention from questions of its genesis in growth and also prevented its recognition as a general phenomenon. It must arise from some characteristic pattern of inhomogeneity in the growing mantle, probably from the generation of interference patterns around regularly spaced centers; simple computer simulations can generate the form in this manner (Waddington and Cowe 1969). The general pattern may not be a direct adaptation at all.

Seilacher then argues that most manifestations of the pattern are probably nonadaptive. His reasons vary but seem generally sound to us. Some are based on field observations: color patterns that remain invisible because clams possessing them either live buried in sediments or remain covered with a periostracum so thick that the colors cannot be seen. Others rely on more general principles: presence only in odd and pathological individuals, rarity as a developmental anomaly, excessive variability compared with much reduced variability when the same general structure assumes a form judged functional on engineering grounds.

In a distinct minority of cases, the divaricate pattern becomes functional in each of the four categories (figure 5.3). Divaricate ribs may act as scoops and anchors in burrowing (Stanley 1970), but they are not properly arranged for such function in

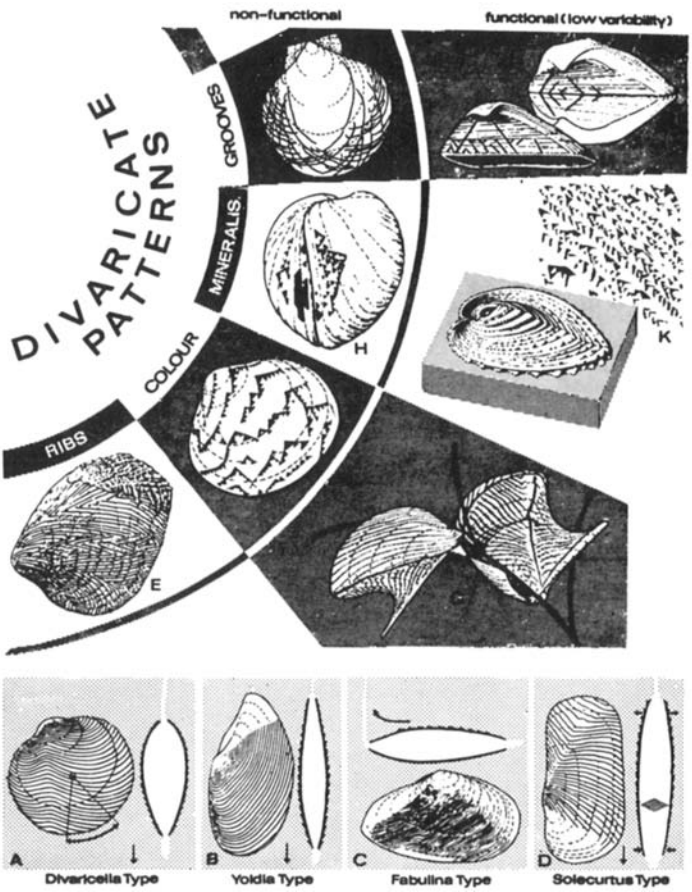


Figure 5.3
The range of divaricate patterns in molluscs. E, F, H, and L are non-functional in Seilacher's judgement. A–D are functional ribs (but these are far less common than non-functional ribs of the form E). G is the mimetic *Arca zebra*. K is *Corculum*. See text for details.

most clams. The color chevrons are mimetic in one species (*Pteria zebra*) that lives on hydrozoan branches; here the variability is strongly reduced. The mineralization chevrons are probably adaptive in only one remarkable creature, the peculiar bivalve *Corculum cardissa* (in other species they either appear in odd specimens or only as postmortem products of shell erosion). This clam is uniquely flattened in an anterio-posterior direction. It lies on the substrate, posterior up. Distributed over its rear end are divaricate triangles of mineralization. They are translucent, while the rest of the shell is opaque. Under these windows dwell endosymbiotic algae!

All previous literature on divaricate structure has focused on its adaptive significance (and failed to find any in most cases). But Seilacher is probably right in representing this case as the spandrels, ceiling holes, and sacrificed bodies of our first section. The divaricate pattern is a fundamental architectural constraint. Occasionally, since it is there, it is used to beneficial effect. But we cannot understand the pattern or its evolutionary meaning by viewing these infrequent and secondary adaptations as a reason for the pattern itself.

Galton (1909, p. 257) contrasted the adaptationist program with a focus on constraints and modes of development by citing a telling anecdote about Herbert Spencer's fingerprints:

> Much has been written, but the last word has not been said, on the rationale of these curious papillary ridges; why in one man and in one finger they form whorls and in another loops. I may mention a characteristic anecdote of Herbert Spencer in connection with this. He asked me to show him my Laboratory and to take his prints, which I did. Then I spoke of the failure to discover the origin of these patterns, and how the fingers of unborn children had been dissected to ascertain their earliest stages, and so forth. Spencer remarked that this was beginning in the wrong way; that I ought to consider the purpose the ridges had to fulfil, and to work backwards. Here, he said, it was obvious that the delicate mouths of the sudorific glands required the protection given to them by the ridges on either side of them, and therefrom he elaborated a consistent and ingenious hypothesis at great length. I replied that his arguments were beautiful and deserved to be true, but it happened that the mouths of the ducts did not run in the valleys between the crests, but along the crests of the ridges themselves.

We feel that the potential rewards of abandoning exclusive focus on the adaptationist program are very great indeed. We do not offer a counsel of despair, as adaptationists have charged; for nonadaptive does not mean nonintelligible. We welcome the richness that a pluralistic approach, so akin to Darwin's spirit, can provide. Under the adaptationist program, the great historic themes of developmental morphology and *Bauplan* were largely abandoned; for if selection can break any correlation and optimize parts separately, then an organism's integration counts for little. Too often, the adaptationist program gave us an evolutionary biology of parts and genes, but not of organisms. It assumed that all transitions could occur step by step and underrated the importance of integrated developmental blocks and pervasive constraints of history and architecture. A pluralistic view could put organisms, with all their recalcitrant yet intelligible complexity, back into evolutionary theory.

References

Baer, K. E. von. 1828. *Entwicklungsgeschichte der Tiere*, Königsberg: Bornträger.

Barash, D. P. 1976. Male response to apparent female adultery in the mountain-bluebird: an evolutionary interpretation, *Am. Nat.*, 110: 1097–1101.

Coon, C. S., Garn, S. M., and Birdsell, J. B. 1950. *Races*, Springfield, Oh., C. Thomas.

Costa, R., and Bisol, P. M. 1978. Genetic variability in deep-sea organisms, *Biol. Bull.*, 155: 125–133.

Darwin, C. 1872. *The origin of species*, London, John Murray.

———. 1880. Sir Wyville Thomson and natural selection, *Nature*, London, 23: 32.

Davitashvili, L. S. 1961. *Teoriya polovogo otbora* [Theory of sexual selection], Moscow, Akademii Nauk.

Falconer, D. S. 1973. Replicated selection for body weight in mice, *Genet. Res.*, 22: 291–321.

Galton, F. 1909. *Memories of my life*, London, Methuen.

Gould, S. J. 1966. Allometry and size in ontogeny and phylogeny, *Biol. Rev.*, 41: 587–640.

———. 1971. D'Arcy Thompson and the science of form, *New Literary Hist.*, 2, no. 2, 229–258.

———. 1974. Allometry in primates, with emphasis on scaling and the evolution of the brain. In *Approaches to primate paleobiology, Contrib. Primatol.*, 5: 244–292.

———. 1977. *Ontogeny and phylogeny*, Cambridge, Ma., Belknap Press.

———. 1978. Sociobiology: the art of storytelling, *New Scient.*, 80: 530–533.

Grassé, P. P. 1977. *Evolution of living organisms*, New York, Academic Press.

Gregory, W. K. 1936. Habitus factors in the skeleton fossil and recent mammals, *Proc. Am. phil. Soc.*, 76: 429–444.

Harner, M. 1977. The ecological basis for Aztec sacrifice. *Am. Ethnologist*, 4: 117–135.

Jerison, H. J. 1973. *Evolution of the brain and intelligence*, New York, Academic Press.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution, *Evolution*, 30: 314–334.

———. 1978. Evolutionary mechanisms of limb loss in tetrapods, *Evolution*, 32: 73–92.

Lewontin, R. C. 1978. Adaptation, *Scient. Am.*, 239 (3): 156–169.

———. 1979. Sociobiology as an adaptationist program, *Behav. Sci.*, 24: 5–14.

Morton, E. S., Geitgey, M. S., and McGrath, S. 1978. On bluebird "responses to apparent female adultery." *Am. Nat.*, 112: 968–971.

Ortiz de Montellano, B. R. 1978. Aztec cannibalism: an ecological necessity? *Science*, 200: 611–617.

Remane, A. 1971. *Die Grundlagen des natürlichen Systems der vergleichenden Anatomie und der Phylogenetik*. Königstein-Taunus: Koeltz.

Rensch, B. 1959. *Evolution above the species level*, New York, Columbia University Press.

Riedl, R. 1975. *Die Ordnung des Lebendigen*, Hamburg, Paul Parey, tr. R. P. S. Jefferies, *Order in living systems: A systems analysis of evolution*, New York, Wiley, 1978.

———. 1977. A systems-analytical approach to macro-evolutionary phenomena, *Q. Rev. Biol.*, 52: 351–370.

Romanes, G. J. 1900. The Darwinism of Darwin and of the post-Darwinian schools. In *Darwin, and after Darwin*, vol. 2, new ed., London, Longmans, Green and Co.

Rudwick, M. J. S. 1964. The function of zig-zag deflections in the commissures of fossil brachiopods, *Palaeontology*, 7: 135–171.

Sahlins, M. 1978. Culture as protein and profit, *New York Review of Books*, 23: Nov., pp. 45–53.

Schindewolf, O. H. 1950. *Grundfragen der Paläontologie*, Stuttgart, Schweizerbart.

Seilacher, A. 1970. Arbeitskonzept zur Konstruktionsmorphologie, *Lethaia*, 3: 393–396.

———. 1972. Divaricate patterns in pelecypod shells, *Lethaia*, 5: 325–343.

Shea, B. T. 1977. Eskimo craniofacial morphology, cold stress and the maxillary sinus, *Am. J. Phys. Anthrop.*, 47: 289–300.

Stanley, S. M. 1970. Relation of shell form to life habits in the Bivalvia (Mollusca). *Mem. Geol. Soc. Am.*, no. 125, 296 pp.

Sweeney, B. W., and Vannote, R. L. 1978. Size variation and the distribution of hemimetabolous aquatic insects: two thermal equilibrium hypotheses. *Science*, 200: 444–446.

Thompson, D. W. 1942. *Growth and form*, New York, Macmillan.

Waddington, C. H., and Cowe, J. R. 1969. Computer simulation of a molluscan pigmentation pattern, *J. Theor. Biol.*, 25: 219–225.

Wallace, A. R. 1899. *Darwinism*, London, Macmillan.

Wilson, E. O. 1978. *On human nature*, Cambridge, Ma., Harvard University Press.

6 Optimization Theory in Evolution

John Maynard Smith

Introduction

In recent years there has been a growing attempt to use mathematical methods borrowed from engineering and economics in interpreting the diversity of life. It is assumed that evolution has occurred by natural selection, hence that complex structures and behaviors are to be interpreted in terms of the contribution they make to the survival and reproduction of their possessors—that is, to Darwinian fitness. There is nothing particularly new in this logic, which is also the basis of functional anatomy, and indeed of much physiology and molecular biology. It was followed by Darwin himself in his studies of climbing and insectivorous plants, of fertilization mechanisms and devices to ensure cross-pollination.

What is new is the use of such mathematical techniques as control theory, dynamic programming, and the theory of games to generate a priori hypotheses, and the application of the method to behaviors and life history strategies. This change in method has led to the criticism (e.g., Lewontin, 54, 55) that the basic hypothesis of adaptation is untestable and therefore unscientific, and that the whole program of functional explanation through optimization has become a test of ingenuity rather than an inquiry into truth. Related to this is the criticism that there is no theoretical justification for any maximization principles in biology, and therefore that optimization is no substitute for an adequate genetic model.

My aim in this review is not to summarize the most important conclusions reached by optimization methods, but to discuss the methodology of the program and the criticisms that have been made of it. In doing so, I have taken as my starting point two articles by Lewontin (54, 55). I disagree with some of the views he expresses, but I believe that the development of evolution theory could benefit if workers in optimization paid serious attention to his criticisms.

From *Annual Review of Ecology and Systematics* 9 (1978): 31–56. Reprinted by permission of *Annual Review*.

I first outline the basic structure of optimization arguments, illustrating this with three examples, namely the sex ratio, the locomotion of mammals, and foraging behavior. I then discuss the possibility that some variation may be selectively neutral, and some structures maladaptive. I summarize and comment on criticisms made by Lewontin. The most damaging, undoubtedly, is the difficulty of testing the hypotheses that are generated. The next section therefore discusses the methodology of testing; in this section I have relied heavily on the arguments of Curio (23). Finally I discuss mathematical methods. The intention here is not to give the details of the mathematics, but to identify the kinds of problems that have been attacked and the assumptions that have been made in doing so.

The Structure of Optimization Models

In this section I illustrate the argument with three example: (*a*) the sex ratio, based on Fisher's (28) treatment and later developments by Hamilton (34), Rosado and Robertson (85), Trivers and Willard (96), and Trivers and Hare (95); (*b*) the gaits of mammals—given a preliminary treatment by Maynard Smith and Savage (66), and further analyzed in several papers in Pedley (78); (*c*) foraging strategies. Theoretical work on them originated with the papers of Emlen (27) and MacArthur and Pianka (57). I have relied heavily on a recent review by Pyke et al. (81). These authors suggest that models have in the main been concerned with four problems: choice by the animal of which types of food to eat (optimal diet); choice of which patch type to feed in; allocation of time to different patches; pattern and speed of movement. In what follows I shall refer only to two of those—optimal diet and allocation of time to different patches.

All optimization models contain, implicitly or explicitly, an assumption about the "constraints" that are operating, an optimization criterion, and an assumption about heredity. I consider these in turn.

The Constraints: Phenotype Set and State Equations

The constraints are essentially of two kinds. In engineering applications, they concern the "strategy set," which specifies the range of control actions available, and the "state equations," which specify how the state of the system being controlled changes in time. In biological applications, the strategy set is replaced by an assumption about the set of possible phenotypes on which selection can operate.

It is clearly impossible to say what is the "best" phenotype unless one knows the range of possibilities. If there were no constraints on what is possible, the best phenotypes would live forever, would be impregnable to predators, would lay eggs at an infinite rate, and so on. It is therefore necessary to specify the set of possible phenotypes, or in some other way describe the limits on what can evolve. The "phenotype set" is an assumption about what can evolve and to what extent; the "state equations" describe features of the situation that are assumed not to change. This distinction will become clearer when particular examples are discussed. Let us consider the three problems in turn.

Sex Ratio For the sex ratio, the simplest assumption is that a parent can produce a fixed number N of offspring, and that the probability S that each birth will be a male can vary from parent to parent over the complete range from 0 to 1; the phenotype set is then the set of values of S over this range. Fisher (28) extended this by supposing that males and females "cost" different amounts; i.e. he supposed that a parent could produce α males and β females, where α and β are constrained to lie on or below the line $\alpha + \beta k = N$, and k is the cost of a female relative to that of a male. He then concluded that the parent should equalize expenditure on males and females. MacArthur (56) further broadened the phenotype set by insisting only that α and β lie on or below a line of arbitrary shape, and concluded that a parent should maximize $\alpha\beta$. A similar assumption was used by Charnov et al. (11) to analyze the evolution of hermaphroditism as opposed to dioecy. Finally, it is possible to ask (97) what is the optimal strategy if a parent can choose not merely a value of S, hence of the expected sex ratio, but also the variance of the sex ratio.

The important point in the present context is that the optimal solution depends on the assumption made. For example, Crow and Kimura (21) conclude that the sex ratio should be unity, but they do so for a model that assumes that $N = \alpha + \beta$ is a constant.

Gaits In the analysis of gaits, it is assumed that the shapes of bones can vary but the mechanical properties of bone, muscle, and tendon cannot. It is also assumed that changes must be gradual; thus the gaits of ostrich, antelope, and kangaroo are seen as different solutions to the same problem, not as solutions to different problems—i.e., they are different "adaptive peaks" (101).

Foraging Strategy In models of foraging behavior, a common assumption is that the way in which an animal allocates its time among various activities (e.g., consuming one prey item rather than another, searching in one kind of patch rather than another, moving between patches rather than continuing to search in the same one) can vary, but the efficiency with which it performs each act cannot. Thus, for example, the length of time it takes to "handle" (capture and consume) a given item, the time and energy spent in moving from place to place, and the time taken to find a given prey item at a given prey density are taken as invariant. Thus the models of foraging so far developed treat the phenotype set as the set of possible behavioral strategies, and treat structure and locomotory or perceptual skills as constants contributing to the state equations (which determine how rapidly an animal adopting some strategy acquires food). In principle there is no reason why optimization models should not be applied to the evolution of structure or skill also; it is simply a question of how the phenotype set is defined.

The Optimization Criterion

Some assumption must then be made concerning what quantity is being maximized. The most satisfactory is the inclusive fitness (see the section Games between Relatives, below); in many contexts the individual fitness (expected number of offspring) is equally good. Often, as in the second and third of my examples, neither criterion is possible, and some other assumption is needed. Two points must be made. First, the assumption about what is maximized is an assumption about what selective forces have been responsible for the trait; second, this assumption is part of the hypothesis being tested.

In most theories of sex ratio the basic assumption is that the ratio is determined by a gene acting in a parent, and what is maximized is the number of copies of that gene in future generations. The maximization has therefore a sound basis. Other maximization criteria have been used. For example, Kalmus and Smith (41) propose that the sex ratio maximizes the probability that two individuals meeting will be of different sexes; it is hard to understand such an eccentric choice when the natural one is available.

An equally natural choice—the maximization of the expected number of offspring produced in a lifetime—is available in theories of the evolution of life history strategies. But often no such easy choice is available.

In the analysis of gaits, Maynard Smith and Savage (66) assumed that the energy expenditure at a given speed would be minimized (or, equivalently, that the speed for a given energy expenditure was maximized). This led to the prediction that the proportion of time spent with all four legs off the ground should increase with speed and decrease with size.

In foraging theory, the common assumption is that the animal is maximizing its energy intake per unit time spent foraging. Schoener (87) points out that this is an appropriate choice, whether the animal has a fixed energy requirement and aims to minimize the time spent feeding so as to leave more time for other activities (‘‘time minimizers’’), or has a fixed time in which to feed during which it aims to maximize its energy gain (‘‘energy maximizers’’). There will, however, be situations in which this is not an appropriate choice. For example, there may be a higher risk of predation for some types of foraging than for others. For some animals the problem may be not to maximize energy intake per unit time, but to take in a required amount of energy, protein, etc. without taking an excess of any one of a number of toxins (S. A. Altmann, personal communication).

Pyke et al. (81) point out that the optimal strategy depends on the time scale over which optimization is carried out, for two reasons. First, an animal that has sole access to some resource e.g., a territory-holder) can afford to manage that resource so as to maximize its yield over a whole season. Second, and more general, optimal behavior depends on a knowledge of the environment, which can be acquired only by experience; this means that in order to acquire information of value in the long run, an animal may have to behave in a way that is inefficient in the short run.

Having considered the phenotype set and the optimization criterion, a word must be said about their relationship to Levins's (51) concept of a fitness set. Levins was explicitly concerned with defining fitness "in such a way that interpopulation selection would be expected to change a species towards the optimum (maximum fitness) structure." This essentially group-selectionist approach led him to conclusions (e.g., for the conditions for a stable polymorphism) different from those reached from the classic analysis of gene frequencies (93). Nevertheless, Levins's attempt to unite ecological and genetic approaches did lead him to recognize the need for the concept of a fitness set—i.e., the set of all possible phenotypes, each phenotype being characterized by its (individual) fitness in each of the environments in which it might find itself.

Levins's fitness set is thus a combination of what I have called the phenotype set and of a measure of the fitness of each phenotype in every possible environment. It did not allow for the fact that fitnesses may be frequency-dependent (see the section on Games, below). The valuable insight in Levins's approach is that it is possible to discuss what course phenotypic evolution may take only if one makes explicit assumptions about the constraints on what phenotypes are possible. It may be better to use the term "phenotype set" to define these constraints, both because a description of possible phenotypes is a process prior to and separable from an estimation of their fitnesses, and because of the group-selectionist associations of the term "fitness set."

An Assumption about Heredity

Because natural selection cannot produce adaptation unless there is heredity, some assumption, explicit or otherwise, is always present. The nature of this assumption can be important. Fisher (28) assumed that the sex ratio was determined by autosomal genes expressed in the parent, and that mating was random. Hamilton (34) showed that the predicted optima are greatly changed if these assumptions are altered. In particular, he considered the effects of inbreeding, and of genes for meiotic drive. Rosado and Robertson (85), Trivers and Willard (96), and Trivers and Hare (95) have analyzed the effects of genes acting in the children and (in Hymenoptera) in the sterile castes.

It is unusual for the way in which a trait is inherited to have such a crucial effect. Thus in models of mammalian gaits no explicit assumption is made; the implicit assumption is merely that like begets like. The same is true of models of foraging, although in this case "heredity" can be cultural as well as genetic—e.g. (72), for the feeding behavior of oyster-catchers.

The question of how optimization models can be tested is the main topic of the next three sections. A few preliminary remarks are needed. Clearly, the first requirement of a model is that the conclusions should follow from the assumptions. This seems not to be the case, for example, for Zahavi's (102) theory of sexual selection (61). A more usual difficulty is that the conclusions depend on unstated assumptions. For example, Fisher does not state that his sex ratio argument assumes random mating, and this was not noticed until Hamilton's 1967 paper (34). Maynard Smith and Price (65) do not state that the idea of an ESS (evolutionary stable strategy) assumes asexual inheritance. It is probably true that no model ever states all its assumptions explicitly. One reason for writing this review is to encourage authors to become more aware of their assumptions.

A particular model can be tested either by a direct test of its assumptions or by comparing its predictions with observation. The essential point is that in testing a model we are testing *not* the general proposition that nature optimizes, but the specific hypotheses about constraints, optimization criteria, and heredity. Usually we test whether we have correctly identified the selective forces responsible for the trait in question. But we should not forget hypotheses about constraints or heredity. For example, the weakest feature of theories concerning the sex ratio is that there is little evidence for the existence of genetic variance of the kind assumed by Fisher—for references, see (63). It may be for this reason that the greatest successes of sex ratio theory (34, 95) have concerned Hymenoptera, in which it is easy to see how genes in the female parent can affect the sex of her children.

Neutrality and Maladaptation

I have said that when testing optimization models, one is not testing the hypothesis that nature optimizes. But if it is not the case that the structure and behavior of organisms are nicely adapted to ensure their survival and reproduction, optimization models cannot be useful. What justification have we for assuming this?

The idea of adaptation is older than Darwinism. In the form of the argument from design, it was a buttress of religious belief. For Darwin the problem was not to prove that organisms were adapted but to explain how adaptation could arise without a creator. He was quite willing to accept that some characteristics are "selectively neutral." For example, he says (26) of the sterile dark red flower at the center of the umbel of the wild carrot: "That the modified central flower is of no functional importance to the plant is almost certain." Indeed, Darwin has been chided by Cain (8) for too readily accepting Owen's argument that the homology between bones of limbs of different vertebrates is nonadaptive. For Darwin the argument was welcome, because the resemblance could then be taken as evidence for genetic relationship (or, presumably, for a paucity of imagination on the part of the creator). But Cain points out that the homology would not have been preserved if it were not adaptive.

Biologists differ greatly in the extent to which they expect to find a detailed fit between structure and function. It may be symptomatic of the times that when, in conversation, I raised Darwin's example of the carrot, two different functional explanations were at once suggested. I suspect that these explanations were fanciful. But however much one may be in doubt about the function of the antlers of the irish elk or the tail of the peacock, one can hardly suppose them to be selectively neutral. In general, the structural and behavioral traits chosen for functional analysis are of a kind that rules out neutrality as a plausible explanation. Curio (23) makes the valid point that the ampullae of Lorenzini in elasmobranchs were studied for many years before their role in enabling a fish to locate prey buried in the mud was demonstrated (40), yet the one hypothesis that was never entertained was that the organ was functionless. The same could be said of Curio's own work (24) on the function of mobbing in birds; behavior so widespread, so constant, and so apparently dangerous calls for a functional explanation.

There are, however, exceptions to the rule that functional investigations are carried out with the aim of identifying particular selective forces, and not of demonstrating that traits are adaptive. The work initiated by Cain and Sheppard (9) on shell color and banding in *Cepaea* was in part aimed at refuting the claim that the variation was selectively neutral and explicable by genetic drift. To that extend the work was aimed at demonstrating adaptation as such; it is significant, however, that the work has been most successful when it has been possible to identify a particular selection pressure (e.g., predation by thrushes).

At present, of course, the major argument between neutral and selective theories concerns enzyme polymorphism. I cannot summarize the argument here, but a few points on methodology are relevant. The argument arose because of the formulation by Kimura (43) and King and Jukes (44) of the "neutral" hypothesis; one reason for proposing it was the difficulty of accounting for the extensive variation by selection. Hence the stimulus was quite different from that prompting most functional investigations; it was the existence of widespread variation in a trait of no obvious selective significance.

The neutral hypothesis is a good "Popperian" one; if it is false, it should be possible to show it. In contrast, the hypothesis of adaptation is virtually irrefutable. In practice, however, the statistical predictions of the neutral theory depend on so many unknowns (mutation rates, the past history of population number and structure, hitch-hiking from other loci) that it has proved hard to test (53). The difficulties have led some geneticists (e.g., 14) to propose that the only way in which the matter can be settled is by the classical methods of ecological genetics—i.e., by identifying the specific selection pressures associated with particular enzyme loci. The approach has had some success but is always open to the objection that the loci for which the neutral hypothesis has been falsified are a small and biased sample.

In general, then, the problems raised by the neutral mutation theory and by optimization theory are wholly different. The latter is concerned with traits that differ between species and that can hardly be selectively neutral but whose selective significance is not fully understood.

A more serious difficulty for optimization theory is the occurrence of maladaptive traits. Optimization is based on the assumption that the population is adapted to the contemporary environment, whereas evolution is a process of continuous change. Species lag behind a changing environment. This is particularly serious when studying species in an environment that has recently been drastically changed by man. For example, Lack (48) argued that the number of eggs laid by a bird maximizes the number of surviving young. Although there is much supporting evidence, there are some apparent exceptions. For example, the gannet *Sula bassana* lays a single egg. Studying gannets on the Bass Rock, Nelson (71) found that if a second egg is added, the pair can successfully raise two young. The explanation can hardly be a lack of genetic variability, because species nesting in the Humboldt current off Peru lay two or even three eggs and successfully raise the young.

Lack (48) suggests that the environment for gannets may recently have improved, as evidenced by the recent increase in the population on the Bass Rock. Support for this interpretation comes from the work of Jarvis (39) on the closely related *S. capensis* in South Africa. This species typically lays one egg, but 1 percent of nests contain two. Using methods similar to Nelson's, Jarvis found that a pair can raise two chicks to fledgings, but that the average weight of twins was lower than singles, and in each nest one twin was always considerably lighter than its fellow. There is good evidence that birds fledging below the average weight are more likely to die soon after. Difficulties of a similar kind arise for the glaucous gull (see 45).

The undoubted existence of maladaptive traits, arising because evolutionary change is not instantaneous, is the most serious obstacle to the testing of optimization theories. The difficulty must arise; if species were perfectly adapted, evolution would cease. There is no easy way out. Clearly a wholesale reliance on evolutionary lag to save hypotheses that would otherwise be falsified would be fatal to the whole research program. The best we can do is to invoke evolutionary lag sparingly, and only when there are independent grounds for believing that the environment has changed recently in a relevant way.

What then is the status of the concept of adaptation? In the strong form—that all organs are perfectly adapted—it is clearly false; the vermiform appendix is sufficient to refute it. For Darwin, adaptation was an obvious fact that required an explanation; this still seems a sensible point of view. Adaptation can also be seen as a necessary consequence of natural selection. The latter I regard as a refutable scientific theory (60); but it must be refuted, if at all, by genetic experiment and not by the observation of complex behavior.

Critiques of Optimization Theory

Lewontin (55) raises a number of criticisms, which I discuss in turn.

Do Organs Solve Problems?

Most organs have many functions. Therefore, if a hypothesis concerning function fails correctly to predict behavior, it can always be saved by proposing an additional function. Thus hypotheses become irrefutable and metaphysical, and the whole program merely a test of ingenuity in conceiving possible functions. Three examples follow: the first is one used by Lewontin.

Orians and Pearson (73) calculated the optimal food item size for a bird, on the assumption that food intake is to be maximized. They found that the items diverged from random in the expected direction, but did not fit the prediction quantitatively. They explained the discrepancy be saying that a bird must visit its nest frequently to discourage predators. Lewontin (54) comments:

> This is a paradigm for adaptive reconstruction. The problem is originally posed as efficiency for food-gathering. A deviation of behavior from random, in the direction predicted, is regarded as strong support for the adaptive explanation of the behavior and the discrepancy from the predicted optimum is accounted for by an ad hoc secondary problem which acts as a constraint on the solution to the first. . . . By allowing the theorist to postulate various combinations of "problems" to which manifest traits are optimal "solutions," the adaptationist programme makes of adaptation a metaphysical postulate, not only incapable of refutation, but necessarily confirmed by every observation. This is the caricature that was immanent in Darwin's insight that evolution is the product of natural selection.

It would be unfair to subject Orians alone to such criticism, so I offer two further examples from my own work.

First, as explained earlier, Maynard Smith and Savage (66) predicted qualitative features of mammalian gaits. However, their model failed to give a correct quantitative prediction. I suspect that if the model were modified to allow for wind resistance and the visco-elastic properties of muscle, the quantitative fit would be improved; at present, however, this is pure speculation. In fact, it looks as if a model that gives quantitatively precise predictions will be hard to devise (1).

Second, Maynard Smith and Parker (64) predicted that populations will vary in persistence or aggressiveness in contest situations, but that individuals will not indicate their future behavior by varying levels of intensity of display. Rohwer (84) describes the expected variability in aggressivity in the Harris sparrow in winter flocks, but also finds a close correlation between aggressivity and a signal (amount of black in the plumage). I could point to the first observation as a confirmation of our theory, and explain how, by altering the model (by changing the phenotype set to permit the detection of cheating), one can explain the second.

What these examples, and many others, have in common is that a model gives predictions that are in part confirmed by observation but that are contradicted in some important respect. I agree with Lewontin that such discrepancies are inevitable if a simple model is used, particularly a model that assumes each organ or behavior to serve only one function. I also agree that if the investigator adds assumptions to his model to meet each discrepancy, there is no way in which the hypothesis of adaptation can be refuted. But the hypothesis of adaptation is not under test.

What is under test is the specific set of hypotheses in the particular model. Each of the three example models above has been falsified, at least as a complete explanation of these particular data. But since all have had qualitative success, it seems quite appropriate to modify them (e.g., by allowing for predation, for wind resistance, for detection of cheating). What is not justified is to modify the model and at the same time to claim that the model is confirmed by observation. For example, Orians would have to show that his original model fits more closely in species less exposed to predation. I would have to show that Rohwer's data fit the ''mixed ESS'' model in other ways—in particular, that the fitness of the different morphs is approximately equal. If, as may well be the case, the latter prediction of the ESS model does not hold, it is hard to see how it could be saved.

If the ESS model proves irrelevant to the Harris sparrow, it does not follow, however, that it is never relevant. By analogy, the assertion is logically correct that there will be a stable polymorphism if the heterozygote at a locus with two alleles is fitter than either homozygote. The fact that there are polymorphisms not maintained by heterosis does not invalidate the logic. The (difficult) empirical question is whether polymorphisms are often maintained by heterosis. I claim a similar logical status for the prediction of a mixed ESS.

In population biology we need simple models that make predictions that hold qualitatively in a number of cases, even if they are contradicted in detail in all of them. One can say with some confidence, for example, that no model in May's *Stability and Complexity in Model Ecosystems* describes exactly any actual case, because no model could ever include all relevant features. Yet the models do make qualitative predictions that help to explain real ecosystems. In the analysis of complex systems, the best we can hope for are models that capture some essential feature.

To summarize my comments on this point, Lewontin is undoubtedly right to complain if an optimizer first explains the discrepancy between theory and observation by introducing a new hypothesis, and then claims that his modified theory has been confirmed. I think he is mistaken in supposing that the aim of optimization theories is to confirm a general concept of adaptation.

Is There Genetic Variance?

Natural selection can optimize only if there is appropriate genetic variance. What justification is there for assuming the existence of such variance? The main justification is that, with rare exceptions, artificial selection has always proved effective, whatever the organism or the selected character (53).

A particular difficulty arises because genes have pleiotropic effects, so that selection for trait A may alter trait B; in such cases, any attempt to explain the changes in B in functional terms is doomed to failure. There are good empirical grounds for doubting whether the difficulty is as serious as might be expected from the widespread nature of pleiotropy. The point can best be illustrated by a particular example. Lewontin (54) noted that in primates there is a constant allometric relationship between tooth size and body size. It would be a waste of time, therefore, to seek a functional explanation of the difference between the tooth size of the gorilla and of the rhesus monkey, since the difference is probably a simple consequence of the difference in body size.

It is quite true that for most teeth there is a constant allometric relationship between tooth and body size, but there is more to it than that (36). The canine teeth (and the teeth occluding with them) of male primates are often larger than those of females, even when allowance has been made for the difference in body size. This sex difference is greater in species in which males compete for females than in monogamous species, and greater in ground-living species (which are more exposed to predation) than in arboreal ones. Hence, there is sex-limited genetic variance for canine tooth size, independent of body size, and the behavioral and ecological correlations suggest that this variance has been the basis of adaptation. It would be odd if there were tooth-specific, sex-limited variance, but no variance for the relative size of the teeth as a whole. However, there is some evidence for the latter. The size of the cheek teeth in females (relative to the size predicted from their body size) is significantly greater in those species with a higher proportion of leaves (as opposed to fruit, flowers, or animal matter) in their diets.

Thus, although at first sight the data on primate teeth suggest that there may be nothing to explain in functional terms, a more detailed analysis presents quite a different picture. More generally, changes in allometric relationships can and do occur during evolution (30).

I have quoted Lewontin as a critic of adaptive explanation, but it would misinterpret him to imply that he rejects all such explanations. He remarks (54) that "the serious methodological difficulties in the use of adaptive arguments should not blind us to the fact that many features of organisms are adaptations to obvious environmental 'problems.'" He goes on to argue that if natural selection is to produce adaptation, the mapping of character states into fitnesses must have two characteristics: "continuity" and "quasi-independence." By continuity is meant that small changes in a character result in small changes in the ecological relations of the organism; if this were not so, it would be hard to improve a character for one role without ruining it for another. By quasi-independence is meant that the developmental paths are such that a variety of mutations may occur, all with the same effect on the primary character, but with different effects on other characters. It is hard to think of better evidence for quasi-independence than the evolution of primate canines.

To sum up this point, I accept the logic of Lewontin's argument. If I differ from him (and on this point he is his own strongest critic), it is in thinking that genetic variance of an appropriate kind will usually exist. But it may not always do so.

It has been an implicit assumption of optimization models that the optimal phenotype can breed true. There are two kinds of reasons why this might not be true. The first is that the optimal phenotype may be produced by a heterozygote. This would be a serious difficulty if one attempted to use optimization methods to analyze the genetic structure of populations, but I think that would be an inappropriate use of the method. Optimization models are useful for analyzing phenotypic evolution, but not the genetic structuring of populations. A second reason why the optimal phenotype may not breed true is more serious: the evolutionarily stable population may be phenotypically variable. (This point is discussed further in the section on Games, below).

The assumption concerning the phenotype set is based on the range of variation observable within species, the phenotypes of related species, and on plausible guesses at what phenotypes might arise under selection. It is rare to have any information on the genetic basis of the phenotypic variability. Hence, although it is possible to introduce specific genetic assumptions into optimization models (e.g., 2, 89), this greatly complicates the analysis. In general, the assumption of "breeding true" is reasonable in particular applications; models in which genes appear explicitly need to be analyzed to decide in what situations the assumption may mislead us.

The Effects of History

If, as Wright (101) suggested, there are different "adaptive peaks" in the genetic landscape, then depending on initial conditions, different populations faced with identical "problems" may finish up in different stable states. Such divergence may be exaggerated if evolution takes the form of a "game" in which the optimal phenotype for one individual depends on what others are doing (see the section on Games, below). An example is Fisher's (28) theory of sexual selection, which can lead to an "autocatalytic" exaggeration of initially small differences. Jacob (38) has recently emphasized the importance of such historical accidents in evolution.

As an example of the difficulties that historical factors can raise for functional explanations, consider the evolution of parental care. A simple game-theory model (62) predicts that for a range of ecological parameters either of two patterns would be stable: male parental care only, or female care only. Many fish and amphibia show one or the other of these patterns. At first sight, the explanation of why some species show one pattern and others the other seems historical; the reasons seem lost in an unknown past. However, things may not be quite so bad. At a recent discussion of fish behavior at See-Wiesen the suggestion emerged that if uniparental care evolved from no parental care, it would be male care, whereas if it evolved from biparental care it would be female care. This prediction is plausible in the light of the original game-theory model, although not a necessary consequence of it. It is, however, testable by use of the comparative data; if it is true, male care should occur in families that also include species showing no care, and female care in families that include species showing biparental care. This may not prove to be the case; the example is given to show that even if there are alternative adaptive peaks, and in the absence of a relevant fossil record, it may still be possible to formulate testable hypotheses.

What Optimization Criterion Should One Use?

Suppose that, despite all difficulties, one has correctly identified the ''problem.'' Suppose, for example, that in foraging it is indeed true that an animal should maximize *E*, its rate of energy intake. We must still decide in what circumstances to maximize *E*. If the animal is alone in a uniform environment, no difficulty arises. But if we allow for competition and for a changing environment, several choices of optimization procedure are possible. For example, three possibilities arise if we allow just for competition:

1. The ''maximum'' solution: Each animal maximizes *E* on the assumption that other individuals behave in the least favorable way for it.
2. The ''Pareto'' point: The members of the population behave so that no individual can improve its intake without harming others.
3. The ESS: The members of the population adopt feeding strategy *I* such that no mutant individual adopting a strategy other than *I* could do better than typical members.

These alternatives are discussed further in the section on Games, below. For the moment, it is sufficient to say that the choice among them is not arbitrary, but follows from assumptions about the mode of inheritance and the population structure. For individual selection and parthenogenetic inheritance, the ESS is the appropriate choice.

Lewontin's criticism would be valid if optimizers were in the habit of assuming the truth of what Haldane once called ''Pangloss's theorem,'' which asserts that animals do those things that maximize the chance of survival of their species. If optimization rested on Pangloss's theorem it would be right to reject it. My reason for thinking that Lewontin regards optimization and Pangloss's theorem as equivalent is that he devotes the last section of his paper to showing that in *Drosophila* a characteristic may be established by individual selection and yet may reduce the competitive ability of the population relative to others. The point is correct and important, but in my view does not invalidate most recent applications of optimization.

The Methodology of Testing

The crucial hypothesis under test is usually that the model correctly incorporates the selective forces responsible for the evolution of a trait. Optimization models sometimes make fairly precise quantitiative predictions that can be tested. However, I shall discuss the question how functional explanations can be tested more generally, including cases in which the predictions are only qualitative. It is convenient to distinguish comparative, quantitative, and individual-variation methods.

Comparative Tests

Given a functional hypothesis, there are usually testable predictions about the development of the trait in different species. For example, two main hypotheses have been proposed to account for the greater size of males in many mammalian species: It is a consequence of competition among males for females; or it arises because the two sexes use different resources. If the former hypothesis is true, dimorphism should be greater in harem-holding and groupliving species, whereas if the latter is true it should be greater in monogamous ones, and in those with a relatively equal adult sex ratio.

Clutton-Brock et al. (16) have tested these hypotheses by analyzing 42 species of primates (out of some 200 extant species) for which adequate breeding data are available. The data are consistent with the sexual selection hypothesis, and show no sign of the trend predicted by the resource differentiation hypothesis. The latter can therefore be rejected, at least as a major cause of sexual dimorphism in primates. It does not follow that intermale competition is the only relevant selective factor (82). Nor do their observations say anything about the causes of sexual dimorphism in other groups. It is interesting (though not strictly relevant at this point) that the analysis also showed a strong correlation between female body size and degree of dimorphism. This trend, as was first noted by Rensch (83), occurs in a number of taxa, but has never received an entirely satisfactory explanation.

The comparative method requires some criterion for inclusion of species. This may be purely taxonomic (e.g., all primates, all passerine birds), or jointly taxonomic and geographic (e.g., all African ungulates, all passerines in a particular forest). Usually, some species must be omitted because data are not available. Studies on primates can include a substantial proportion of extant species (16, 68); in contrast, Schoener (86), in one of the earliest studies of this type, included all birds for which data were available and which also met certain criteria of territoriality, but he had to be content with a small fraction of extant species. It is therefore important to ask whether the sample of species is biased in ways likely to affect the hypothesis under test. Most important is that there be some criterion of inclusion, since otherwise species may be included simply because they confirm (or contradict) the hypothesis under test.

Most often, limitations of data will make it necessary to impose both taxonomic and geographic criteria. This need not prevent such data from being valuable, either in generating or in testing hypotheses; examples are analyses of flocking in birds (7, 31) and of breeding systems in forest plants (3, 4).

A second kind of difficulty concerns the design of significance test. Different species cannot always be treated as statistically independent. For example, all gibbons are monogamous, and all are arboreal and frugivorous, but since all may be descended from a single ancestor with these properties, they should be treated as a single case in any test of association (not that any is suspected). To take an actual example of this difficulty, Lack (49) criticized Verner and Willson's (98) conclusion that polygamy in passerines is associated with marsh and prairie habitats on the grounds that many of the species concerned belong to a single family, the Icteridae.

Statistical independence and other methodological problems in analyzing comparative data are discussed by Clutton-Brock and Harvey (17). In analyzing the primate data, they group together as a single observation all congeneric species belonging to the same ecological category. This is a conservative procedure, in that it is unlikely to find spurious cases of statistical significance. Their justification for treating genera, but not families, as units is that for their data there are significant differences between genera within families for seven of the eight ecological and behavioral variables, but significant additional variation between families for only two of them. It may be, however, that a more useful application of statistical methods is their use (17) of partial regression, which enables them to examine the effects of a particular variable when the effects of other variables have been removed, and to ask how much of the total variation in some trait is accounted for by particular vaiables.

Quantitative Tests

Quantitative tests can be illustrated by reference to some of the predictions of foraging theory. Consider first the problem of optimal diet. The following model situation has been widely assumed. There are a number of different kinds of food items. An animal can search simultaneously for all of them. Each item has a characteristic food value and "handling time" (the time taken to capture and consume it). For any given set of densities and hence frequencies of encounter, the animal must only decide which items it should consume and which ignore.

Pyke et al. (81) remark that no fewer than eight authors have independently derived the following basic result. The animal should rank the items in order of $V =$ food value/handling time. Items should be added to the diet in rank order, provided that for each new item the value of V is greater than the rate of food intake for the diet without the addition. This basic result leads to three predictions:

1. Greater food abundance should lead to greater specialization. This qualitative prediction was first demonstrated by Ivlev (37) for various fish species in the laboratory, and data supporting it have been reviewed by Schoener (87). Curio (25) quotes a number of cases that do not fit.
2. For fixed densities, a food type should either be always taken, or never taken.
3. Whether a food item should be taken is independent of its density, and depends on the densities of food items of higher rank.

Werner and Hall (100) allowed blue-gill sunfish to feed on *Daphnia* of three different size classes; the diets observed agreed well with the predictions of the model. Krebs et al. (47) studies great tits foraging for parts of mealworms on a moving conveyor belt. They confirmed prediction 3 but not 2; that is, they found that whether small pieces were taken was independent of the density of small pieces, but, as food abundance rose, small pieces were dropped only gradually from the diet. Goss-Custard (29) has provided field evidence confirming the model from a study of redshank feeding on marine worms of different sizes, and Pulliam (80) has confirmed it for chipping sparrows feeding on seeds.

Turning to the problem of how long an animal should stay in a patch before moving to another, there is again a simple prediction, which Charnov (10) has called the ''Marginal Value Theorem'' (the same theorem was derived independently by Parker and Stuart [77] in a different context). It asserts that an animal should leave a patch when its rate of intake in the patch (its ''marginal'' rate) drops to the average rate of intake for the habitat as a whole. It is a corollary that the marginal rate should be the same for all patches in the habitat. Two laboratory experiments on tits (20, 46) agree well with the prediction.

A more general problem raised by these experiments is discussed by Pyke et al. (81). How does an animal estimate the parameters it needs to know before it can perform the required optimization? How much time should it spend acquiring information? Sometimes these questions may receive a simple answer. Thus the results of Krebs et al. (46) suggest that a bird leaves a patch if it has not found an item of food for some fixed period r (which varied with the overall abundance of food). The bird seems to be using r, or rather $1/r$, as an estimate of its marginal capture rate. But not all cases are so simple.

Individual Variation

The most direct way of testing a hypothesis about adaptation is to compare individuals with different phenotypes, to see whether their fitnesses vary in the way predicted by the hypothesis. This was the basis of Kettlewell's (42) classic demonstration of selection on industrial melanism in moths. In principle, the individual differences may be produced by experimental interference (Curio's [23] ''method of altering a character'') or they may be genetic or of unknown origin (Curio's ''method of variants''). Genetic differences are open to the objection that genes have pleiotropic effects, and occasionally are components of supergenes in which several closely linked loci affecting the same function are held in linkage disequilibrium, so that the phenotypic difference responsible for the change in fitness may not be the one on which attention is concentrated. This difficulty, however, is trivial compared to that which arises when two species are compared.

The real difficulty in applying this method to behavioral differences is that suitable individual differences are often absent and experimental interference is impractical. Although it is hard to alter behavior experimentally, it may be possible to alter its consequences. Tinbergen et al. (94) tested the idea that gulls remove egg shells from the nest because the shells attract predators to their eggs and young; they placed egg shells close to eggs and recorded a higher predation rate.

However, the most obvious field of application of this method arises when a population is naturally variable. Natural variation in a phenotype may be maintained by frequency-dependent selection; in game-theoretical terms, the stable state may be a mixed strategy. If a particular case of phenotypic variability (genetic or not) is thought to be maintained in this way, it is important to measure the fitnesses of individuals with different phenotypes. At a mixed ESS (which assumes parthenogenetic inheritance) these fitnesses are equal; with sexual reproduction, exact equality is not guaranteed, but approximate equality is a reasonable expectation (91). If the differences are not genetic, we still expect a genotype to evolve that adopts the different strategies with frequencies that equalize their payoffs.

The only test of this kind known to me is Parker's (76) measurement of the mating success of male dungflies adopting different strategies. His results are consistent with a ''mixed ESS'' interpretation; it is not known whether the differences are genetic. The importance of tests of this kind lies in the fact that phenotypic variability can have other explanations; for example, it may arise from random environmental effects, or from genes with heterotic effects. In such cases, equality of fitness between phenotypes is not expected.

Mathematical Approaches to Optimization

During the past twenty years there has been a rapid development of mathematical techniques aimed at solving problems of optimization and control arising in economics and engineering. These stem from the concepts of ''dynamic programming'' (5) and of the ''maximum principle'' (79). The former is essentially a computer procedure to seek the best control policy in particular cases without the hopelessly time-consuming task of looking at every possibility. The latter is an extension of the classic methods of the calculus of variations that permits one to allow for ''inequality'' constraints on the state and control variables (e.g., in the resource allocation model discussed below, the proportion u of the available resources allocated to seeds must obey the constraint $u < 1$).

This is not the place to describe these methods, even if I were competent to do so. Instead, I shall describe the kinds of problems that can be attacked. If a biologist has a problem of one of these kinds, he would do best to consult a mathematician. For anyone wishing to learn more of the mathematical background, Clark (12) provides an excellent introduction.

I discuss in turn ''optimization,'' in which the problem is to choose an optimal policy in an environment without competitors; ''games,'' in which the environment includes other "players" who are also attempting to optimize something; and "games of inclusive fitness," in which the "players" have genes in common. I shall use as an illustration the allocation of resources between growth and reproduction.

Optimization

Choice of a Single Value The simplest type of problem, which requires for its solution only the technique of differentiation, is the choice of a value for a single parameter. For example, in discussing the evolution of gaits, Maynard Smith and Savage (66) found an expression for P, the power output, as a function of the speed V, of size S, and of J, the fraction of time for which all four legs are off the ground. By solving the equation $\mathrm{d}P/\mathrm{d}J = 0$, an equation $J = \mathrm{f}(V, S)$ was obtained, describing the optimum gait as a function of speed and size.

Few problems are as simple as this, but some more complex cases can be reduced to problems of this kind, as will appear below.

A Simple Problem in Sequential Control Most optimization theory is concerned with how a series of sequential decisions should be taken. For example, consider the growth of an annual plant (19, 69). The rate at which the plant can accumulate resources depends on its size. The resources can be allocated either for further growth, or to seeds, or divided between them. For a fixed starting size and length of season, how should the plant allocate its resources so as to maximize the total number of seeds produced?

In this problem the "state" of the system at any time is given simply by the plant's size, x; the "control variable" $u(t)$ is the fraction of the incoming resource allocated to seeds at time t; the "constraints" are the initial size, the length of the season, the fact that $u(t)$ must lie between 0 and 1, and the "state equation,"

$$\mathrm{d}x/\mathrm{d}t = \mathrm{F}[x(t), u(t)], \tag{1}$$

which describes how the system changes as a function of its state and of the control variable.

If equation 1 is linear in u, it can be shown that the optimal control is "bang-bang"—that is, $u(t) = 0$ up to some critical time t^*, and subsequently $u(t) = 1$. The problem is thus reduced to finding the single value, t^*. But if equation 1 is nonlinear, or has stochastic elements, the optimal control may be graded.

More Complex Control Problems Consider first the "state" of the system. This may require description by a vector rather than by a single variable. Thus suppose the plant could also allocate resources to the production of toxins that increased its chance of survival. Then its state would require measures of both size and toxicity. The state description must be sufficient for the production of a state equation analogous to equation 1. The state must also include any information used in determining the control function $u(t)$. This is particularly important when analyzing the behavior of an animal that can learn. Thus suppose that an animal is foraging and that its decisions on whether to stay in a given patch or to move depend on information it has acquired about the distribution of food in patches; then this information is part of the state of the animal. For a discussion, see (20).

Just as the state description may be multidimensional, so may the control function; for example, for the toxic plant the control function must specify the allocation both to seeds and to toxins.

The state equation may be stochastic. Thus the growth of a plant depends on whether it rains. A plant may be supposed to "know" the probability of rain (i.e., its genotype may be adapted to the frequency of rain in previous generations) but not whether it will actually rain. In this case, a stochastic state equation may require a graded control. This connection between stochasticity and a "compromise" response as opposed to an all-or-none one is a common feature of optimal control. A second example is the analysis by Oster and Wilson (75) of the optimal division into castes in social insects: A predictable environment is likely to call for a single of worker, while an uncertain one probably calls for a division into several castes.

Reverse Optimality McFarland (67) has suggested an alternative approach. The typical one is to ask how an organism should behave in order to maximize its fitness. Mathematically, this requires that one define an "objective function" that must be "maximized" ("objective" here means "aim" or "goal"); in the plant example, the objective function is the number of seeds produced, expressed as a function of x and $u(t)$. But a biologist may be faced with a different problem. Suppose that he knew, by experiment, how the plant actually allocates its resources. He could then ask what the plant is actually maximizing. If the plant is perfectly adapted, the objective function so obtained should correspond to what Sibly and McFarland (88) call the "cost function"—that is, the function that should be maximized if the organism is maximizing its fitness. A discrepancy would indicate maladaptation.

There are difficulties in seeing how this process of reverse optimality can be used. Given that the organism's behavior is "consistent" (i.e., if the prefers A to B and B to C, it prefers A to C), it is certain that its behavior maximizes *some* objective function; in general there will be a set of functions maximized. Perfect adaptation then requires only that the cost function correspond to one member of this set. A more serious difficulty is that it is not clear what question is being asked. If a discrepancy is found, it would be hard to say whether this was because costs had been wrongly measured or because the organism was maladapted. This is a particular example of my general point that it is not sensible to test the hypothesis that animals optimize. But it may be that the reverse optimality approach will help to analyze how animals in fact make decisions.

Games

Optimization of the kind just discussed treats the environment as fixed, or as having fixed stochastic properties. It corresponds to that part of population genetics that assumes fitnesses to be independent of genotype frequencies. A number of selective processes have been proposed as frequency-dependent, including predation (13, 70) and disease (15, 32). The maintenance of polymorphism in a varied environment (50) is also best seen as a case of frequency-dependence (59). The concept can be applied directly to phenotypes.

The problem is best formulated in terms of the theory of games, first developed (99) to analyze human conflicts. The essence of a game is that the best strategy to adopt depends on what one's opponent will do; in the context of evolution, this means that the fitness of a phenotype depends on what others are present; i.e., fitnesses are frequency dependent.

The essential concepts are those of a "strategy" and a "payoff matrix." A strategy is a specification of what a "player" will do in every situation in which it may find itself; in the plant example, a typical strategy would be to allocate all resources to growth for twenty days, and then divide resources equally between growth and seeds. A strategy may be "pure" (i.e., without chance elements) or "mixed" (i.e., of the form "do A with probability p and B with probability $1 - p$," where A and B are pure strategies).

The "payoff" to an individual adopting strategy A in competition to one adopting B is written $E(A, B)$, which expresses the expected *change* in the fitness of the player adopting A if his opponent adopts B. The evolutionary model is then of a population of individuals adopting different strategies. They pair off at random, and their fitnesses change according to the payoff matrix. Each individual then produces offspring identical to itself, in numbers proportional to the payoff it has accumulated. Inheritance is thus parthenogenetic, and selection acts on the population is infinite, so that the chance of meeting an opponent adopting a particular strategy is independent of one's own strategy.

The population will evolve to an evolutionarily stable strategy, or ESS, if one exists (64). An ESS is a strategy that, if almost all individuals adopt it, no rare mutant can invade. Thus let I be an ESS, and J a rare mutant strategy of frequency $p \ll 1$. Writing the fitnesses of I and J as $W(I)$ and $W(J)$,

$$W(I) = C + (1 - p)E(I, I) + pE(I, J);$$

$$W(J) = C + (1 - p)E(J, I) + pE(J, J).$$

In these equations C is the fitness of an individual before engaging in a contest. Since I is an ESS, $W(I) > W(J)$ for all $J \neq I$; that is, remembering that p is small, either

Table 6.1
Payoff matrix for a game

	Player 2	
Player 1	*A*	*B*
A	1	5
B	2	4

Note: The values in the matrix give the payoff to Player 1.

$$E(I,I) > E(J,I), \qquad \textit{or}$$
$$E(I,I) = E(J,I), \qquad \text{and} \qquad E(I,J) > E(J,J). \tag{2}$$

These conditions (expressions 2) are the definition of an ESS.

Consider the matrix in table 6.1. For readers who prefer a biological interpretation, *A* is "Hawk" and *B* is "Dove"; thus *A* is a bad strategy to adopt against *A*, because of the risk of serious injury, but a good strategy to adopt against *B*, and so on.

The game has no pure ESS, because $E(A,A) < E(B,A)$ and $E(B,B) < E(A,B)$. It is easy to show that the mixed strategy—playing *A* and *B* with equal probability—is an ESS. It is useful to compare this with other "solutions," each of which has a possible biological interpretation:

The Maximin Solution This is the pessimist's solution, playing the strategy that minimizes your losses if your opponent does what is worst for you. For our matrix, the maximin strategy is always to play *B*. Lewontin (52) suggested that this strategy is appropriate if the "player" is a species and its opponent nature: The species should minimize its chance of extinction when nature does its worst. This is the "existential game" of Slobodkin and Rapoport (92). It is hard to see how a species could evolve this strategy, except by group selection. (Note that individual selection will not necessarily minimize the chance of death: A mutant that doubled the chance that an individual would die before maturity, but that quadrupled its fecundity if it did survive, would increase in frequency.)

The Nash Equilibrium This is a pair of strategies, one for each player, such that neither would be tempted to change his strategy as long as the other continues with his. If in our matrix, player 1 plays *A* and 2 plays *B*, we have a Nash equilibrium; this is also the case if 1 plays *B* and 2 plays *A*. A population can evolve to the Nash point if it is divided into two classes, and if members of one class compete only with members of the other. Hence it is the appropriate equilibrium in the "parental investment" game (62), in which all contests are between a male and a female. The ESS is subject to the added constraint that both players must adopt the same strategy.

The Group Selection Equilibrium If the two players have the same genotype, genes in either will be favored that maximize the sum of their payoffs. For our matrix both must play strategy *B*. The problem of the stable strategy when the players are related but not identical is discussed in the section on Games between Relatives, below.

It is possible to combine the game-theoretical and optimization approaches. Mirmirani and Oster (69) make this extension in their model of resource allocation in plants. They ask two questions. What is the ESS for a plant growing in competition with members of its own species? What is the ESS when two species compete with one another?

Thus consider two competing plants whose sizes at time t are P_1 and P_2. The effects of competition are allowed for by writing

$$\mathrm{d}P_1/\mathrm{d}t = (r_1 - e_1 P_2)(1 - u_1)P_1,$$

$$\mathrm{d}P_2/\mathrm{d}t = (r_2 - e_2 P_1)(1 - u_2)P_2, \tag{3}$$

where u_1 and u_2 are the fractions of the available resources allocated to seeds. Let $J_1[u_1(t), u_2(t)]$ be the total seed production of plant 1 if it adopts the allocation strategy $u_1(t)$ and its competitor adopts $u_2(t)$. Mirmirani and Oster seek a stable pair of strategies $u_1^*(t)$, $u_2^*(t)$, such that

$$J_1[u_1(t), u_2^*(t)] \leq J_1[u_1^*(t), u_2^*(t)], \qquad \text{and}$$

$$J_2[u_1^*(t), u_2(t)] \leq J_2[u_1^*(t), u_2^*(t)]. \tag{4}$$

That is, they seek a Nash equilibrium, such that neither competitor could benefit by unilaterally altering its strategy. They find that the optimal strategies are again "bang-bang," but with earlier switching times than in the absence of competition. Strictly, the conditions indicated by expressions 4 are correct only when there is competition between species, and when individuals of one species compete only with individuals of the other; formally this would be so if the plants grew alternately in a linear array. The conditions indicated by expressions 4 are not appropriate for intraspecific competition, since they permit $u_1^*(t)$ and $u_2^*(t)$ to be different, which could not be the case unless individuals of one genotype competed only with individuals of the other. For intraspecific competition ($r_1 = r_2$, $e_1 = e_2$), the ESS is given by

$$J_1[u_1(t), u_1^*(t)] \leq J_1[u_1^*(t), u_1^*(t)]. \tag{5}$$

As it happens, for the plant growth example equations 4 and 5 give the same control function, but in general this need not be so.

The ESS model assumes parthenogenetic inheritance, whereas most interesting populations are sexual. If the ESS is a pure stategy, no difficulty arises; a genetically homogeneous sexual population adopting the strategy will also be stable. If the ESS is a mixed strategy that can be achieved by a single individual with a variable behavior, there is again no difficulty. If the ESS is a mixed one that can be achieved only by a population of pure strategists in the appropriate frequencies, two difficulties arise:

1. Even with the parthenogenetic model, the conditions expressed in expressions 2 do not guarantee stability. (This was first pointed out to me by Dr. C. Strobeck.) In such cases, therefore, it is best to check the stability of the equilibrium, if necessary by simulation; so far, experience suggests that stability, although not guaranteed, will usually be found.
2. The frequency distribution may be one that is incompatible with the genetic mechanism. This difficulty, first pointed out by Lewontin (52), has recently been investigated by Slatkin (89–91) and by Auslander et al. (2). It is hard to say at present how serious it will prove to be; my hope is that a sexual population will usually evolve a frequency distribution as close to the ESS as its genetic mechanism will allow.

Games between Relatives

The central concept is that of "inclusive fitness" (33). In classical population genetics we ascribe to a genotype I a "fitness" W, corresponding to the expected number of offspring produced by I. If, averaged over environments and genetic backgrounds, the effect of substituting allele A for a is to increase W, allele A will increase in frequency. Following Oster et al. (74) but ignoring unequal sex ratios, Hamilton's proposal is that we should replace W_i by the inclusive fitness, Z_i, where

$$Z_i = \sum_{j=1}^{R} r_{ij} W_j, \tag{6}$$

where the summation is over all R relatives of I; r_{ij} is the fraction of J's genome that is identical by descent to alleles in I; and W_j is the expected number of offspring of the jth relative of I. (If $J = I$, then equation 6 refers to the component of inclusive fitness from an individual's own offspring.)

An allele A will increase in frequency if it increases Z rather than just W. Three warnings are needed:

1. It is usual to calculate r_{ij} from the pedigree connecting I and J (as carried out, for example, by Malécot (58)). However, if selection is occurring, r_{ij} so estimated is only approximate, as are predictions based on equation 6 (35).
2. Some difficulties arose in calculating appropriate values of r_{ij} for haplodiploids; these were resolved by Crozier (22).
3. If the sex ratio is not unity, additional difficulties arise (74).

Mirmirani and Oster (69) have extended their plant-growth model along these lines to cover the case when the two competitors are genetically related. They show that as r increases, the switching time becomes earlier and the total yield higher.

Conclusion

The role of optimization theories in biology is not to demonstrate that organisms optimize. Rather, they are an attempt to understand the diversity of life.

Three sets of assumptions underlie an optimization model. First, there is an assumption about the kinds of phenotypes or strategies possible (i.e., a ''phenotype set''). Second, there is an assumption about what is being maximized; ideally this should be the inclusive fitness of the individual, but often one must be satisfied with some component of fitness (e.g., rate of energy intake while foraging). Finally, there is an assumption, often tacit, about the mode of inheritance and the population structure; this will determine the type of equilibrium to which the population will move.

In testing an optimization model, one is testing the adequacy of these hypotheses to account for the evolution of the particular structures or patterns of behavior under study. In most cases the hypothesis that variation in the relevant phenotypes is selectively neutral is not a plausible alternative, because of the nature of the phenotypes chosen for study. However, it is often a plausible alternative that the phenotypes are not well adapted to current circumstances because the population is lagging behind a changing environment; this is a serious difficulty in testing optimization theories.

The most damaging criticism of optimization theories is that they are untestable. There is a real danger that the search for functional explanations in biology will degenerate into a test of ingenuity. An important task, therefore, is the development to an adequate methodology of testing. In many cases the comparative method is the most powerful; it is essential, however, to have clear criteria for inclusion or exclusion of species in comparative tests, and to use statistical methods with the same care as in the analysis of experimental results.

Tests of the quantitative predictions of optimization models in particular populations are beginning to be made. It is commonly found that a model correctly predicts qualitative features of the observations, but is contradicted in detail. In such cases the Popperian view would be that the original model has been falsified. This is correct, but it does not follow that the model should be abandoned. In the analysis of complex systems it is most unlikely that any simple model, taking into account only a few factors, can give quantitatively exact predictions. Given that a simple model has been falsified by observations, the choice lies between abandoning it and modifying it, usually by adding hypotheses. There can be no simple rule by which to make this choice; it will depend on how persuasive the qualitative predictions are, and on the availability of alternative models.

Mathematical methods of optimization have been developed with engineering and economic applications in mind. Two theoretical questions arise in applying these methods in biology. First, in those cases in which the fitnesses of phenotypes are frequency-dependent, the problem must be formulated in game-theoretical terms;

some difficulties then arise in deciding to what type of equilibrium a population will tend. A second and related set of questions arise when specific genetic assumptions are incorporated into the model, because it may be that a population with the optimal phenotype cannot breed true. These questions need further study, but at present there is no reason to doubt the adequacy of the concepts of optimization and of evolutionary stability for studying phenotypic evolution.

Acknowledgments

My thanks are due Dr. R. C. Lewontin for sending me two manuscripts that formed the starting point of this review, and Drs. G. Oster and R. Pulliam for their comments on and earlier draft. I was also greatly helped by preliminary discussions with Dr. E. Curio.

References

1. Alexander, R. M. 1977. Mechanics and scaling of terrestrial locomotion. In *Scale Effects in Animal Locomotion*, ed. T. J. Pedley, London, Academic Press, pp. 93–110.

2. Auslander, D., J. Guckenheimer, and G. Oster. 1978. Random evolutionarily stable strategies. *Theor. Pop. Biol.*, 13(2): 276–293.

3. Baker, H. G. 1959. Reproductive methods as factors in speciation in flowering plants. *Cold Spring Harbor Symp. Quant. Biol.*, 24: 177–191.

4. Bawa, K. S., and P. A. Opler. 1975. Dioecism in tropical forest trees. *Evolution*, 29: 167–179.

5. Bellman, R. 1957. *Dynamic Programming*, Princeton, N.J., Princeton University Press.

6. Bishop, D. T., and C. Cannings. 1978. A generalized war of attrition. *J. Theor. Biol.*, 70: 85–124.

7. Buskirk, W. H. 1976. Social systems in tropical forest avifauna. *Am. Nat.*, 110: 293–310.

8. Cain, A. J. 1964. The perfection of animals. In *Viewpoints in Biology*, ed. J. D. Carthy and C. L. Duddington, 3: 36–63.

9. ———, and P. H. Sheppard. 1954. Natural selection in *Cepaea*. *Genetics*, 39: 89–116.

10. Charnov, E. L. 1976. Optimal foraging, the marginal value theorem. *Theor. Pop. Biol.*, 9: 129–136.

11. ———, J. Maynard Smith, and J. J. Bull. 1976. Why be an hermaphrodite? *Nature*, 263: 125–126.

12. Clark, C. W. 1976. *Mathematical Bioeconomics*, N.Y., Wiley.

13. Clarke, B. 1962. Balanced polymorphism and the diversity of sympatric species. In *Taxonomy and Geography*, ed. d. Nichols, London, Syst. Assoc. Publ., 4: 47–70.

14. ———. 1975. The contribution of ecological genetics to evolutionary theory: detecting the direct effects of natural selection on particular polymorphic loci. *Genetics*, 79: 101–113.

15. ———. 1976. The ecological genetics of host-parasite relationships. In *Genetic Aspects of Host-Parasite Relationships*, ed. A. E. R. Taylor and R. Muller, Oxford, Blackwell, pp. 87–103.

16. Clutton-Brock, T. H., and P. H. Harvey. 1977. Primate ecology and social organisation. *J. Zool.*, London, 183: 1–39.

17. ———. 1977. Species differences in feeding and ranging behaviour in primates. In *Primate Ecology*, ed. T. H. Clutton-Brock, London, Academic, pp. 557–584.

18. ———, and B. Rudder. 1977. Sexual dimorphism, socioeconomic sex ratio and body weight in primates. *Nature*, 269: 797–800.

19. Cohen, D. 1971. Maximising final yield when growth is limited by time or by limiting resources. *J. Theor. Biol.*, 33: 299–307.

20. Cowie, R. J. 1977. Optimal foraging in great tits (*Parus major*). *Nature*, 268: 137–139.

21. Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory*, N.Y., Harper & Row.

22. Crozier, R. H. 1970. Coefficients of relationship and the identity of genes by descent in the Hymenoptera. *Am. Nat.*, 104: 216–217.

23. Curio, E. 1973. Towards a methodology of teleonomy. *Experientia*, 29: 1045–1058.

24. ———. 1975. The functional organization of anti-predator behaviour in the pied flycatcher: a study of avian visual perception. *Anim. Behav.*, 23: 1–115.

25. ———. 1976. *The Ethology of Predation*, Berlin, Springer-Verlag.

26. Darwin, C. 1877. *The Different Forms of Flowers on Plants of the Same Species*, London, John Murray.

27. Emlen, J. M. 1966. The role of time and energy in food preference. *Am. Nat.*, 100: 611–617.

28. Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*, London: Oxford Univ. Press, 291 pp.

29. Goss-Custard, J. D. 1977. Optimal foraging and the size selection of worms by redshank, *Tringa totanus*, in the field. *Anim. Behav.*, 25: 10–29.

30. Gould, S. J. 1971. Geometric scaling in allometric growth: a contribution to the problem of scaling in the evolution of size. *Am. Nat.*, 105: 113–116.

31. Grieg-Smith, P. W. 1978. The formation, structure and feeding of insectivorous bird flocks in West African savanna woodland. *Ibis*, 121(3): 284–297.

32. Haldane, J. B. S. 1949. Disease and evolution. *Ric. Sci.*, Suppl., 19: 68–76.

33. Hamilton, W. D. 1964. The genetical theory of social behavior. I and II. *J. Theor. Biol.*, 7: 1–16; 17–32.

34. ———. 1967. Extraordinary sex ratios. *Science*, 156: 477–488.

35. ———. 1972. Altruism and related phenomena, mainly in social insects. *Ann. Rev. Ecol. Syst.*, 3: 193–232.

36. Harvey, P. H., M. Kavanagh, and T. H. Clutton-Brock. 1978. Sexual dimorphism in primate teeth. *J. Zool.*, 186: 475–485.

37. Ivlev, V. S. 1961. *Experimental Ecology of the Feeding of Fishes*, New Haven, Yale Univ. Press.

38. Jacob, F. 1977. Evolution and tinkering. *Science*, 196: 1161–1166.

39. Jarvis, M. J. F. 1974. The ecological significance of clutch size in the South African gannet [*Sula capensis*. (Lichtenstein)]. *J. Anim. Ecol.*, 43: 1–17.

40. Kalmijn, A. J. 1971. The electric sense of sharks and rays. *J. Exp. Biol.*, 55: 371–383.

41. Kalmus, H., and C. A. B. Smith. 1960. Evolutionary origin of sexual differentiation and the sex-ratio. *Nature*, 186: 1004–1006.

42. Kettlewell, H. B. D. 1956. Further selection experiments on industrial melanism in the Lepidoptera. *Heredity*, 10: 287–301.

43. Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature*, 217: 624–626.

44. King, J. L., and T. H. Jukes. 1969. Non-Darwinian evolution: random fixation of selectively neutral mutations. *Science*, 164: 788–798.

45. Krebs, C. J. 1972. *Ecology*, N.Y., Harper & Row.

46. Krebs, J. R., J. C. Ryan, and E. L. Charnov. 1974. Hunting by expectation or optimal foraging? A study of patch use by chickadees. *Anim. Behav.*, 22: 953–964.

47. ———, J. T. Ericksen, M. I. Webber, and E. L. Charnov. 1977. Optimal prey selection in the great tit (*Parus major*). *Anim. Behav.*, 25: 30–38.

48. Lack, D. 1966. *Population Studies of Birds*, Oxford, Clarendon Press.

49. ———. 1968. *Ecological Adaptiations for Breeding in Birds*, London, Methuen.

50. Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. *Am. Nat.*, 87: 131–133.

51. Levins, R. 1962. Theory of fitness in a heterogeneous environment. I. The fitness set and adaptive function. *Am. Nat.*, 96: 361–373.

52. Lewontin, R. C. 1961. Evolution and the theory of games. *J. Theor. Biol.*, 1: 382–403.

53. ———. 1974. *The Genetic Basis of Evolutionary Change*, N.Y., Columbia Univ. Press.

54. ———. 1977. Adaptation. In *The Encyclopedia Einaudi*, Torino, Giulio Einaudi Edition.

55. ———. 1978. Fitness, survival and optimality. In *Analysis of Ecological Systems*, ed. D. H. Horn, R. Mitchell, G. R. Stairs, Columbus, Oh., Ohio State Univ. Press.

56. MacArthur, R. H. 1965. Ecological consequences of natural selection. In *Theoretical and Mathematical Biology*, ed. T. Waterman, H. Morowitz, N.Y., Blaisdell.

57. ———, and E. R. Pianka. 1966. On optimal use of a patch environment. *Am. Nat.*, 100: 603–609.

58. Malécot, G. 1969. *The Mathematics of Heredity*, transl. D. M. Yermanos. San Francisco, W. H. Freeman, 88 pp.

59. Maynard Smith, J. 1962. Disruptive selection, polymorphism and sympatric speciation. *Nature*, 195: 60–62.

60. ———. 1969. The status of neo-Darwinism. In *Towards a Theoretical Biology. 2: Sketches*, ed. C. H. Waddington, Edinburgh, Edinburgh Univ. Press, pp. 82–89.

61. ———. 1976. Sexual selection and the handicap principle. *J. Theor. Biol.*, 57: 239–242.

62. ———. 1977. Parental investment—a prospective analysis. *Anim. Behav.*, 25: 1–9.

63. ———. 1978. *The Evolution of Sex*, London, Cambridge Univ. Press.

64. ———, and G. A. Parker. 1976. The logic of asymmetric contests. *Anim. Behav.*, 24: 159–175.

65. ———, and G. R. Price. 1973. The logic of animal conflict. *Nature*, 246: 15–18.

66. ———, and R. J. G. Savage. 1956. Some locomotory adaptations in mammals. *Zool, J. Linn. Soc.*, 42: 603–622.

67. McFarland, D. J. 1977. Decision making in animals. *Nature*, 269: 15–21.

68. Milton, K., and M. L. May. 1976. Body-weight, diet and home range area in primates. *Nature*, 259: 459–462.

69. Mirmirani, M., and G. Oster. 1978. Competition, kin selection and evolutionarily stable strategies. *Theor. Pop. Biol.*, 13(3): 304–339.

70. Moment, G. 1962. Reflexive selection: a possible answer to an old puzzle. *Science*, 136: 262–263.

71. Nelson, J. B. 1964. Factors influencing clutch size and chick growth in the North Atlantic Gannet. *Sula bassana. Ibis*, 106: 63–77.

72. Norton-Griffiths, M. 1969. The organization, control and development of parental feeding in the oystercatcher (*Haematopus ostralegus*). *Behavior*, 34: 55–114.

73. Orians, G. H., and N. E. Pearson. 1978. On the theory of central place foraging. In *Analysis of Ecological Systems*, ed. D. H. Horn, R. Mitchell, G. R. Stairs, Columbus, Ohio State Univ. Press.

74. Oster, G., I. Eshel, and D. Cohen. 1977. Worker-queen conflicts and the evolution of social insects. *Theor. Pop. Biol.*, 12: 49–85.

75. ———, and E. O. Wilson. 1978. *Caste and Ecology in the Social Insects*, Princeton, N.J., Princeton Univ. Press.

76. Parker, G. A. 1974. The reproductive behaviour and the nature of sexual selection in *Scatophaga stercoraria* L. IX. Spatial distribution of fertilization rates and evolution of male search strategy within the reproductive area. *Evolution*, 28: 93–108.

77. ———, and R. A. Stuart. 1976. Animal behaviour as a strategy optimizer: evolution of resource assessment strategies and optimal emigration thresholds. *Am. Nat.*, 110: 1055–1076.

78. Pedley, T. J. 1977. *Scale Effects in Animal Locomotion*, London, Academic Press.

79. Pontryagin, L. S., V. S. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. 1962. *The Mathematical Theory of Optimal Processes*, N.Y., Wiley.

80. Pulliam, H. R. 1978. Do chipping sparrows forage optimally? A test of optimal foraging theory in nature. *Am. Nat.*

81. Pyke, G. H., H. R. Pullian, and E. L. Charnov. 1977. Optimal foraging: a selective review of theory and tests. *Q. Rev. Biol.*, 52: 137–154.

82. Ralls, K. 1976. Mammals in which females are larger than males. *Q. Rev. Biol.*, 51: 245–276.

83. Rensch, B. 1959. *Evolution above the Species Level*, New York, Columbia Univ. Press.

84. Rohwer, S. 1977. Status signaling in Harris sparrows: some experiments in deception. *Behaviour*, 61: 107–129.

85. Rosado, J. M. C., and A. Robertson. 1966. The genetic control of sex ratio. *J. Theor. Biol.*, 13: 324–329.

86. Schoener, T. W. 1968. Sizes of feeding territories among birds. *Ecology*, 49: 123–141.

87. ———. 1971. Theory of feeding strategies. *Ann. Rev. Ecol. Syst.*, 2: 369–404.

88. Sibly, R. and D. McFarland. 1976. On the fitness of behaviour sequences. *Am. Nat.*, 110: 601–617.

89. Slatkin, M. 1978. On the equilibration of fitnesses by natural selection. *Am. Nat.*, 112: 845–859.

90. ———. 1979. The evolutionary response to frequency- and density-dependent interactions. *Am. Nat.*, 114: 384–398.

91. ———. 1979. Frequency- and density-dependent selection on a quantitative character. *Genetics*, 93: 755–771.

92. Slobodkin, L. B., and A. Rapoport. 1974. An optimal strategy of evolution. *Q. Rev. Biol.*, 49: 181–200.

93. Strobeck, C. 1975. Selection in a fine-grained environment. *Am. Nat.*, 109: 419–425.

94. Tinbergen, N., G. J. Broekhuysen, F. Feekes, J. C. W. Houghton, H. Kruuk, and E. Szule. 1963. Egg shell removal by the black-headed gull, *Larus ribidundus* L.: a behaviour component of camouflage. *Behaviour*, 19: 74–117.

95. Trivers, R. L., and H. Hare. 1976. Haplodiploidy and the evolution of social insects. *Science*, 191: 249–263.

96. ———, and D. E. Willard. 1973. Natural selection of parental ability to vary the sex ratio of offspring. *Science*, 179: 90–92.

97. Verner, J. 1965. Selection for sex ratio. *Am. Nat.*, 19: 419–421.

98. ———, and M. F. Willson. 1966. The influence of habitats on mating systems of North American passerine birds. *Ecology*, 47: 143–147.

99. Von Neumann, J., and O. Morgenstern. 1953. *Theory of Games and Economic Behavior*, Princeton, N. J., Princeton Univ. Press.

100. Werner, E. E., and D. J. Hall. 1974. Optimal foraging and size selection of prey by the bluegill sunfish (*Lepomis mochrochirus*). *Ecology*, 55: 1042–1052.

101. Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Int. Congr. Genet.*, 1: 356–366.

102. Zahavi, A. 1975. Mate selection—a selection for a handicap. *J. Theor. Biol.*, 53: 205–214.

IV Women in the Evolutionary Process

7 Empathy, Polyandry, and the Myth of the Coy Female

Sarah Blaffer Hrdy

> Sexual selection theory (Bateman, 1948; Darwin, 1871; Trivers, 1972; Williams, 1966) is one of the crown jewels of the Darwinian approach basic to sociobiology. Yet so scintillating were some of the revelations offered by the theory, that they tended to outshine the rest of the wreath and to impede comprehension of the total design, in this instance, the intertwined, sometimes opposing, strategies and counter strategies of both sexes which together compose the social and reproductive behavior of the species. (Hrdy & Williams, 1983, p. 7)

But why did that happen, and how? And what processes led to the current destabilization of the model and reformulation of our thinking about sexual selection?

Introduction

For over three decades, a handful of partially true assumptions were permitted to shape the construction of general evolutionary theories about sexual selection. These theories of sexual selection presupposed the existence of a highly discriminating, sexually ''coy,'' female who was courted by sexually undiscriminating males. Assumptions underlying these stereotypes included, first, the idea that relative male contribution to offspring was small, second, that little variance exists in female reproductive success compared to the very great variance among males, and third, that fertilization was the only reason for females to mate. While appropriate in some contexts, these conditions are far from universal. Uncritical acceptance of such assumptions has greatly hampered our understanding of animal breeding systems particularly, perhaps, those of primates.

These assumptions have only begun to be revised in the last decade, as researchers began to consider the way Darwinian selection operates on females as well as males. This paper traces the shift away from the stereotype of female as sexually passive and discriminating to current models in which females are seen to play an active role in managing sexual consortships that go beyond traditional ''mate choice.'' It is impossible to understand this history without taking into account the background, including

From R. Bleier (ed.), *Feminist Approaches to Science* (New York: Pergamon Press, 1986), pp. 119–146.

the gender, of the researchers involved. Serious consideration is given to the possibility that the empathy for other females subjectively felt by women researchers may have been instrumental in expanding the scope of sexual selection theory.

Anisogamy and the Bateman Paradigm

In one of the more curious inconsistencies in modern evolutionary biology, a theoretical formulation about the basic nature of males and females has persisted for over three decades, from 1948 until recently, despite the accumulation of abundant openly available evidence contradicting it. This is the presumption basic to many contemporary versions of sexual selection theory that males are ardent and sexually undiscriminating while females are sexually restrained and reluctant to mate. My aims in this paper will be to examine this stereotype of "the coy female," to trace its route of entry into modern evolutionary thinking and to examine some of the processes that are only now, in the last decade, causing us to rethink this erroneous corollary to a body of theory (Darwin, 1871) that has otherwise been widely substantiated. In the course of this examination, I will speculate about the role that empathy and identification by researchers with same-sex individuals may have played in this strange saga.

Obviously, the initial dichotomy between actively courting, promiscuous males and passively choosing, monandrous females dates back to Victorian times. "The males are almost always the wooers," Darwin wrote in 1871, and he was very clear in his own writings that the main activity of females was to choose the single best suitor from among these wooers. As he wrote in *The Descent of Man and Selection in Relation to Sex* (1871), "It is shown by various facts, given hereafter, and by the results fairly attributable to sexual selection, that the female, though comparatively passive, generally exerts some choice and accepts one male in preference to the others." However the particular form in which these ideas were incorporated into modern and ostensibly more "empirical" versions of post-Darwinian evolutionary thought derived from a 1948 paper about animals by a distinguished plant geneticist, Angus John Bateman.

Like so much in genetics, Bateman's ideas about the workings of nature were based primarily on experiments with *Drosophila*, the minuscule flies that materialize in the vicinity of rotting fruit. Among the merits of fruitflies rarely appreciated by housekeepers are the myriad of small genetic differences that determine a fruitfly's looks. Bred over generations in a laboratory, distinctive strains of *Drosophila* sporting odd-colored eyes, various bristles, peculiar crenulations here and there, grotesquely shaped eyes, and so forth can be produced by scientists, and these markers are put to use in tracing genealogies.

Bateman obtained various lots of differently decorated *Drosophila* all belonging to the one species, *Drosophila melanogaster*. He housed three to five flies of each sex in glass containers and allowed them to breed. On the basis of 64 such experiments, he found (by counting the offspring bearing their parents' peculiar genetic trademarks) that while 21% of his males failed to fertilize any female, only 4% of his females failed to produce offspring.

A highly successful male, he found, could produce nearly three times as many offspring as the most successful female. Furthermore, the difference between the most successful and the least successful male, what is called the *variance* in male reproductive success, was always far greater than the variance among females. Building upon these findings, Bateman constructed the centerpiece to his paradigm: whereas a male could always gain by mating just one more time, and hence benefit from a nature that made him undiscriminatingly eager to mate, a female, already breeding near capacity after just one copulation, could gain little from multiple mating and should be quite uninterested in mating more than once or twice.

From these 64 experiments with *Drosophila*, Bateman extrapolated to nature at large: selection pressures brought about by competition among same-sexed individuals for representation in the gene pools of succeeding generations would almost always operate more strongly upon the male than upon the female. This asymmetry in breeding potential would lead to a nearly universal dichotomy in the sexual nature of the male and female:

> One would therefore expect to find in all but a few very primitive organisms ... that males would show greater intra-sexual selection than females. This would explain why ... there is nearly always a combination of an undiscriminating eagerness in the males and a discriminating passivity in the females. Even in a derived monogamous species (e.g. man) this sex difference might be expected to persist as a rule. (Bateman, 1948, p. 365)

This dichotomy was uncritically incorporated into modern thinking about sexual selection. In his classic 1972 essay on ''Parental Investment and Sexual Selection,'' Harvard biologist Robert Trivers acknowledged Bateman's paper as ''the key reference'' (provided him, as it happens by one of the major evolutionary biologists of our time, and Trivers' main mentor at Harvard, Ernst Mayr). Trivers' essay on parental investment, carrying with it Bateman's model, was to become the second most widely cited paper in all of sociobiology, after Hamilton's 1964 paper on kin selection.

Expanding on Bateman's original formulation, Trivers argued that whichever sex invests least in offspring will compete to mate with the sex investing most. At the root of this generalization concerning the sexually discriminating female (apart from Victorian ideology at large) is the fact of anisogamy (gametes unequal in size) and the perceived need for a female to protect her already substantial investment in each maternal gamete; she is under selective pressure to select the best available male to fertilize it. The male, by contrast, produces myriad gametes (sperm), which are assumed to be physiologically cheap to produce (note, however, that costs to males of competing for females are rarely factored in), and he disseminates them indiscriminately.

Two central themes in contemporary sociobiology then derive directly from Bateman. The first theme is the dichotomy between the "nurturing female," who invests very much more per offspring than males, and "the competitive male," who invests little or nothing beyond sperm but who actively competes for access to any additional female (see for example Daly & Margo Wilson,[1] 1983, pp. 78–79; Trivers, 1985, p. 207). As Trivers noted in his summary of Bateman's experiments with *Drosophila*, "A female's reproductive success did not increase much, if any, after the first copulation and not at all after the second; most females were uninterested in copulating more than once or twice" (1972, p. 138). And so it was that "coyness" came to be the single most commonly mentioned attribute of females in the literature on sociobiology. Unlike the male, who, if he makes a mistake can move on to another female, the female's investment was initially considered to be so great that she was constrained from aborting a bad bet and attempting to conceive again. (Criticisms and recent revisions of the notion are discussed later in the section, "The Females Who Forgot to be Coy.") In this respect, contemporary theory remains fairly faithful to Darwin's original (1871) two-part definition of sexual selection. The first part of the theory predicts competition between males for mates; the second, female choice of the best competitor.

The second main sociobiological theme to derive from Bateman is not explicitly discussed in Darwin but is certainly implicit in much that Darwin wrote (or more precisely, did not write) about females. This is the notion that female investment is already so large that it can not be increased and the idea that most females are already breeding close to capacity. If this were so, the variance in female reproductive success would be small, making one female virtually interchangeable with another. A logical corollary of this notion is the incorrect conclusion that selection operates primarily on males.

The conviction that intrasexual selection will weigh heavily upon males while scarcely affecting females was explicitly stated by Bateman, but also appears in implicit form in the writings of contemporary sociobiologists (Daly & Margo Wilson, 1983, Chapter 5; Wilson, 1978, p. 125). It is undeniable that males have the capacity to inseminate multiple females while females (except in species such as those squirrels, fish, insects, and cats, where several fathers can sire a single brood) are inseminated—at most—once each breeding period. But a difficulty arises when the occasionally true assumption that females are not competing among themselves to get fertilized is then interpreted to mean that there will be reduced within-sex competition among females generally (e.g., Freedman, 1979, p. 33).

Until about 1980—and even occasionally after that—some theoreticians were writing about females as though each one was relatively identical in both her reproductive potential and in her realization of that potential. This erroneous generalization lead some workers (perhaps especially those whose training was not in evolutionary biology per se) to the erroneous and patently non-Darwinian conclusion that females are not subject to selection pressure at all and the idea that competition among males is somehow more critical because ''leaving offspring is at stake'' (Carol Cronin, 1980, p. 302; see also Virginia Abernethy, 1978, p. 132). To make an unfortunate situation worse, the close conformity between these notions and post-Victorian popular prejudice meant that ideas about competitive, promiscuous men and choosey women were selectively picked up in popular writing about sociobiology. An article in *Playboy Magazine* celebrating ''Darwin and the Double Standard'' (Morris, 1979) comes most vividly to mind, but there were many others.

The Females Who Forgot to Be Coy

Field studies of a number of animal groups provide abundant examples of females who, unlike Bateman's *Drosophila*, ardently seek to mate more than once or twice. Furthermore, fertilization by the best male can scarcely be viewed as their universal goal since in many of these cases females were not ovulating or else were actually pregnant at the time they solicit males.

It has been known for years (among some circles) that female birds were less than chaste, especially since 1975 when Bray, Kennelly, and Guarino demonstrated that when the ''master'' of the blackbird harem was vasectomized, his females nevertheless conceived (see also Lumpkin, 1983). Evelyn Shaw and Joan Darling (1985) review some of this literature on ''promiscuous'' females, particularly for marine organisms. Among shiner perch, for example, a female who is not currently producing eggs will nevertheless court and mate with numbers of males, collecting from each male sperm that are then stored in the female's ovaries till seasonal conditions promote ovulation. Female cats, including leopards, lions, and pumas are notorious for their frequency of matings. A lioness may mate 100 times a day with multiple partners over a 6–7-day period each time she is in estrus (Eaton, 1976). Best known of all, perhaps, are such primate examples as savanna baboons, where females initiate multiple brief consortships, or chimpanzees, where females alternate between prolonged consortships with one male and communal mating with all males in the vicinity (DeVore, 1965; Hausfater, 1975; Caroline Tutin, 1975). However, only since 1979 or so has female promiscuity been a subject of much theoretical interest (see for example Alatalo, Lundberg, & Stahlbrandt, 1982; Sandy Andelman, in press; Gladstone, 1979; Sarah Blaffer Hrdy, 1979; Susan Lumpkin, 1983; Meredith Small, forthcoming; R. Smith, 1984; Wirtz, 1983), largely I believe because theoretically the phenomenon should not have existed and therefore there was little theoretical infrastructure for studying it, certainly not the sort of study that could lead to a PhD (or a job).

In terms of the order Primates, evidence has been building since the 1960s that females in a variety of prosimian, monkey, and ape species were managing their own reproductive careers so as actively to solicit and mate with a number of different males, both males within their (supposed) breeding unit and those outside it. As theoretical interest increased, so has the quality of the data.

But before turning to such evidence, it is first critical to put sex in perspective. To correct the stereotype of "coyness," I emphasize female sexual activity but, as always in such debates, reality exists in a plane distinct from that predefined by the debate. In this case, reality is hours and hours, sometimes months and months, of existence where sexual behavior is not even an issue, hours where animals are walking, feeding, resting, grooming. Among baboons (as in some human societies) months pass when a pregnant or lactating mother engages in no sexual behavior at all. The same is generally true for langurs, except that females under particular conditions possess a *capacity* to solicit and copulate with males even if pregnant or lactating, and they sometimes do so. At such times, the patterning of sexual receptivity among langurs could not be easily distinguished from that of a modern woman. The same could be said for the relatively noncyclical, semicontinuous, situation-dependent receptivity of a marmoset or tamarin.

With this qualification in mind—that is the low frequencies of sexual behavior in the lives of *all* mammals, who for the most part are doing other things—let's consider the tamarins.

Tamarins are tiny South American monkeys, long thought to be monogamous. Indeed, in captivity, tamarins do breed best when a single female is paired with one mate. Add a second female and the presence of the dominant female suppresses ovulation in the subordinate. (The consequences of adding a second male to the cage are unknown, since such an addition was thought to violate good management practices.) Nevertheless, in the recent (and first) long-term study of individually marked tamarins in the wild, Anne Wilson Goldizen discovered that given the option, supposedly monogamous saddle-backed tamarins (*Saguinus fusicollis*) will mate with several adult males, each of whom subsequently help to care for her twin offspring in an arrangement more nearly "polyandrous" than monogamous (Goldizen & Terborgh, forthcoming). Furthermore the presence of additional males, and their assistance in rearing young may be critical for offspring survival.) One of the ironies here, pointed out in another context by Janet Sayers (1982), is that females are thus presumed to commit what is known in sociobiology as a *Concorde fallacy*; that is, pouring good money after bad. Although in other contexts (e.g., Dawkins, 1976) it has been argued that creatures are selected to cut bait rather than commit Concorde fallacies, mothers were somehow excluded from this reasoning (however, see Trivers, 1985, p. 268, for a specific acknowledgement and correction of the error). I happen to believe that the resolution to this contradiction lies in recognizing that gamete producers and mothers do indeed "cut bait" far more often than is generally realized, and that skipped ovulations, spontaneous abortion, and abandonment of young by mothers are fairly routine events in nature. That is, the reasoning about the Concorde fallacy is right enough, but our thinking about the commitment of mothers to nurture no matter what has been faulty.)

Indeed, on the basis of what I believe today (cf. Hrdy, 1981, p. 59), I would argue that a polyandrous component[2] is at the core of the breeding systems of most troop-dwelling primates: females mate with many males, each of whom may contribute a little bit toward the survival of offspring. Barbary macaques provide the most extreme example (Taub, 1980), but the very well-studied savanna baboons also yield a similar, if more moderate, pattern. David Stein (1981) and Jeanne Altmann (1980) studied the complex interactions between adult males and infants. They found that (as suggested years ago by Tim Ransom and Bonnie Ransom, 1972) former, or sometimes future, consorts of the mother develop special relationships with that female's infant, carrying it in times of danger and protecting it from conspecifics, possibly creating enhanced feeding opportunities for the infant. These relationships are made possible by the mother's frequent proximity to males with whom she has special relationships and by the fact that the infant itself comes to trust these males and seek them out; more is at issue than simply male predilections. Altmann aptly refers to such males as *god-fathers*. Infants, then, are often the focal-point of elaborate male–female–infant relationships, relationships that are often initiated by the females themselves (Barbara Smuts, 1985).

Even species such as Hanuman langurs, blue monkeys, or redtail monkeys, all primates traditionally thought to have "monandrous" or "uni-male" breeding systems, are far more promiscuous than that designation implies. Indeed, mating with outsiders is so common under certain circumstances as to throw the whole notion of one-male breeding units into question (Cords, 1984; Tsingalia & Thelma Rowell, 1984). My own first glimpse of a langur, the species I was to spend nearly 10 years studying intermittently, was of a female near the Great Indian Desert in Rajasthan moving rapidly through a steep granite canyon, moving away from her natal group to approach and solicit males in an all-male band. At the time, I had no context for interpreting behavior that merely seemed strange and incomprehensible to my Harvard-trained eyes. Only in time, did I come to realize that such wandering and such seemingly "wanton" behavior were recurring events in the lives of langurs.

In at least three different sets of circumstances female langurs solicit males other than their so-called *harem-leaders*: first, when males from nomadic all-male bands temporarily join a breeding troop; second, when *females* leave their natal troops to travel temporarily with all-male bands and mate with males there; and third, when a female for reasons unknown to any one, simply takes a shine to the resident male of a neighboring troop (Hrdy 1977; Moore 1985; filmed in Hrdy, Hrdy, & Bishop, 1977). It may be to abet langurs in such projects that nature has provided them attributes characteristic of relatively few mammals. A female langur exhibits no visible sign when she is in estrus other than to present to a male and to shudder her head. When she encounters strange males, she has the capacity to shift from cyclical receptivity (that is, a bout of heat every 28 days) into a state of semicontinuous receptivity that can last for weeks. Monkeys with similar capacities include vervets, several of the guenons, and gelada baboons, to mention only a few (reviewed in Hrdy & Whitten, 1986).

A number of questions are raised by these examples. First, just exactly why might females bother to be other than coy, that is why should they actively seek out partners including males outside of their apparent breeding units (mate "promiscuously," seek "excess" copulations, beyond what are necessary for fertilization)? Second, why should this vast category of behaviors be, until recently, so generally ignored by evolutionary theorists? As John Maynard Smith noted, in the context of mobbing behavior by birds, "behavior so widespread, so constant, and so apparently dangerous calls for a functional explanation" (1984, p. 294).

To be fair, it should be acknowledged that mobbing behavior in birds is more stereotyped than sexual behavior in wild cats or monkeys, and it can be more systematically studied. Nevertheless, at issue here are behaviors exhibited by the majority of species in the order primates, the best studied order of animals in the world, and the order specifically included by Bateman in his extrapolation from coyness in arthropods to coyness in anthropoids. Furthermore, females engaged in such "promiscuous matings" entail obvious risks ranging from retaliatory attacks by males, venereal disease, the energetic costs of multiple solicitations, predation risks from leaving the troop, all the way to the risk of lost investment by a male consort who has been selected to avoid investing in other males' offspring (Trivers, 1972). In retrospect, one really does have to wonder why it was nearly 1980 before promiscuity among females attracted more than cursory theoretical interest.

Once the initial conceptual block was overcome (and I will argue in the last section that the contributions of women researchers was critical to this phase, at least in primatology), once it was recognized that oh yes, females mate promiscuously and this is a most curious and fascinating phenomenon, the question began to be vigorously pursued. (Note though that the focus of this paper is on male-centered theoretical formulations, readers should be aware that there are other issues here, such as the gap between theoreticians and fieldworkers, which I do not discuss.)

In my opinion, no conscious effort was ever made to leave out female sides to stories. The Bateman paradigm was very useful, indeed theoretically quite powerful, in explaining such phenomena as male promiscuity. But, although the theory was useful in explaining male behavior, by definition (i.e., *sexual selection* refers to competition between one sex for *access* to the other sex) it excluded much within-sex reproductive competition among females, which was not over fertilizations per se but which also did not fall neatly into the realm of the survival-related phenomena normally considered as due to natural selection. (The evolution of sexual swellings might be an example of a phenomenon that fell between definitional cracks and hence went unexplained until recently [Clutton-Brock & Harvey, 1976; Hrdy, 1981].) To understand female promiscuity, for example, we first needed to recognize the limitations of sexual selection theory and then needed to construct a new theoretical base for explaining selection pressures on females.

The realization that male–male competition and female choice explains only a small part of the evolution of breeding systems has led to much new work (e.g., Wasser, 1983, and work reviewed therein). We now have, for example, no fewer than six different models to explain how females might benefit from mating with different males (see Smith, R., 1984, for a recent review).

These hypotheses, most of them published in 1979 or later, can be divided into two categories, first those postulating genetic benefits for the offspring of sexually assertive mothers, and second, those postulating nongenetic benefits for either the female herself or her progeny. All but one of these (the oldest, "prostitution hypothesis") was arrived at by considering the world from a female's point of view.

Whereas all the hypotheses specifying genetic benefits predict that the female should be fertile when she solicits various male partners (except in those species where females have the capacity to store sperm), this condition is not required for the nongenetic hypotheses. It should be noted, too, that only functional explanations for multiple matings are listed. The idea that females simply "enjoy" sex begs the question of why females in a genus such as *Drosophila* do not appear highly motivated to mate repeatedly, while females in other species apparently are so motivated and have evolved specific physiological apparatus making promiscuity more likely (e.g., a clitoris, a capacity for orgasm brought about by prolonged or multiple sources of stimulation, a capacity to expand receptivity beyond the period of ovulation, and so forth; see Hrdy, 1981, Chapter 7 for discussion). Nevertheless, the possibility persists that promiscuous behaviors arise as endocrinological accidents or perhaps that females have orgasms simply because males do (Symons, 1979), and it is worth remembering that an act of faith is involved in assuming that there is any function at all. (I mention this qualifier because I am not interested in arguing a point that can not currently be resolved.)

Assuming that promiscuous behaviors and the physiological paraphernalia leading to them have evolved, four hypotheses are predicated on genetic benefits for the offspring of sexually assertive mothers: (a) the "fertility backup hypothesis," which assumes that females will need sperm from a number of males to assure conception (Meredith Small, forthcoming; Smith, R., 1984); (b) "the inferior cuckold hypothesis," in which a female paired with an inferior mate surreptitiously solicits genetically superior males when conception is likely (e.g., Benshoof & Thornhill, 1979); (c) "the diverse paternity" hypothesis, whereby females confronted with unpredictable fluctuations in the environment produce clutches sired by multiple partners to diversity paternity of offspring produced over a lifetime (Parker, 1970; Williams, 1975); or (d) in a somewhat obscure twist of the preceding, females in species where litters can have more than one father alter the degree of relatedness between sibs and maternal half-sibs by collecting sperm from several fathers (Davies & Boersma, 1984).

The remaining explanations are predicated on nongenetic benefits for females and do not assume the existence of either genetic differences between males or the existence of female capacities to detect them: (e) the ''prostitution'' hypothesis, whereby females are thought to exchange sexual access for resources, enhanced status, etc.—the oldest of all the explanations (first proposed by Sir Solly Zuckerman, 1932, recently restated by Symons, 1979; see also, Nancy Burley & Symanski, 1981, for discussion); (f) the ''therapeutic hypothesis'' that multiple matings and resulting orgasm are physiologically beneficial to females or make conception more likely (Mary Jane Sherfey, 1973); (g) the ''keep 'em around'' hypothesis whereby females (with the connivance of dominant males in the group) solicit subordinate males to discourage these disadvantaged animals from leaving the group (Stacey, 1982); and (h) the ''manipulation hypothesis,'' suggesting that females mate with a number of males in order to confuse information available to males about paternity and thereby extract investment in, or tolerance for, their infants from different males (Hrdy, D. B., 1979; Stacey, 1982).

It is this last hypothesis that I now want to focus on, not because that hypothesis is inherently any better than others, but because I know the most about it and about the assumptions that needed to be changed before it could be dreamed up.

The ''manipulation hypothesis,'' first conceived in relation to monkeys, grew out of a dawning awareness that, first of all, individual females could do a great deal that would affect the survival of their offspring, and second, that males, far from mere dispensers of sperm, were critical features on the landscape where infants died or survived. That is, females were more political, males more nurturing (or at least not neutral), than some earlier versions of sexual selection theory would lead us to suppose.

A Female Is Not a Female Is Not a Female

To his credit, A. J. Bateman was a very empirical scientist. He was at pains to measure ''actual'' and not just ''potential'' genetic contribution made by parents. Not for him the practice—still prevalent in primatology several decades later—of counting up some male's copulations and calling them *reproductive success*. Bateman counted offspring actually produced. And, in a genus such as *Drosophila*, where infant mortality is probably fairly random and a stretch of bad weather accounts for far more deaths than a spate of bad parenting, the assumption that one mother is equivalent to another mother is probably not farfetched. Such factors as the social status of the mother, her body size, her expertise in child-rearing, or the protection and care elicited from other animals may indeed make little difference. But what if he had been studying monkeys or even somebody's favorite fish? Even for *Drosophila* conditions exist in which females benefit from multiple copulations. In a series of experiments with *Dro-*

sophila pseudoobscura, Turner and Anderson (1983) have shown that the number of offspring that survive to maturity was significantly higher for females allowed to mate for longer periods and with more partners than for females isolated from males after brief mating periods. This effect was most pronounced in laboratory groups that were nutritionally stressed.

The female coho salmon buries her eggs in nests, which she guards for as long as she lives. Females fight over the best nest sites, and about one out of three times, a female will usurp another female's nest and destroy her eggs. Females vary greatly in size, and their differing dimensions may be translated into different degrees of fecundity. A big female may produce more than three times as many eggs as a small one. Differences in the survival of eggs to hatching lead to even greater variance in female reproductive success; there may be as much as a 30-fold difference in number of surviving offspring (Van den Berghe, 1984).

But the mother salmon only breed once; consider an iteroparous monkey mother who, although she produces only one or two infants at a time, breeds over many years and who, like a macaque or baboon, may inherit her feeding range and troop rank from her mother at birth. These legacies will affect her reproductive output and will, in turn, pass to her own daughters. Males of course enter this system, and vary among themselves, but in most instances they are transients, breeding briefly, and indeed, possibly living shorter lives on average than females. Take the extreme example of the gelada baboon who has only one chance for controlling access to a small ''harem'' of females (who by the way have about as much to do with controlling the male, as he does in controlling them). The male gelada baboon breeds in his unit for several years before another male enters, pushing him into forced retirement. The former ''harem-leader'' lingers on in the troop, but as a celibate watcher, possibly babysitting, but breeding no more (Dunbar, 1984). It is a tale of the tortoise and the hare. After the male hare is dismissed, the female tortoise breeds on year after year.

Although we do not yet have data on the lifetime reproductive success of males or females from any species of wild primate, I will be surprised if the variance among males exceeds the variance among females by as much as traditionally thought in species such as Japanese or rhesus macaques or gelada baboons. In the most polyandrous species, such as tamarins, variance in the reproductive success of twin-producing females may actually be greater than that for males. If we carry out our calculations over generations, remembering that every male, however wildly reproductively successful, has a mother and a grandmother (e.g., see Hartung, in press) differences in the degree of variance between the two sexes grow even smaller, though extremes of variance in reproductive success will of course crop up one generation sooner for fathers than for mothers.

The anisogamy paradigm of Bateman offered powerful insights into the selective pressures that operate on males; for many mammals, selection weighs heaviest on males in competition with other males for access to females. In addition, Bateman provided the framework that eventually led to an understanding of why males tend to compete for mates while females compete for resources. But the Bateman and the anisogamy paradigm also led us to overlook the full range of possible sources of variance in female reproductive success; not only variance arising from female–female competition over resources to translate into large gametes, but also variance arising from other factors as well. Not all females conceive. In some cases, such as marmosets, the presence of the dominant female suppresses ovulation in her subordinates. Some offspring, once conceived, are not carried to term. Among the factors leading to spontaneous abortion in baboons may be harassment by other females or the arrival of strange males (Mori & Dunbar, in press; Wasser & Barash, 1984). And of course, offspring once born need not survive. If born to a low-ranking toque macaque mother, a juvenile daughter may die of starvation, or if born to a mother chimp who for some reason is incapacitated, an offspring may be killed by a higher-ranking female. Having survived, a maturing female howler monkey may nevertheless find herself unable to join a breeding group and never have a chance to reproduce. A mother's condition, her competitive abilities, and her maternal skills are all very much at issue in the case of creatures such as primates. Yet, as amazing as it sounds, only relatively recently have primatologists begun to examine behaviors other than direct mother–infant interactions that affect the fates of infants (for elaboration see Hrdy, 1981; Small, 1984). Not the least among the variables affecting their survival is the role played by males, and the capacity of females to influence this male performance.

Male Involvement with Infants

Even for *Drosophila* it was a mistake to imagine that male investment never went further than chromosomes. Recent research makes it clear that, as in various butterflies and cockroaches, male fruitflies may sometimes transmit along with their sperm essential nutrients that otherwise would be in short supply (Markow & Ankney, 1984). When assumptions about minimal male involvement are extrapolated to species such as primates, however, far more than underestimation of male involvement is at stake. I would argue that it is not only ill-advised but impossible to understand primate breeding systems without taking into account the role of males in determining the survival or demise of infants.

There is probably no order of mammals in which male involvement with infants is more varied, more complex, or more crucial than among primates. About 10% of all mammalian genera exhibit some form of direct male care, that is the male carries the infant or provisions it. Among primates, however, the percentage of genera with direct, positive (if also sometimes infrequent) interactions between males and infants is roughly four times that, the highest figure reported for any order of mammals (Devra

Kleiman & Malcolm, 1981; Vogt, 1984). Conversely, infanticide has been reported for over 15 different species of primates belonging to 8 genera and is probably widespread among apes and monkeys (Hausfater & Hrdy, 1984). Indeed, some male care is probably a direct outgrowth of the need by males to protect infants from other males (Busse & Hamilton, 1981). Yet, oddly, after two decades of intensive study of wild primates, we are only now beginning to scratch the surface of the rich interactions that exist between infants and adult males, which seem to have such critical repercussions for infant survival (see Hrdy, 1976; and especially, Taub, 1984a, 1984b). Effects of these relationships for infants after they grow up have rarely been investigated, although several researchers have recently suggested the possibility that fathers among gibbons and orangutans may play a role in helping their sons to set up or defend territories (MacKinnon, 1978; Tilson, 1981). These cases are of special importance because apart from intervention by brothers or by fathers in adopting an orphan (among gorillas and chimpanzees) direct, ''maternal-like'' care of infants by males is not typically seen among apes. But, the fact that parental investment by males does not take the same form as investment by females does not lessen its importance for offspring or its cost to the parent. My focus here is on primates, but I believe I could make many of the same points if I were a student of amphibians or fish in which male care is very common. One critical role of males is to protect immatures from distantly related conspecifics. It has long been assumed that one reason for male care among these species was the greater certainty of paternity permitted in species with external fertilization (i.e., the male can *know* which eggs he fertilized). But surely among these groups, as among primates, there has been selection on females to manipulate this situation.

The main exception to a general pattern of ignoring interactions between males and infants was of course the study of male care among monogamous primates. It has been known for over 200 years, ever since a zoologist-illustrator named George Edwards decided to watch the behavior of pet marmosets in a London garden, that among certain species of New World monkeys males contributed direct care for infants that equalled or exceeded that given by females (Edwards, 1758). Mothers among marmosets and tamarins typically give birth to twins, as often as twice a year, and to ease the female in her staggering reproductive burden the male carries the infant at all times except when the mother is actually suckling it. It was assumed that monogamy and male confidence of paternity was essential for the evolution of such care (Kleiman, 1977), and at the same time, it was assumed that monogamy among primates must be fairly rare (e.g., see Symons, 1979, or virtually any textbook on physical anthropology prior to 1981).

Recent findings, however, make it necessary to revise this picture. First of all, monogamy among primates turns out to be rather more frequent than previously believed (either obligate or facultative monogamy can be documented for some 17–20% of extant primates) and, second, male care turns out to be far more extensive than previously thought and not necessarily confined to monogamous species (Hrdy, 1981). Whereas, previously, it was assumed that monogamy and male certainty of paternity facilitated the evolution of male care, it now seems appropriate to consider the alternative possibility, whether the extraordinary capacity of male primates to look out for the fates of infants did not in some way pre-adapt members of this order for the sort of close, long-term relationships between males and females that, under some ecological circumstances, leads to monogamy! Either scenario could be true. The point is that on the basis of present knowledge there is no reason to view male care as a restricted or specialized phenomenon. In sum, though it remains true that mothers among virtually all primates devote more time and/or energy to rearing infants than do males, males nonetheless play a more varied and critical role in infant survival than is generally realized.

Male–infant interactions are weakly developed among prosimians, and in these primitive primates, male care more or less (but not completely) coincides with monogamy (Vogt, 1984). Direct male care occurs in 7 out of 17 genera, including one of the most primitive of all lemurs, the nest-building ruffed lemur (*Lemur variegatus*), where the male diligently tends the nest while the mother forages (personal communication from Patricia Wright). Among New World monkeys, 12 of 16 genera (Vogt, 1984) or, calculated differently, 50% of all species (Wright, 1984) exhibit direct male care, often with the male as the primary caretaker. That is, shortly after birth, an adult male—often with the help of various immatures in the group or other males—will take the infant, carry it (or them, in the frequent case of twins) on his back, share food with infants, either adult males or juveniles may catch beetles to feed them, or assist them by cracking the casing of tough fruit.

The role of males as primary caretakers for single (nontwin) infants is very richly developed among the night monkeys, *Aotus trivirgatus*. These small, monogamously mated, South American monkeys are the only nocturnal higher primate. Because of the difficulty in watching them, their behavior in the wild has gone virtually undocumented until detailed behavioral studies were undertaken by Patricia Wright using an image intensifier and other gear to allow her to work at night. Combining her observations of captive *Aotus* with field observations, a picture emerges in which the male is primary caretaker (in terms of carrying the infant) from the infant's first day of life, although the mother, of course, still is providing physiologically very costly milk. Based on captive observations, the mother carried the infant 33% of the time during the first week of life, the male 51%, and a juvenile group member 15%. In the wild, the infant was still being carried by the male at 4 months of age, although "weaning" tantrums were seen, as the male would try to push the infant off his back. By 5 months, the infant was relatively independent of either parent (Patricia Wright, 1984).

There is little question that there is an association between monogamy and extensive male care. Nevertheless, this does not mean that the evolution of male care is precluded by situations in which females mate with more than one male, as discussed for the case of savannah baboons.

Recent research on male–infant relations among baboons reveals that during their first week of life infant baboons at Amboseli spend about a third of their daylight hours within 5 feet of an adult male, often, but not always, a former sexual consort of the mother. This level of proximity was maintained throughout the first 7 weeks and then dropped sharply. At the same time, the amount of time infants spend in actual contact with an adult male, which is never much, is rising from 1% in the first week to 3% by the eighth week. During their first half-year of life, infants spend .5% of their time connected with an adult male, a low figure (Stein, 1984). Averaging together data from a number of different baboon field studies, David Taub calculates that a male–infant interaction takes place only about once every 19 hours (or, adjusting for the number of males in a multi-male troop, one interaction per male every 344 hours). However, Taub concurs with Busse and Hamilton (1981) and others, that the proximity of these males may be crucial for infant survival, particularly critical for discouraging attacks on the infant either by incoming males, unfamiliar with the infant's mother or, as suggested by Wasser (1983) for forestalling harassment by female troop members belonging to competing matrilines. That is, when the cost of care is fairly low (the male need only remain in the vicinity of the infant but can engage in other activities) and when it is rendered nonexclusively to several infants (e.g., to the offspring of each of the male's special female friends), male care certainly does occur in nonmonogamous systems. What is offered may not be ''quantity'' time, but it may well be ''quality'' time—''quality'' in a very real sense: enhancing infant survival.

Yet, even these caveats can be dispensed with in the unusual case of the polyandrous tamarin species (*Saguinus fuscicollis*) studied by Goldizen. The female mates with several males and each of them subsequently helps rear the infant. Indeed, preliminary data from Goldizen's continuing research suggests that infants with several male caretakers are more likely to survive than infants born in small groups with only one adult male. Here, then, is both quality and quantity time, combined in a nonmonogamous breeding system, a system where males have a probability but no certainty of paternity. If we pause for a moment and consider the tamarin case from the male's point of view, the system Goldizen reports almost certainly derived initially from a monogamous one in which males were indeed caring for offspring likely to be their own. Only after such a system was established could a female have plausibly manipulated the situation to enlist the aid of two helpers.

Assuming that primate males do indeed remember the identity of past consorts and that they respond differentially to the offspring of familiar and unfamiliar females, females would derive obvious benefits from mating with more than one male. A researcher with this model in mind has quite different expectations about female behavior than one expecting females to save themselves in order to mate with the best available male. The resulting research questions will be very different.

The Role of Women Researchers

When generalizations persist for decades after evidence invalidating them is also known, can there be much doubt that some bias was involved? We were predisposed to imagine males as ardent, females as coy; males as polygynists, females monandrous. How else could the *Drosophila* to primate extrapolation have entered modern evolutionary thinking unchallenged?

Assuming, then, this bias, a preconstituted reality in which males played central roles, what factors motivated researchers to revise invalid assumptions? What changes in the last decade brought about the new focus on female reproductive strategies and, with it, the recognition that certain assumptions and corollaries of the Bateman paradigm, and especially female monandry, were seriously limited and even, if applied universally, quite wrong.

The fact that there is relatively less intrasexual selection for mates among females does not mean reduced intrasexual competition or reduced selection among females in other spheres of activity. To understand male–male competition for mates is to understand only a small part of what leads to the evolution of particular primate breeding systems. We need also consider the many sources of variance in female reproductive success, including a whole range of female behaviors not directly related to "mothering" that may have repercussions on the fates of their infants.

Polyandrous mating with multiple males, mating with males when conception is not possible—what from the males' point of view might be termed "excessive" matings—can only be understood within this new framework, but it requires a whole new set of assumptions and research questions. As a result, sexual selection theory is currently in a state of flux; it is being rethought as actively as any area in evolutionary biology. What processes contributed to this destabilization of a long-held paradigm? And in particular, what led us to rethink the myth of the coy or monandrous female?

Improved methodologies and longer studies would not by themselves have led us to revise the myth of the coy female, simply because the relevant information about "female promiscuity" was already in hand long before researchers began to ask why females might be mating with more than one male. Indeed, at least one writer, working in a framework well outside of primatology and evolutionary biology, picked up on the reports of female promiscuity in baboons and chimpanzees at an early date (1966) and asked why it had evolved. This of course was the feminist psychiatrist Mary Jane Sherfey in her book, *The Evolution of Female Sexuality* (1973). Sherfey's vision of the "sexually insatiable" female primate was generally ignored by primatologists and biologists both because of her ideological perspective and because her standards of evidence were far from scientific. If her ideas were mentioned, it was typically with sarcasm and derision (Symons, 1979, pp. 76–77, 94, 262, 311). And, yet, it is important to note that however extreme her views (and scholarly balance was not Sherfey's strong point), they provided a valuable antidote to equally extreme ideas about universally coy females that were widely held by scientists within the academic mainstream of evolutionary biology. Elsewhere, I wrote about the various factors which caused us to recognize the importance of female dominance hierarchies in the lives of cercopithecine monkeys (Hrdy, 1984). Changes in methodology (e.g., focal animal sampling of all individuals in a group) and the emergence of long-term studies played critical roles in revising male-centered models of primate social organization. In that case as well, some of the relevant information was available long before we decided it was significant (e.g., the detailed Japanese studies indicating matrilineal inheritance of rank, Kawai, 1958; Kawamura, 1958). But, in the "coy female" case, I don't think that the duration of the studies or the field methods made as much difference as the particular research questions being asked. Ultimately, however, long-term studies are going to be very important for testing the various hypotheses to explain why females mate with multiple males.

New or better data alone did not change the framework in which we asked questions; rather, I believe, something motivational changed. Among the factors leading to a reevaluation of the myth of the coy female, the role of women researchers must be considered. That is, I seriously question whether it could have been just chance or just historical sequence that caused a small group of primatologists in the 1960s, who happened to be mostly male, to focus on male–male competition and on the number of matings males obtained, while a subsequent group of researchers, including many women (beginning in the 1970s), started to shift the focus to female behaviors having long-term consequences for the fates of infants (reviewed in Hrdy & Williams, 1983).

In this paper, I deliberately included first names whenever the work of a woman was cited. I did this to emphasize just how many women are currently working specifically in this area. Even a casual inspection reveals that women are disproportionately represented among primatologists compared to their representation in science generally. For example, in 1984, just over a third of the members (36%) of the American Society of Primatologists were women.[3] As we reconstruct the journey from Bateman (1948) to the recognition that the adjectives *coy* and *female* are something less than synonymous, it seems clear that the insights of women are implicated at every stage along the way and that their involvement exceeds their representation in the field. Having said this, I need to remind readers that as history my account here is biased by a conscious focus on contributions by women. A broader treatment would also have to describe the pioneering research on long-term male–female relations by T. M. Ransom and Robert Seyfarth and the extensive studies of male–infant relations by Mason, Mitchell, Redican, Stein, Taub, and others (see Taub, 1984a, 1984b, for reviews). I am acutely aware that my treatment here is biased both by my particular purpose (discussing the role of empathy by females for other females in causing us to revise old assumptions) and by my own involvement in the transition of primatology from the study of primate "behavior" to the study of primate "sociobiology." Hence, I leave to someone else the task of writing a balanced history of primatology in this period (e.g., see Alison Jolly, 1985).

The contributions of women researchers can be interpreted in several ways. Perhaps, women are simply better observers. As Louis Leakey used to say in an effort to justify his all-too-evident preference for women researchers, "You can send a man and a woman to church, but it is the woman who will be able to tell you what everyone had on" (personal communication, 1970). Or perhaps women are by temperament more pragmatic or more empirical, less open to theoretical bias. A difficulty with both ideas, of course, is that a few women were present in primatology in the 1960s, and both sexes participated in perpetuating myths about monkeys living in male-centered societies, where the primary activities of females had to do with mothering (e.g., Jane Goodall, 1971 or Phyllis Jay, 1963; but see Jane Lancaster, 1975; and Thelma Rowell, 1972, for exceptions). Women seemed just as vulnerable to bias as men.

If the presence of women was a constant but our ideas changed, perhaps, as Donna Haraway (1976) likes to remind us, the interpretations of primatologists simply mirror ideological phases in the history of the Western world. Indeed, it is disconcerting to note that primatologists are beginning to find politically motivated females and nurturing males at roughly the same time that a woman runs for vice president of the United States and Garry Trudeau starts to poke fun at "caring males" in his cartoons.

Or, perhaps, as Thelma Rowell (1984) suggested it was easier "for females to empathize with females, and ... empathy is a covertly accepted aspect of primate studies" (p. 16). Perhaps, the insights were there all along but it took longer to challenge and correct male-centered paradigms because the perceptions of women fieldworkers lacked the authority of male theorists.

In *A Feeling for the Organism*, Evelyn Fox Keller (1983) hints at the possibility that women biologists may have some special sensibility concerning the creatures that they study, an ability to enter into the lives of their subjects—a suggestion that maize geneticist Barbara McClintock, the subject of her biography, would surely deny. Among other things, such a singular "gift" for women might be thought to confine women to particular areas of science or to diminish their accomplishments. That is, as primatologist Linda Fedigan wrote recently,

> I do admit to some misgivings about the wider implications of female empathy. Rowell may be correct about our sense of identification with other female primates, but I well remember my dismay when, having put many hours of effort into learning to identify the individual female monkeys of a large group, my ability was dismissed as being inherent in my sex by a respected and senior male colleague. (p. 308)

To put Fedigan's concern in perspective one needs to realize that in conversations with primatologists and, indeed, among ethologists generally, it is fairly commonplace to hear it said that women seem better able than men to learn to individually identify large numbers of animals. In a now legendary study, the seemingly incredible capacity of British ornithologist Dafila Scott to identify and remember hundreds of unmarked swans was tested by a male colleague. Indeed, it is occasionally suggested that the difficulty men have learning individuals is one reason why more men go into the ecological side of primatology.

Similarly, and I believe justifiably, women primatologists have worried about identifying too closely with the study of mothers and infants for fear that this area would become the ''home economics'' of primatology, a devalued women's domain within the discipline, or for fear that it would exacerbate the already common view that women study monkeys because it satisfies a deep-felt need to be around cuddly creatures.

Yet, suppose that there is some truth to the idea that women identified with same-sex subjects and allowed this identification to influence research focus? After all, isn't this what male primatologists, and many other ethologists as well, were doing throughout the 1960s and, occasionally, into the 1980s?

Even today, one can encounter lovely examples of what I call the *punch line phenomenon*, when a covert identification by researchers with same-sex individuals suddenly becomes overt in a last paragraph or emphatic comment. For example, in a seemingly impartial 1982 paper entitled, ''Why Do Pied Flycatcher Females Mate with Already Mated Males?,'' the authors present data to show that females who mate with already mated males rear fewer offspring than female flycatchers who are the sole mates of males, regardless of the kind of territories they had to offer her. Surely, this modern, post-''coy female'' paper, focused as it is upon the reproductive success of females, a paper essentially about female strategies, will not succumb to a male-centered perspective. Yet by the end of the paper, by some imperceptible process, the female has become object, the male protagonist: ''Our conclusion is that polygamous pied flycatcher males deceive their secondary females'' (p. 591) and the strategy works, according to the authors, because the females lack the time to check out whether the male already has a mate whose offspring he will invest in: ''it pays for a pied flycatcher female to be fast rather than coy, and therefore *she* [italics mine] can be deceived. . . .''

My own work, before I began consciously to consider such matters, provides another example. The last line of *The Langurs of Abu: Female and Male Strategies of Reproduction* (1977), a book in which I scrupulously devoted equal space to both sexes, reads, ''For generations, langur females have possessed the means to control their own destinies: caught in an evolutionary trap they have never been able to use them'' (p. 309). I might as well have said *we*.

On a conversational level, few primatologists bother to deny this phenomenon. As a colleague remarked recently when the subject came up, "Of course I identify with them. I sometimes identify with female baboons more than I do with males of my own species." But why, we still need to ask, was the process of same-sex identification by women different in the 1970s and 1980s than in the early years of primatology?

I leave the general answers to such questions to social historians, who are more qualified than I to deal with them. At this point in the chapter, I abandon scholarship and attempt briefly to trace my own experiences as I remember them, particularly as they relate to the recognition of the active roles females were playing in the evolution of primate breeding systems.

Reminiscence

In 1970, as a first-year graduate student at Harvard, I began research on infanticidal behavior by males and ended, a decade later, almost entirely focused on the reproductive strategies of females. What processes were involved? Some months after starting my fieldwork in Rajasthan, India, I abandoned my original hypothesis (that infanticide was a response to crowding) and adopted an interpretation based on classical sexual selection theory: infanticide was an outcome of male–male competition for access to females. That is, males only killed infants when they (the males) invaded breeding units from outside; mothers whose infants were killed subsequently mated with the killer sooner than if the mothers had continued to lactate (Hrdy, 1974). By killing infants sired by other males, the usurpers increased their own opportunities to mate with fertile females.

The story was straightforward enough and in line with everything I had been taught at Harvard. But, there were loose ends, not the least of which was my growing emotional involvement with the plight of female langurs. Every 27 months, on average, some male was liable to show up and attempt to kill a female's infant, and increasingly, my identification was with the female victimized in this way, not with the male who, according to the sexual selection hypothesis, was thereby increasing his reproductive success. If infanticide really was an inherited male trait that could be elicited by particular conditions (as I believed was the case), why would females put up with this system? Why not refuse to breed with an infanticidal male and wait until a male without any genetic propensity for infanticide showed up? Consideration of this question led to many others related to the question of intrasexual competition among females generally (Hrdy, 1981).

First came an unconscious process of identification with the problems a female langur confronts followed by the formulation of conscious questions about how a female copes with them. This, in turn, led to the desire to collect data relevant to those questions. Once asked, the new questions and new observations forced reassessment of old assumptions and led to still more questions. Even events I had seen many times before (e.g., females leaving their troops to solicit extratroop males) raised questions as they never had before.

If it was really true that females did not benefit from additional matings, why were female langurs taking such risks to solicit males outside their troop? Why would already pregnant females solicit and mate with males? What influence might such behavior have for the eventual fate of the female's offspring? What were the main sources of variance in female reproductive success and what role did nonreproductive sexuality play in all this? Why is situation-dependent receptivity, as opposed to strictly defined cyclical receptivity or estrus, so richly developed in the order primates? Where did the idea of the coy female ever come from anyway? These are the questions that preoccupied me since 1977 and all of them grow out of an ability to imagine females as active strategists.

Yet, identification with same-sex individuals in another primate species may not be quite so simple as it sounds. This history of primatology suggests that the nature of this identification was changing over time as the self-image of women researchers also changed. In my own case, changes in the way I looked at female langurs were linked to a dawning awareness of male–female power relationships in my own life, though ''dawning'' perhaps overstates the case.

It would be difficult to explain to an audience of political activists how intelligent human beings could be as politically unaware as many field biologists and primatologists are. Almost by definition, we are people who lead isolated lives and, by and large, avoid joining groups or movements. In addition, I was the sole woman in my cohort, since I was the first woman graduate student my particular advisor had taken on and only toward the end of the 1970s did I begin to read anything by feminist scholars like Carolyn Heilbrun and Jean Baker Miller. Each step in understanding what, for example, might be meant by a term like *androcentric* was embarked upon very slowly and dimly, sometimes resentfully, as some savage on the fringe of civilization might awkwardly rediscover the wheel. When I did encounter feminist writings, I was often put off by the poor quality of the scholarship. Sherfey's book is a case in point: highly original insights were imbedded in what seemed to me a confused and often erroneous matrix. Nevertheless, the notion of ''solidarity'' with other women and, indeed, the possibility that female primates generally might confront shared problems was beginning to stir and to raise explicit questions about male–female relations in the animals I studied. That is, there were two (possibly more) interconnected processes: an identification with other females among monkeys taking place at roughly the same time as a change in my definition of women and my ability to identify and articulate the problems women confront.

Such an admission raises special problems for primatologists. My discipline has the choice of either dismissing me as a particularly subjective member of the tribe or else acknowledging that the tribe has some problems with objectivity. It is almost a cliché to mention now how male-biased the early animal behavior studies were (see Wasser, 1983). But, in the course of the last decade of revision, are we simply substituting a new set of biases for the old ones?

The feminist charge that most fields, including psychology, biology, and animal behavior, have been male-centered, is, I think, by now undeniable. Yet to me, the noteworthy and encouraging thing is how little resistance researchers in my own field have exhibited when biases are pointed out. Although I still sense in Britain a reluctance to admit that male bias was ever actually a problem, among primatologists in the United States it is now widely acknowledged, and this has to be a healthy sign. Indeed, in animal behavior and primatology, there has been something more like a small stampede by members of both sexes to study female reproductive strategies, as well as perhaps a rush to substitute a new set of biases for the old. (That is, among feminist scholars it is now permissible to say that males and females are different, provided one also stipulates that females are more cooperative, more nurturing, more supportive—not to mention equipped with unique moral sensibilities; among sociobiologists *kudos* accrue to the author of the most Machiavellian scenario conceivable.)

There are of course antidotes to the all-too-human element that plagues our efforts to study the natural world. Common sense in methodology is one. No one will ever again be permitted to make pronouncements about primate breeding systems after having studied only one sex or after watching only the conspicuous animals. A recognition of the sources of bias is another. If, for example, we suspect that identification with same-sex individuals goes on or that certain researchers identify with the dominant and others with the oppressed and so forth, we would do well to encourage multiple studies, restudies, and challenges to current theories by a broad array of observers. We would also do well to distinguish explicitly between what we know and what we know is only interpretation. But really (being generous) this is science as currently practiced: inefficient, biased, frustrating, replete with false starts and red herrings, but nevertheless responsive to criticism and self-correcting, and hence better than any of the other more unabashedly ideological programs currently being advocated.[4]

Acknowledgments

In the preface to her recent book *Mother Care: Other Care*, my colleague in behavioral biology Sandra Scarr (1984, p. xi) notes, "I wish I could thank all the wonderful graduate school professors who helped me to realize the joys of combining profession and motherhood; unfortunately there weren't any at Harvard in the early 1960s." A decade later, things at Harvard—at least in the biologically oriented part of Harvard that I encountered—had changed remarkably little. As I think back on those postgraduate years (my undergraduate experience at Harvard was a wonderful and very different story), I can not recall a single moment's fear of success, but what I do distinctly recall was the painful perception that there were professors and fellow students (no women in those years) who acted as if *they* feared that I might succeed. Intellectually, it was a tremendously exciting environment, filled with stimulating and occasionally inspirational teachers and coworkers. It was also an environment that was socially and psychologically hostile to the professional aspirations of women. But there were exceptions, exceptions made all the more significant because they were rare. In particular, I remain deeply grateful to Ed Wilson for encouragement and for unfailing behind-the-scenes support (probably the best kind) offered not only to me, but to other women ethologists both younger and older than myself.

Writing now from an ecological niche so benign as to cause me to wonder if perhaps my zeal as a feminist won't now lapse as a consequence, it is a pleasure to acknowledge discussions with Leo Berenstain, Daniel Hrdy, Jon Marks, David Olmsted, Peter Rodman, and Judy Stamps, who read and gave me detailed criticisms of this paper. I also thank Jeanne Altmann, Alison Jolly, Jane Lancaster, Linda Partridge, Joannie Silk, Meredith Small, Barbara Smuts, and the Pats, Whitten and Wright, for valuable discussion, and thank Ruth Bleier and Erik Eckholm for inciting me to think along the lines that I do at the end of this paper. Not least, I thank Nancy McLaughlin for her assistance in preparing the manuscript.

Notes

1. In this chapter, I designate women researchers by spelling out their first names; the point of using this admittedly odd convention will become clear in the section on "The Role of Women Researchers."

2. For want of a better term, *polyandrous* is used here to refer to a female with more than one established mate. The term *promiscuous* will be used to refer to multiple, brief consortships, some of which may last longer. Whereas *polyandrous* is a poor term because it suggests some stable, institutionalized relationship, which is probably wrong for describing tamarins, *promiscuous* is also problematic. It implies a lack of selectivity among females, which may or may not be the case. Davies and Lundberg (1984, p. 898) have recently proposed using the term *polygnandry* to refer to "two or three males sharing access to two, three or four females." Such a term applies to Barbary macaques and might be a good one for the baboon situation except that there is not a 100% overlap in the females with which each male mates. Clearly, the terminology needs to be cleared up, but for the time being the important point is to emphasize the contrast between what we now know and the old stereotype of monandrous females selecting a single mate.

3. It should be noted however that membership in the ASP signals the *motivation* of women to join, since all one has to do is sign up and pay dues. Recognition and acceptance may be quite different. Contrast, for example, the position of women on editorial boards (4 of 40 on the *International Journal of Primatology* are women; 0 of 19 on the editorial board of the journal *Behavioral*

Ecology and Sociobiology). When we examine the prestigious roster of *elected* fellows of the Animal Behavior Society for 1985, 1 of 62 is a woman. All 19 autobiographical chapters in *Leaders in the Study of Animal Behavior* are by men.

4. Recent feminist programs advocating "conscious partiality" come to mind. If an unbiased knowledge is impossible, this argument runs, an explicitly biased, politically motivated approach is preferable to the illusion of impartial research.

References

Abernethy, V. (1978). Female hierarchy: An evolutionary perspective. In L. Tiger & H. Fowler (Eds.), *Female Hierarchies*. Chicago: Beresford Book Service.

Alatalo, R. V., Lundberg, A., & Stahlbrandt, K. (1982). Why do pied flycatcher females mate with already mated males. *Animal Behaviour, 30*, 585–593.

Altmann, J. (1980). *Baboon mothers and infants*. Cambridge: Harvard University Press.

Altmann, S. (Ed.). (1965). *Japanese monkeys: A collection of translations*. Edmonton, Canada: The editor.

Andelman, S. (forthcoming). Concealed ovulation and prolonged receptivity in vervet monkeys (*Cercopithecus aethiops*).

Bateman, A. J. (1948). Intra-sexual selection in drosophila. *Heredity, 2*, 349–368.

Benshoof, L., & Thornhill, R. (1979). The evolution of monogamy and concealed ovulation in humans. *Journal of Biological Structures, 2*, 95–106.

Bleier, R. (1984). *Science and gender*. Elmsford, NY: Pergamon.

Bray, O. E., Kennelly, J. J., & Guarino, J. L. (1975). Fertility of eggs produced on territories of vasectomized red-winged blackbirds. *Wilson Bulletin, 87*, no. 2, 187–195.

Burley, N., & Symanski, R. (1981). Women without: An evolutionary perspective on prostitution. In *The immoral landscape: Female prostitution in Western societies*. Toronto: Butterworth.

Busse, C., & Hamilton, W. J., III. (1981). Infant carrying by male chacma baboons. *Science, 212*, 1281–1283.

Clutton-Brock, T. H., & Harvey, P. (1976). Evolutionary rules and primate societies. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology*. Cambridge: Cambridge University Press.

Cords, M. (1984). Mating patterns and social structure in redtail monkeys (*Cercopithecus ascanius*). *Zeitschrift für Tierpsychologie, 64*, 313–329.

Cronin, C. (1980). Dominance relations and females. In D. R. Omark, F. F. Strayer, and D. G. Freeman (Eds.), *Dominance relations*. New York: Garland Press.

Daly, M., & Wilson, M. (1983). *Sex, evolution and behavior*. Boston: Willard Grant Press.

Darwin, C. (1871). *The descent of man and selection in relation to sex* (1887 edition). New York: D. Appleton and Co.

Davies, E. M., & Boersma, P. D. (1984). Why lionesses copulate with more than one male. *The American Naturalist, 123*, no. 5, 594–611.

Davies, N. B., & Lundberg, A. (1984). Food distribution and a variable mating system in the dunnock, *Prunella modularis*. *Journal of Animal Ecology, 53*, 895–912.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

DeVore, I. (Ed.). (1965). *Primate behavior*. New York: Holt, Rinehart and Winston.

Diamond, J. (1984). Theory and practice of extramarital sex. *Nature, 312*, 196.

Dunbar, R. (1984). *Reproductive decisions: An economic analysis of gelada baboon social strategies*. Princeton, NJ: Princeton University Press.

Eaton, R. (Ed.). (1976). *The world's cats II*. Seattle, WA: Feline Research Group, Woodland Park Zoo.

Edwards, G. (1758). *Gleanings of Natural History* (Vol. 5). London: College of Physicians.

Fedigan, L. (1984). Sex ratios and sex differences in primatology (book review of *Female primates*). *American Journal of Primatology, 7*, 305–308.

Freedman, D. (1979). *Human sociobiology: A holistic approach*. New York: The Free Press.

Fujioka, M., & Tamagishi, S. (1981). Extramarital and pair copulations in the cattle egret. *Auk, 98*, 134–144.

Gladstone, D. (1979). Promiscuity in monogamous colonial birds. *The American Naturalist, 114*, no. 4, 545–557.

Goldizen, A. W., & Terborgh, J. (in press). Cooperative polyandry and helping behavior in saddle-backed tamarins (*Saguinus fuscicollis*). Proceedings of the IXth Congress of the International Primatological Society. Cambridge: Cambridge University Press.

Goodall, J. (1971). *In the shadow of man*. Boston: Houghton Mifflin.

Haraway, D. (1976). The contest for primate nature: Daughters of man-the-hunter in the field. In M. Kann (Ed.), *The future of American democracy: Views from the left*. Philadelphia, PA: Temple University Press.

Hartung, J. (in press). Matrilineal inheritance: New theory and analysis. *The Behavioral and Brain Sciences*.

Hausfater, G. (1975). Dominance and reproduction in baboons (*Papio cynocephalus*). *Contributions to Primatology* (Vol. 7). Basel, Switzerland: S. Karger.

Hausfater, G., & Hrdy, S. B. (Eds.). (1984). *Infanticide: Comparative and evolutionary perspectives*. New York: Aldine.

Hrdy, D. B. (1979). Integrated field study of the behavior, genetics and diseases of the Hanuman langur in Rajasthan, India. Proposal submitted to the National Science Foundation.

Hrdy, S. B. (1974). Male–male competition and infanticide among the langurs (*Presbytis entellus*) of Abu, Rajasthan. *Folia Primatologica, 22*, 19–58.

Hrdy, S. B. (1976). The care and exploitation of nonhuman primates by conspecifics other than the mother. *Advances in the Study of Behavior, VI*, 101–158.

Hrdy, S. B. (1977). *The langurs of Abu: Female and male strategies of reproduction*. Cambridge: Harvard University Press.

Hrdy, S. B. (1979). Infanticide among animals: A review, classification and examination of the implications for the reproductive strategies of females. *Ethology and Sociobiology, 1*, 3–40.

Hrdy, S. B. (1981). *The woman that never evolved*. Cambridge: Harvard University Press.

Hrdy, S. B. (1984). Introduction: Female reproductive strategies. In M. Small, (Ed.), *Female primates: Studies by women primatologists*. New York: Alan Liss.

Hrdy, S. B., Hrdy, D. B., & Bishop, J. (1977). *Stolen copulations*. 16 mm color film. Peabody Museum.

Hrdy, S. B., & Whitten, P. (1986). The patterning of sexual activity. In D. Cheney, R. Seyfarth, B. Smuts, R. Wrangham, & T. Struhsaker (Eds.), *Primate societies*. Chicago: University of Chicago Press.

Hrdy, S. B., & Williams, G. C. (1983). Behavioral biology and the double standard. In S. K. Wasser (Ed.), *Social behavior of female vertebrates*. New York: Academic Press.

Jay, P. (1963). The female primate. In S. Farber & R. Wilson (Eds.), *The potential of woman*. New York: McGraw-Hill.

Jolly, A. (1985). *The evolution of primate behavior*. New York: Macmillan.

Kawai, M. (1958). On the system of social ranks in a natural troop of Japanese monkeys: I. Basic rank and dependent rank. *Primates, 1–2*, 111–130.

Kawamura, S. (1958). Matriarchal social ranks in the Minoo-B troop: A study of the rank system of Japanese monkeys. *Primates, 1–2*, 149–156.

Keller, E. F. (1983). *A feeling for the organism: The life and work of Barbara McClintock*. New York: W. H. Freeman.

Kleiman, D. (1977). Monogamy in mammals. *Quarterly Review of Biology, 52*, 39–69.

Kleiman, D., & Malcolm, J. (1981). The evolution of male parental investment in mammals. In D. J. Gubernick & P. H. Klopfer (Eds.), *Parental care in mammals*. New York: Plenum Press.

Koyama, N. (1967). On dominance rank and kinship of a wild Japanese monkey in Arashiyama. *Primates, 8*, 189–216.

Lamb, M. (1984). Observational studies of father-child relationships in humans. In D. Taub (Ed.), *Primate paternalism*. New York: Van Nostrand Reinhold.

Lancaster, J. (1975). *Primate behavior and the emergence of human culture*. New York: Holt, Rinehart and Winston.

Lott, D. (1981). Sexual behavior and intersexual strategies in American Bison. *Zeitschrift für Tierpsychologie, 56*, 97–114.

Lumpkin, S. (1983). Female manipulation of male avoidance of cuckoldry behavior in the ring dove. In S. C. Wasser (Ed.), *The social behavior of female vertebrates*. New York: Academic Press.

MacKinnon, J. (1978). *The ape within us*. New York: Holt, Rinehart and Winston.

Markow, T. A., & Ankney, P. F. (1984). *Drosophila* males contribute to oogenesis in a multiple mating species. *Nature, 224*, 302–303.

Moore, J. (1985). Demography and sociality in primates. Doctoral dissertation, Harvard University. Cambridge.

Mori, U., & Dunbar, R. I. M. (in press). Changes in the reproductive condition of female gelada baboons following the takeover of one-male units. *Zeitschrift für Tierpsychologie*.

Morris, S. (1979, August). Darwin and the double standard. *Playboy Magazine*.

Parker, G. A. (1970). Sperm competition and its evolutionary consequences in the insects. *Biological Review, 45*, 525–567.

Ransom, T., & Ransom, B. (1971). Adult-male–infant interactions among baboons (*Papio anubis*). *Folia Primatologica, 16*, 179–195.

Rowell, T. (1972). *Social behaviour of monkeys*. Baltimore, MD: Penguin Books.

Rowell, T. (1984). Introduction: Mothers, infants and adolescents. In M. Small (Ed.), *Female primates*. New York: Alan Liss.

Sayers, J. (1982). *Biological politics*. London: Tavistock.

Scarr, S. (1984). *Mother care: Other care*. New York: Basic Books.

Seyfarth, R. (1978). Social relationships between adult male and female baboons, part 2: Behavior throughout the female reproductive cycle. *Behaviour, 64*, nos. 3–4, 227–247.

Shaw, E., & Darling, J. (1985). *Female strategies*. New York: Walker.

Sherfey, M. J. (1973). *The evolution of female sexuality* (first published 1966). New York: Vintage Books.

Small, M. (Ed.). (1984). *Female primates*. New York: Alan Liss.

Small, M. (Forthcoming). Primate female sexual behavior and conception: Is there really sperm to spare?

Smith, J. M. (1984). Optimization theory in evolution. In E. Sober (Ed.), *Conceptual issues in evolutionary biology*. Cambridge, MA: The MIT Press.

Smith, R. (1984). Sperm competition. In *Sperm competition and the evolution of animal mating systems*. New York: Academic Press.

Smuts, B. B. (1985). *Sex and friendship in baboons*. New York: Aldine Publishing Co.

Stacey, P. B. (1982). Female promiscuity and male reproductive success in social birds and mammals. *The American Naturalist, 120*, no. 1, 51–64.

Stein, D. (1981). The nature and function of social interactions between infant and adult male yellow baboons (*Papio cynocephalus*). Doctoral dissertation, University of Chicago.

Stein, D. (1984). Ontogeny of infant–adult male relationships during the first year of life for yellow baboons (*Papio cynocephalus*). In D. Taub (Ed.), *Primate paternalism*. New York: Van Nostrand Reinhold.

Symons, D. (1979). *The evolution of human sexuality*. Oxford: Oxford University Press.

Taub, D. (1980). Female choice and mating strategies among wild Barbary macaques (*Macaca sylvana*). In D. Lindburg (Ed.), *The macaques*. New York: Van Nostrand Reinhold.

Taub, D. (1984a). Male–infant interactions in baboons and macaques: A critique and reevaluation. Paper presented at the American Zoological Society Meetings, Philadelphia, PA.

Taub, D. (1984b). *Primate paternalism*. New York: Van Nostrand Reinhold.

Tiger, L. (1977). The possible biological origins of sexual discrimination. In D. W. Brothwell (Ed.), *Biosocial man*. London: The Eugenics Society.

Tilson, R. (1981). Family formation strategies of Kloss' gibbons. *Folia Primatologica, 35*, 259–287.

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man*. Chicago: Aldine.

Trivers, R. L. (1985). *Social evolution*. Menlo Park, CA: Benjamin/Cummings.

Tsingalia, H. M., & Rowell, T. E. (1984). The behaviour of adult male blue monkeys. *Zeitschrift für Tierpsychologie, 64*, 253–268.

Turner, M. E., & Anderson, W. W. (1983). Multiple mating and female fitness in *Drosophilia pseudoobscura*. *Evolution, 37*, no. 4, 714–723.

Tutin, C. (1975). Sexual behaviour and mating patterns in a community of wild chimpanzees (*Pan troglodytes schweinfurthii*). Doctoral dissertation submitted to the University of Edinburgh, Edinburgh.

Van den Berghe, E. (1984). Female competition, parental care, and reproductive success in salmon. Paper presented at Animal Behavior Society Meetings, Cheney, Washington, August 13–17.

Vogt, J. (1984). Interactions between adult males and infants in prosimians and New World monkeys. In D. Taub (Ed.), *Primate paternalism*. New York: Van Nostrand Reinhold.

Wasser, S. C. (Ed.). (1983). *The social behavior of female vertebrates*. New York: Academic Press.

Wasser, S. C., & Barash, D. (1984). Reproductive suppression among female mammals. *Quarterly Review of Biology*, 513–538.

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.

Williams, G. C. (1975). *Sex and evolution*. Princeton, NJ: Princeton University Press.

Wilson, E. O. (1978). *On human nature*. Cambridge: Harvard University Press.

Wirtz, P. (1983). Multiple copulations in the Waterbuck. *Zeitschrift für Tierpsychologie, 61*, 78–82.

Wright, P. (1984). Biparental care in *Aotus trivirgatus* and *Callicebus molloch*. In M. Small (Ed.), *Female primates*. New York: Alan Liss.

Zuckerman, Sir S. (1932). *The social life of monkeys and apes*. London: Butler and Turner, Ltd.

8 Pre-theoretical Assumptions in Evolutionary Explanations of Female Sexuality

Elisabeth Lloyd

My contribution to this Symposium focuses on the links between sexuality and reproduction from the evolutionary point of view.[1] The relation between women's sexuality and reproduction is particularly important because of a vital intersection between politics and biology—feminists have noticed, for more than a century, that women's identity is often defined in terms of her reproductive capacity. More recently, in the second wave of the feminist movement in the United States, debates about women's identity have explicitly included sexuality; much feminist argument in the late 1960's and early 1970's involved an attempt to separate out an autonomous female sexuality from women's reproductive functions.

It is especially relevant, then, to examine biological arguments, particularly evolutionary arguments, to see what they say about *whether* and *how* women's sexuality is related to reproduction. We shall find that many evolutionary arguments seem to support the direct linking of female sexuality and reproduction. Yet I will argue that this support is not well-grounded. In fact, I think evolutionary explanations of female sexuality exemplify how social beliefs and social agendas can influence very *basic* biological explanations of fundamental physiological processes. In this paper, I shall spend some time spelling out a few examples in which assumptions about the close link between reproduction and sexuality yield misleading results, then I shall conclude with a discussion of the consequences of this case study for issues in the philosophy of science.

The fundamental problem is that it is simply *assumed* that every aspect of female sexuality should be explained in terms of reproductive functions. But there is quite a bit of biological evidence that this is an empirically incorrect assumption to make. This raises the question of why autonomous female sexuality, distinct from reproductive functions, got left out of these explanations. I shall ultimately conclude that social context is playing a large and unacknowledged role in the practice of this science.

Philosophical Studies 69: 139–153, Kluwer Academic Publishers, with kind permission from Springer Science and Business Media.

Perhaps the notion of the potential independence of female sexuality and reproduction may be unclear: I suggest thinking in terms of two distinct models, one in which all basic aspects of sexuality are *explained* in terms of reproduction, and the other in which sexuality is seen as an autonomous set of functions and activities, which are only *partially* explained in terms of reproductive functions. The difference may seem minor, but the two models have significantly disparate consequences when used in scientific explanation.

Let us begin with a classic and wide-spread model representing the hormonal determination of sexual behavior. In this model, female animals are only willing to have sexual intercourse when they are fertile—their sexual interest and activity are completely hormonally controlled. Typical and familiar examples of this type of set-up include rats, dogs, and cats. When these animals are in estrus, they are willing and eager to mate, otherwise not. Technically, estrus is defined hormonally—that is, estrus is a particular phase of the menstrual cycle, in which the animal is fertile, and certain hormone measures are very high. This model embodies a *very tight* link between sexuality and reproduction: female sexuality functions completely in the service of reproduction.

Some interesting problems arise, however, in the application of this hormonally deterministic picture to human and non-human primate behavior.

First of all, although estrus is biologically defined as a hormonal state, it is very common for estrus to be defined *operationally* as the period in which "the female is willing to participate in sex." In one species, the bonobos, this behavioral definition led to the comic conclusion that this species is in estrus 57–86% of the time.[2] Notice that identifying estrus in this manner amounts to an *enforcement* of the belief that sexual behavior is tightly linked to reproduction. It becomes impossible even to *ask* whether these primates have an active sexual interest outside of their peak hormonal periods.

It turns out that when independent studies are made, sexual activity is not confined to the fertile phase for a number of non-human primates, including rhesus monkeys, several species of baboons, and common chimpanzees.[3] Social factors such as partner preferences can be as influential as hormonal factors in regulating sexual behavior in several of these species.[4]

Female homosexual activity provides a good test for the assumed dependence of female sexuality on hormonal status. In addition, homosexual behaviors are clearly independent of reproduction *per se*, and might be interpreted as an indicator of an autonomous female sexuality. It turns out that female homosexual activities, which are widely observed in non-human primates, seem to be *independent* of the hormonal status of the participants. This independence has led some researchers to ignore such behaviors, or to declare that they are not, in fact, sexual. For example, pygmy chimpanzee females are commonly observed engaging in "genito-genital rubbing" (called "GG rubbing") in which two females hold each other and "swing their hips laterally while keeping the front tips of vulvae, where the clitorises protrude, in touch with each other."[5] Kano argues that this behavior is not sexual, because non-human primates can only be "sexual" during estrus; the fact that pygmy chimps engage in GG rubbing outside of estrus, claims Kano, itself "suggests that this behavior does not occur exclusively in a sexual context, but has some other social significance."[6] Generally, some caution about the interpretation of apparently sexual behaviors is appropriate; the misunderstanding of many dominance behaviors as sexual ones plagued primatology in its first decades. At stake in this case, however, is the very *possibility* of hormonally independent female sexuality. The issue was resolved in 1984, when Mori, using a detailed study of statistical relations among behaviors, concluded that GG rubbing was, in fact, sexual behavior, since the same cluster of behavior surrounded both mating and GG rubbing.[7]

A more blatant example of researcher bias typing reproduction and female sexuality tightly together appears in an experiment being done on female orgasm in stumptail macaques. The original studies on female macaque orgasm, completed in the 1970's, documented female orgasm in the context of female homosexual mounting—that is, one female mounts another female, and stimulates herself to orgasm.[8] One very interesting result of these studies was the finding that the mounting, orgasmic female was *never* in estrus when these orgasms occurred. This is a provocative result for several reasons. First, according to the hormonal determinism model, female macaques are not supposed to be interested in any sexual activity outside of estrus; Second, these same female macaques *never* evidenced any sign of orgasm when they were participating in heterosexual coitus. A later study of the same species documented the same basic patterns, with the exception that four out of ten females in the group seemed, occasionally, to have orgasm during heterosexual coitus.[9]

I was surprised, therefore, when I spoke with a researcher who was working on the evolution of female orgasm in stumptail macaques.[10] He described his experimental set-up to me with some enthusiasm: the females are radio-wired to record orgasmic muscle contractions and increased heartrate, etc. This sounds like the ideal experiment, because it can record the sex lives of the females mechanically, without needing a human observer. In fact, the project had been funded by the NIH, and had presumably gone through the outside referee and panel reviews necessary for funding. But then the researcher described to me the clever way he had set up his equipment to record the female orgasms—he wired up the heart rate of the *male* macaques as the signal to start recording the *female* orgasms. When I pointed out that the vast majority of female stumptail orgasms occurred during sex among the females alone, he replied that yes, he knew that, but he was only interested in the *important* orgasms.

Obviously, this is a very unfortunate case. But it is not an isolated incident. Observations, measurements, interpretations, and experimental design are all affected by the background assumptions of the scientists. There is a pervasive and undefended assumption that female sexuality in non-human primates is tightly linked to reproduction. I would like now to explore briefly the situation regarding human beings.

Human Cases

In most of the literature on the evolution of human sexuality, much attention is paid to the distinct attributes of human beings. The continual sexual ''receptivity'' of the human female is contrasted with the (supposed) strict hormonal restrictions on sexual activity in non-human animals. Human beings are supposed to be uniquely adapted to be sexually free from hormonal dictates, the possessors of a separate and self-constructed sexuality. When it comes to evolutionary explanations of women's sexuality, though, the tight connection between reproduction and sexuality remains firmly in place.

To continue with the hormonal theme, we can begin by looking at beliefs about the distribution of female sexual interest during the menstrual cycle. Many researchers, in evolutionary biology, behavior, and physiology, have *deduced* that it must be the case in human females that peak sexual interest and desire occur at the same time as peak fertility. This conclusion is a simple extension of the hormonal determinism model from mice and dogs. While this may have the ring of a reasonable assumption, it is not supported by the clinical literature. Kinsey, for example, found that 59% of his female sample experienced patterns of fluctuation in their sexual desire during their cycle—but only 11% experience a peak of sexual desire in mid-cycle, when they are most likely to be fertile.[11] More recently, Singer and Singer, in a survey of studies, found that only 6–8% of women experience an increase in sexual desire around the time of ovulation. Most studies found peaks of sexual desire right before and after menstruation, when the woman is almost invariably infertile.[12]

Hence, the majority of evidence supports a picture in which female sexual interest and activity is clearly *decoupled* from her reproductive state. Sexual interest and motivation is highest when the woman is least likely to conceive. Unfortunately, a number of researchers working in the area of the evolution of sexuality have not taken this on board, and continue to assert that peak sexual desire *must* be around the time of ovulation—otherwise it would not make any sense.

This ''making sense'' is precisely what I'm interested in. According to these researchers, female sexuality doesn't *make sense* unless it is in the service of reproduction. There is no scientific defense offered for this assumption. A similar assumption is also present in the evolutionary explanations offered for female orgasm.

I have examined thirteen stories for the evolution of human female orgasm, and all except one of these stories assume that orgasm is an evolutionary adaptation. That is, they assume that orgasm conferred a *direct selective* advantage on its possessors, and that is how it came to be prevalent among women. The most common general formula for explaining the evolution of human female orgasm is through the pair-bond. Here, the pair-bond means more-or-less monogamous heterosexual coupling, and it is argued that such coupling increases the potential reproductive success of both parties through mutual cooperation and assistance with rearing offspring. The idea is that the male and the female in the pair bond provide mutual support to one another, and assist each other in rearing offspring, and that offspring raised under these conditions will tend themselves to have higher survival and reproductive success than those raised under other circumstances.

Hence, pair-bonding is seen as an adaptation in the evolutionary sense—it exists *because* it confers better chances of surviving and reproducing to those who display the trait. Under the assumption that pair-bonds are adaptive, frequent intercourse is also seen as adaptive, since it helps "cement the pair bond." And this is where orgasm comes in. Orgasm evolved, according to these pair-bond theorists, because it gave the female a reward and motivation to engage in frequent intercourse, which is itself adaptive, because it helps cement the pair bond. A number of different theorists have developed permutations of this basic story, but it remains the most widely accepted evolutionary story for female orgasm.[13]

Now, there is a glaring problem with this story—It assumes that intercourse is reliably connected to orgasm in females. All of the available clinical studies on women's sexual response indicate that this is a problematic assumption. Somewhere between 20–35% of women always or almost always experience orgasm with unassisted intercourse.[14] I should add that this figure is supported by what cross-cultural information exists.[15] This figure is very low, and it is especially striking given that somewhere around 90% of women do experience orgasm. Furthermore, about 30% of women *never* have orgasm with intercourse—this figure is taken from a population of women who do have regular intercourse, and of whom almost all are orgasmic.[16] What this means is that *not* to have orgasm from intercourse is the experience of the majority of women the majority of the time. Not to put too fine a point on it, if orgasm is an adaptation which is a reward for engaging in frequent intercourse, it does not seem to work very well.

Obviously, this observation does not rule out the possibility that there is some selective advantage to female orgasm, but the salient point is that *none of these pair bond theorists even address this problem*, which I call the orgasm-intercourse discrepancy. Rather they simply assume that when intercourse occurs, so does orgasm.[17]

In general, the association of intercourse with orgasm is relatively unproblematic among males. Hence, what is being assumed here is that female sexual response is like male sexual response to the same situation. There is little or no awareness, among the pair-bond theorists, of the orgasm-intercourse discrepancy, in spite of the fact that they cite or refer to the very studies which document this fact, including Kinsey's 1953 report on women's sexual response.

There is one obvious and understandable reason for this slip. They are, after all, trying to explain orgasm through evolutionary theory, which involves showing that the trait gave a reproductive advantage to its owner. It's easy to see how the equation of reproduction through intercourse and orgasm went by unnoticed. Nevertheless, this case does illustrate the main thesis, that female sexuality is unquestioningly equated with reproduction, and with the sort of sex that leads to reproduction.

There is another intriguing line of argument for the adaptive value of female orgasm, which was first published by Desmond Morris in 1967, though Shirley Strum tells me that Sherwood Washburn was teaching this in his classes at Berkeley earlier. Morris claimed that orgasm had a special function related to bipedalism (that is, walking on our hind legs), because it would increase chances of fertilization. Here again we have the direct link between female sexuality and reproduction.

> It does this in a rather special way that applies only to our own peculiar species. To understand this, we must look back at our primate relatives. When a female monkey has been inseminated by a male, she can wander away without any fear of losing the seminal fluid that now lies in the innermost part of her vaginal tract. She walks on all fours. The angle of her vaginal passage is still more or less horizontal. If a female of our own species were so unmoved by the experience of copulation that she too was likely to get up and wander off immediately afterwards, the situation would be different, for she walks bipedally and the angle of her vaginal passage during normal locomotion is almost vertical. Under the simple influence of gravity the seminal fluid would flow back down the vaginal tract and much of it would be lost. There is therefore . . . a great advantage in any reaction that tends to keep the female horizontal when the male ejaculates and stops copulation. The violent response of female orgasm, leaving the female sexually satiated and exhausted, has precisely this effect.[18]

Morris' view is in turn based on his understanding of physiological response—he says earlier . . . "after both partners have experienced orgasm [in intercourse] there normally follows a considerable period of exhaustion, relaxation, rest and frequently sleep."[19] Similarly, he claims, "once the climax has been reached, all the [physiological] changes noted are rapidly reversed and the resting, post-sexual individual quickly returns to the normal quiescent physiological state."[20]

Now let us refer to the clinical sex literature, which is cited by Morris and by others. According to this literature, the tendencies to states of sleepiness and exhaustion following orgasm, are, in fact, true for men but not for women. Regarding Morris's claim that the physiological changes are "rapidly reversed," this is also true for men but not for women—women return to the plateau phase of sexual excitement, and not to the original unexcited phase, as men do. This was one of the most noted conclusions of Masters and Johnson, whose picture of sexual response was enthusiastically adopted by Morris—but, it seems, only in part.[21]

In fact, Masters and Johnson publicized an interesting and important difference between men's and women's sexuality, and that is the capacity of many women to have more than one orgasm without a significant break. 47% of the women in Hite's survey did not feel that a single orgasm was always satisfying to them, and many women wanted more, some as many as 15–25. If, at this point, you are concerned about Hite's bad reputation as a statistician and researcher, I'd like to point out that many of Hite's findings in that first study, published as the Hite report, were consistent with Kinsey's figures, and the Kinsey reports are considered, to this day, and in spite of any problems they might have, to be the best general studies ever done on the topic of women's sexuality.[22] Masters and Johnson contrast the ability of many women to have five or six orgasms within a matter of minutes with the adult male's usual inability to have more than one orgasm in a short period.[23] This female ability is linked to the fact that, following orgasm, women do not return to the pre-aroused state, as men do, but instead to the plateau phase of excitement.

Hence, Morris's story is in trouble. He claims that the physiological changes are rapidly reversed for women as well as for men. He also neglects the sizable percentage of women who are not satisfied by a single orgasm. Given the documented tendency in men to sleep and exhaustion following a single orgasm, it's not at all clear that a female desire to have orgasm wouldn't have exactly the opposite effect from that described by Morris—perhaps the woman would jump right up and cruise for a little more action at precisely the time when the sperm are most likely to leak out.

Actually, another serious problem with this story was recently pointed out by Shirley Strum, an expert on baboon behavior.[24] Supposedly, the selection pressure shaping female sexual response here is the potential loss of sperm that is threatened because human beings walk on two legs, and because the vaginal position is thus changed from horizontal to almost vertical. One would think, then, that our relatives walking on four legs would be protected against this occurrence, for anatomical reasons. But Strum says that immediately following intercourse, female baboons like to go off and *sit down* for ten or fifteen minutes. When they get up, she says, they inevitably leave a visible puddle of semen on the ground. Perhaps, then, the loss of semen is not the serious evolutionary challenge that Desmond Morris and others take it to be.

Summary

I claim that social agendas appear in these stories through the obliteration of any female sexual response that is independent from her function as a reproducer. Autonomous, distinct female sexual response just disappears.

In these explanations women are presumed to have orgasms nearly always with intercourse, as men do. Women are presumed to return to the resting state following orgasm, as men do. One could object that Morris is a relatively easy target, so I will offer the following tidbit in defense of my analysis. Gordon Gallup and Susan Suarez published, in 1983, a technical discussion on optimal reproductive strategies for bipedalism, and took up Morris' anti-gravity line of argument. They argue that orgasm would be adaptive because it would keep the woman lying down, and hence keep the semen from escaping. In the context of these paragraphs on female orgasm, they state, "it is widely acknowledged that intercourse frequently acts as a mild sedative. The average individual requires about five minutes of repose before returning to a normal state after orgasm."[25] The scientific reference they offer for this particular generalization is Kinsey 1948, which is, in fact, exclusively on *male* sexual response. In other words, this "average individual" which figures in their story about female orgasm, is, in fact, explicitly male.

An Alternative Explanation

Donald Symons, in his book *The Evolution of Human Sexuality* (1979), argues that female orgasm is not an adaptation. He develops a story parallel to the one about male nipples—female orgasm exists because orgasm is strongly selected in males, and because of their common embryological form, women are born with the potential for having orgasms, too.[26] Part of the story, then, is that orgasm is strongly selected in males; this is fairly plausible, since it is difficult for male mammals to reproduce without ejaculation, which requires a reflex response in certain muscles. These muscles are, in fact, the same (homologous) muscles that are involved in female orgasm. It is also significant that the intervals between contractions in orgasm is 4/5 of a second in both men and women. This is considered evidence that orgasm is a reflex with the same developmental origin in both sexes.

One of the consequences of Symons' theory is that it would be expected that similar stimulation of the clitoris and penis would be required to achieve the same reaction or reflex response. This similarity shows especially in the figures on masturbation. Only 1.5% of women masturbate by vaginal entry, which provides stimulation similar to the act of intercourse; the rest do so by direct or indirect stimulation of the clitoris itself.[27] Also, on the developmental theory, one would *not* expect similar reactions to intercourse, given the differences in stimulation of the homologous organs.

Finally, this theory is also supported by the evidence of orgasm in non-human primates. The observed orgasms occur almost exclusively when the female monkeys are themselves mounting other monkeys, and not during copulation. On the non-adaptive view of orgasm, this is almost to be expected. There, female orgasm is defined as a potential, which, if the female gets the right sort and amount of stimulation, is activated. Hence, it is not at all surprising that this does not occur often during copulation, which in these monkeys includes very little, if any, stimulation of the clitoris, but occurs rather with analogous stimulation of the homologous organs that they get in mounting.

Symons' proposal, which I found very powerful and plausible, has been sharply criticized by a number of feminists. For instance, a leading feminist sociobiologist, Sarah Blaffer Hrdy, claims that this non-adaptive explanation is dismissive of female sexuality.[28] Similarly, Mina Caulfield accuses Symons of denying the ''significance of female pleasure.''[29]

I view these criticisms as misguided, because they are based on the assumption that *only* adaptive explanations can provide for the significance of a trait. But why should we believe this? Musical and singing ability are not adaptations, but they are very important to human culture and human life. One must have adopted the idea, not merely that ''what is natural is good,'' but further, that ''only what is adaptive is good.'' The evolutionarily derivative role of female orgasm implies absolutely nothing about its importance unless you are a committed adaptationist. Finally, I wonder why these feminists are so eager to get orgasm defined as an adaptation—several of the serious evidential problems with evolutionary explanations about female orgasm arose, I have argued, from making an easy connection between sexuality and reproduction.

I would like to just mention a possible alternative interpretation. The conclusion that orgasm is not an adaptation *could* be interpreted as emancipatory. After all, the message here is that orgasm is a freebie. It can be used in any way that people want; there is no ''natural'' restriction on female sexual activities, nor is there any scientific ground for such a notion. Under the developmental view, the constraints are loosened on possible explanations about women's sexuality that are consistent with accepted clinical conclusions and with evolutionary theory. Hence, the realm formerly belonging exclusively to reproductive drive would now be open to much, much more.

Discussion

I would like to draw two conclusions.

First, I believe that prior assumptions have more influence in these areas of science than is commonly acknowledged in the usual philosophical and scientific pictures of scientific theorizing and testing. In the cases examined here, science is not very separate from the social and cultural context. Rather, social assumptions and prior commitments of the scientists play a major role in the practice of science itself, at many levels—experimental design, data collection, predictions, hypothesis formulation, and the evaluation of explanations.

To understand this area of scientific practice, we need a view of science that is more sophisticated, one that has more moving parts, than the pictures typically presented by philosophers of science. Under the usual approaches, science is seen as involving relations purely between theory and data, or between theory, data, and explainer. But this is not enough. We need a way to recognize and analyze the vital role of pretheoretical beliefs and categories in *all* stages of scientific research.

One might object that the subject matter of this part of science makes social influence inevitable, and that one would not expect this same level of cultural bias in other scientific contexts. That's probably right. But we do not need to show social forces at work in every possible case of scientific inquiry in order to insist on having a theory of science with enough flexibility to work in many areas. The cases I have presented here are definitely "science," with plenty of funding, backing, authority, influence, and prestige. Philosophers who insist on a *pure* view of science, based on isolated and idealized examples of physics, are voting themselves out of the action. There are very interesting and important things going on in other areas as well, as the cases I have outlined above attest. Developing a view of science which can account for these other fields is vital.

My suggestion does *not* involve commitment to a relativist position. In a complete analysis of evolutionary explanations of human sexuality, I would adopt Helen Longino's general approach, in which she characterizes objectivity in science as resulting from the critical interaction of different groups and individuals with different social and cultural assumptions and different stakes. Under this view, the irreducibility of the social components of the scientific situation is accounted for—these social assumptions are, in fact, an essential part of the picture of scientific practice.

At any rate, I take it that the cases I have described above violate our common philosophical understandings of how we arrive at scientific beliefs, how knowledge is created, and how science words. If philosophers go the route of labelling as "science" *only* that which obeys the demands of current philosophy, we will end up discussing only some parts of physics and maybe some math. Meanwhile, what about the rest of science—biology, social sciences, anthropology, psychology, biochemistry? I suggest adopting and developing recent contextualist and feminist views of science, which take explicit account of pretheoretical assumptions and preconceptions, and their social origins.

This case involving female sexuality is very interesting because there are *two* very strong forces working to put sex and reproduction together. Adaptationism, within biology, promotes the easy linking of all sexual activity with reproduction success, the measure of relative fitness. Secondly, the long social tradition of *defining* women in terms of their sexual and reproductive functions alone also tends to link sexuality and reproduction more tightly than the evidence indicates.

The long struggle by various women's movements to separate sex and reproduction seems to have had very little effect on the practice of the science we have examined in this paper. This is especially ironic, because politically, ever since the late Nineteenth Century, scientific views about gender differences and the biology of women have been the single most powerful political tool against the women's movements. My second and more controversial conclusion is that current "purist" philosophy of science actually *contributes to* that political power by reinforcing myths of the insulation of scientific endeavors from social influences. A more sophisticated understanding of the production and evaluation of scientific knowledge would mean seeing science as (partly) a continuation of politics. Science would then lose at least *some* independent authority in the political arena. Judging by the scientific work that I have discussed in this paper, I think that would be a good thing.

Notes

1. This paper contains the text of an invited lecture delivered at the Symposium, Sex and Reproduction, Pacific Division APA, 1992. Further documentation and discussion will be found in *All About Eve: Bias in Evolutionary Explanations of Women's Sexuality* (Harvard University Press, forthcoming).

2. Kano, T. (1982) "The social group of Pygmy Chimpanzees of Wamba," *Primates* 23(2), 171–188.

3. Hafez, E. S. E. (1971) "Reproductive cycles," in *Comparative Reproduction of Non-human Primates*, ed. E. S. E. Hafez. Springfield, IL: Charles C. Thomas.

4. Wolfe, L. (1979) "Behavioral patterns of estrous females of the Arachiyama West troop of Japanese Macaques (*Macaca fuscata*)," *Primates* 20(4), 525–534.

5. Koruda, S. (1980) "Social behavior of the Pygmy Chimpanzees," *Primates* 21(2), 181–197. Quote from p. 189.

6. Kano, T. (1980) "Special behavior of wild Pygmy Chimpanzees (*Pan paniscus*) of Wambe: A preliminary report," *Journal of Human Evolution* 9, 243–260. Quote from p. 243.

7. Mori, A. (1984) "A ethological study of Pygmy Chimpanzees in Wambe Zaire: A comparison with Chimpanzees," *Primates* 25(3), 255–278.

8. Chevalier-Skolnikoff, S. (1974) "Male-female, female-female, and male-male sexual behavior in the Stumptail Monkey, with special attention to the female orgasm," *Archives of Sexual Behavior* 3(2), 95–116; (1976) "Homosexual behavior in a laboratory group of Stumptail monkeys (*Macaca arctoides*): Forms, contexts, and possible social functions," *Archives of Sexual Behavior* 5(6), 511–527.

9. Goldfoot, D., J. Westerborg-van Loon, W. Groeneveld, and A. Koos Slob (1980) "Behavioral and physiological evidence of sexual climax in the female stump-tailed macaque (*Macaca arctoides*)," *Science* 208, 1477–1479.

10. The identity of this researcher is not included for publication. The information stated here was obtained through personal communication.

11. Kinsey, A. C. et al. (1953) *Sexual Behavior in the Human Female*. Philadelphia: W. B. Saunders.

12. Singer, I. and J. Singer (1972) "Periodicity of sexual desire in relation to time of ovulation in women," *Journal of Biosocial Science* 4, 471–481.

13. Morris' work has been criticized by later researchers as being methodologically flawed (e.g., Wilson (1975) *Sociobiology*, Harvard UP; Crook (1972) "Sexual selection, dimorphism, and social organization in the primates," in *Sexual Selection and the Descent of Man*, ed. B. Campbell. Chicago: Aldine) but it is still widely cited, and its basic premises are accepted or slightly modified in other respected accounts such as: F. Beach (1973) "Human sexuality and evolution," in *Advances in Behavioral Biology*, eds. Wm. Montanga and Wm. Sadler. NY: Plenum Press, pp. 333–365; G. Pugh (1977) *Biological Origins of Human Values*, NY: Basic Books; Crook (1972); and B. Campbell (1967) *Human Evolution: An Introduction to Man's Adaptations*. Chicago: Aldine.

14. Hite, S. (1976) The Hite Report, NY: Macmillan; Kinsey, A. et al. op cit.

15. See, e.g., Davenport, W. (1977) "Sex in cross-cultural perspective," in *Human Sexuality in Four Perspectives*, ed. F. Beach. Johns Hopkins U. Press, pp. 115–163.

16. Hite, S. op cit.

17. Typically, in evolutionary explanations, if a trait is taken to have evolved as an adaptation, yet is rarely used in the adaptive context, some explanation of the details of the selection pressure or the extreme adaptive value of the trait is offered.

18. Morris, D. (1967) *The Naked Ape*. London: Jonathan Cape, p. 79.

19. Ibid., p. 55.

20. Ibid., p. 59.

21. Masters, W. H. and V. Johnson (1966) *Human Sexual Response*. Boston: Little, Brown.

22. Kinsey (1953, pp. 375–376); Hite (1976, p. 417); Masters and Johnson (1966, p. 65).

23. Masters, W. H. and V. Johnson (1961) "Orgasm, anatomy of the female," in *Encyclopedia of Sexual Behavior*, Vol. II, eds. A. Ellis and A. Abar-banal. New York: Hawthorn, p. 792.

24. Personal Communication.

25. Gallup and Suarez (1983) "Optimal reproductive strategies for bipedalism," *Journal of Human Evolution* 12, 195.

26. This argument is spelled out in more detail by Stephen Jay Gould, in an essay that was based on my research and arguments ("Freudian Slip," *Natural History*, Feb. 1987. pp. 14–21).

27. Kinsey, 1953; Hite (1976), pp. 410–411.

28. Hrdy, S. B. (1981) *The Woman That Never Evolved*. Cambridge, MA: Harvard University Press, p. 165.

29. Caulfield, M. D. (1985) "Sexuality in human evolution: What is 'natural' in sex?" *Feminist Studies* 11(2), 343–363.

V Evolutionary Psychology

9 Toward Mapping the Evolved Functional Organization of Mind and Brain

John Tooby and Leda Cosmides

The human brain is a biological system produced by the evolutionary process, and thus, cognitive neuroscience is itself a branch of evolutionary biology. Accordingly, cognitive neuroscientists can benefit by learning about and applying the technical advances made in modern evolutionary biology. Among other things, evolutionary biology can supply researchers with (1) the biologically rigorous concept of function appropriate to neural and cognitive systems, (2) a growing list of the specialized functions the human brain evolved to perform, and (3) the ability to distinguish the narrowly functional aspects of the neural and cognitive architecture that are responsible for its organization from the much larger set of properties that are byproducts or noise. With these and other tools, researchers can construct experimental stimuli and tasks that activate and are meaningful to functionally dedicated subunits of the brain. The brain is comprised of many such subunits: evolutionarily meaningful stimuli and tasks are far more likely than arbitrary ones to elicit responses that can illuminate their complex functional organization.

Nothing in biology makes sense except in the light of evolution.
—T. Dobzhansky

It is the theory which decides what we can observe.
—A. Einstein

Seeing with New Eyes: Toward an Evolutionarily Informed Cognitive Neuroscience

The task of cognitive neuroscience is to map the information-processing structure of the human mind and to discover how this computational organization is implemented in the physical organization of the brain. The central impediment to progress is obvious: The human brain is, by many orders of magnitude, the most complex system that humans have yet investigated. Purely as a physical system, the vast intricacy of

From Michael Gazzaniga, ed., *The New Cognitive Neurosciences*, 2d ed., Cambridge: The MIT Press, 2000). Reprinted by permission of The MIT Press.

chemical and electrical interactions among hundreds of billions of neurons and glial cells defeats any straightforward attempt to build a comprehensive model, as one might attempt to do with particle collisions, geological processes, protein folding, or host-parasite interactions. Combinatorial explosion makes the task of elucidating the brain's computational structure even more overwhelming: There is an indefinitely large number of specifiable inputs, measurable outputs, and possible relationships between them. Even worse, no one yet knows with certainty how computations are physically realized. They depend on individuated events within the detailed structure of neural microcircuitry largely beyond the capacity of current technologies to observe or resolve. Finally, the underlying logic of the system has been obscured by the torrent of recently generated data.

Historically, however, well-established theories from one discipline have functioned as organs of perception for others (e.g., statistical mechanics for thermodynamics). They allow new relationships to be observed and make visible elegant systems of organization that had previously eluded detection. It seems worth exploring whether evolutionary biology could provide a rigorous metatheoretical framework for the brain sciences, as they have recently begun to do for psychology (Shepard, 1984, 1987a, 1987b; Gallistel, 1990; Cosmides and Tooby, 1987; Pinker, 1994, 1997; Marr, 1982; Tooby and Cosmides, 1992).

Cognitive neuroscience began with the recognition that the brain is an organ designed to process information and that studying it as such would offer important new insights. Cognitive neuroscientists also recognize that the brain is an evolved system, but few realize that anything follows from this second fact. Yet these two views of the brain are intimately related and, when considered jointly, can be very illuminating.

Why Brains Exist

The brain is an organ of computation that was built by the evolutionary process. To say that the brain is an organ of computation means that (1) its physical structure embodies a set of programs that process information, and (2) that physical structure is there *because* it embodies these programs. To say that the brain was built by the evolutionary process means that its functional components—its programs—are there *because* they solved a particular problem-type in the past. In systems designed by natural selection, function determines structure.

Among living things, there are whole kingdoms filled with organisms that lack brains (plants, Monera, fungi). The sole reason that evolution introduced brains into the designs of some organisms—the reason brains exist at all—is because brains performed computations that regulated these organisms' internal processes and external activities in ways that promoted their fitness. For a randomly generated modification in design to be selected—that is, for a mutation to be incorporated by means of a nonrandom process into a species-typical brain design—it had to improve the ability of organisms to solve adaptive problems. That is, the modification had to have a certain kind of effect: It had to improve the organisms' performance of some activity that systematically enhanced the propagation of that modification, summed across the species' range and across many generations. This means that the design of the circuits, components, systems, or modules that make up our neural architecture must reflect, to an unknown but high degree, (1) the computational task demands inherent in the performance of those ancestral activities and (2) the evolutionarily long-enduring structure of those task environments (Marr, 1982; Shepard, 1987a; Tooby and Cosmides, 1992).

Activities that promoted fitness in hominid ancestral environments differ in many ways from activities that capture our attention in the modern world, and they were certainly performed under radically different circumstances. (Consider: hunting *vs.* grocery shopping; walking everywhere *vs.* driving and flying; cooperating within a social world of ~200 relatives and friends *vs.* 50,000 strangers in a medium-sized city). The design features of the brain were built to specifications inherent in ancestral adaptive problems and selection pressures, often resulting in talents or deficits that seem out of place or irrational in our world. A baby cries—alerting her parents—when she is left to sleep alone in the dark, not because hyenas roam her suburban household, but because her brain is designed to keep her from being eaten under the circumstances in which our species evolved.

There is no single algorithm or computational procedure that can solve every adaptive problem (Cosmides and Tooby, 1987; Tooby and Cosmides, 1990a, 1992). The human mind (it will turn out) is composed of many different programs for the same reason that a carpenter's toolbox contains many different tools: Different problems require different solutions. To reverse-engineer the brain, one needs to discover functional units that are native to its organization. To do this, it is useful to know, as specifically as possible, what the brain is for—which specific families of computations it was built to accomplish and what counted as a biologically successful outcome for each problem-type. The answers to this question must be phrased in computational terms because that is the only language that can capture or express the functions that neural properties were naturally selected to embody. They must also refer to the ancestral activities, problems, selection pressures, and environments of the species in question because jointly these define the computational problems each component was configured to solve (Cosmides and Tooby, 1987; Tooby and Cosmides, 1990a, 1992).

For these reasons, evolutionary biology, biological anthropology, and cognitive psychology (when integrated, called *evolutionary psychology*) have the potential to supply to cognitive neuroscientists what might prove to be a key missing element in their research program: a partial list of the native information-processing functions that the human brain was built to execute, as well as clues and principles about how to discover or evaluate adaptive problems that might be proposed in the future.

Just as the fields of electrical and mechanical engineering summarize our knowledge of principles that govern the design of human-built machines, the field of evolutionary biology summarizes our knowledge of the engineering principles that govern the design of organisms, which can be thought of as machines built by the evolutionary process (for overviews, see Daly and Wilson, 1984; Dawkins, 1976, 1982, 1986; Krebs and Davies, 1997). Modern evolutionary biology constitutes, in effect, a foundational "organism design theory" whose principles can be used to fit together research findings into coherent models of specific cognitive and neural mechanisms (Tooby and Cosmides, 1992). To apply these theories to a particular species, one integrates analyses of selection pressures with models of the natural history and ancestral environments of the species. For humans, the latter are provided by hunter-gatherer studies, biological anthropology, paleoanthropology, and primatology (Lee and DeVore, 1968).

First Principles: Reproduction, Feedback, and the Antientropic Construction of Organic Design

Within an evolutionary framework, an organism can be described as a self-reproducing machine. From this perspective, the defining property of life is the presence in a system of "devices" (organized components) that cause the system to construct new and similarly reproducing systems. From this defining property—self-reproduction—the entire deductive structure of modern Darwinism logically follows (Dawkins, 1976; Williams, 1985; Tooby and Cosmides, 1990a). Because the replication of the design of the parental machine is not always error free, randomly modified designs (i.e., mutants) are introduced into populations of reproducers. Because such machines are highly organized so that they cause the otherwise improbable outcome of constructing offspring machines, most random modifications interfere with the complex sequence of actions necessary for self-reproduction. Consequently, such modified designs will tend to remove themselves from the population—a case of negative feedback.

However, a small residual subset of design modifications will, by chance, happen to constitute improvements in the design's machinery for causing its own reproduction. Such improved designs (by definition) cause their own increasing frequency in the population—a case of positive feedback. This increase continues until (usually) such modified designs outreproduce and thereby replace all alternative designs in the population, leading to a new species-standard design. After such an event, the population of reproducing machines is different from the ancestral population: The population- or species-standard design has taken a step "uphill" toward a greater degree of functional organization for reproduction than it had previously. This spontaneous feedback process—natural selection—causes functional organization to emerge *naturally*, that is, without the intervention of an intelligent "designer" or supernatural forces.

Over the long run, down chains of descent, this feedback cycle pushes designs through state-space toward increasingly well-organized—and otherwise improbable—functional arrangements (Dawkins, 1986; Williams, 1966, 1985). These arrangements are functional in a specific sense: the elements are improbably well organized to cause their own reproduction in the environment in which the species evolved. Because the reproductive fates of the inherited traits that coexist in the same organism are linked together, traits will be selected to enhance each other's functionality (however, see Cosmides and Tooby, 1981, and Tooby and Cosmides, 1990a, for the relevant genetic analysis and qualifications). As design features accumulate, they will tend to sequentially fit themselves together into increasingly functionally elaborated machines for reproduction, composed of constituent mechanisms—called *adaptations*—that solve problems that either are necessary for reproduction or increase its likelihood (Darwin, 1859; Dawkins, 1986; Thornhill, 1991; Tooby and Cosmides, 1990a; Williams, 1966, 1985). Significantly, in species like humans, genetic processes ensure that complex adaptations virtually always are species-typical (unlike nonfunctional aspects of the system). This means that *functional* aspects of the architecture will tend to be universal at the genetic level, even though their expression may often be sex or age limited, or environmentally contingent (Tooby and Cosmides, 1990b).[1]

Because design features are embodied in individual organisms, they can, generally speaking, propagate themselves in only two ways: by solving problems that increase the probability that offspring will be produced either by the organism they are situated in or by that organism's kin (Hamilton, 1964; Williams and Williams, 1957; however, see Cosmides and Tooby, 1981, and Haig, 1993, for intragenomic methods). An individual's relatives, by virtue of having descended from a recent common ancestor, have an increased likelihood of having the same design feature as compared to other conspecifics. This means that a design modification in an individual that causes an increase in the reproductive rate of that individual's kin will, by so doing, tend to increase its own frequency in the population. Accordingly, design features that promote both direct reproduction and kin reproduction, and that make efficient tradeoffs between the two, will replace those that do not. To put this in standard biological terminology, design features are selected to the extent that they promote their inclusive fitness (Hamilton, 1964).

In addition to selection, mutations can become incorporated into species-typical designs by means of chance processes. For example, the sheer impact of many random accidents may cumulatively propel a useless mutation upward in frequency until it crowds out all alternative design features from the population. Clearly, the presence of such a trait in the architecture is not explained by the (nonexistent) functional consequences that it had over many generations on the design's reproduction; as a result, chance-injected traits will not tend to be coordinated with the rest of the organism's architecture in a functional way.

Although such chance events play a restricted role in evolution and explain the existence and distribution of many simple and trivial properties, organisms are not primarily chance agglomerations of stray properties. Reproduction is a highly improbable outcome in the absence of functional machinery designed to bring it about, and only designs that retain all the necessary machinery avoid being selected out. To be invisible to selection and, therefore, not organized by it a modification must be so minor that its effects on reproduction are negligible. As a result, chance properties do indeed drift through the standard designs of species in a random way, but they are unable to account for the complex organized design in organisms and are, correspondingly, usually peripheralized into those aspects that do not make a significant impact on the functional operation of the system (Tooby and Cosmides, 1990a, 1990b, 1992). Random walks do not systematically build intricate and improbably functional arrangements such as the visual system, the language faculty, face recognition programs, emotion recognition modules, food aversion circuits, cheater detection devices, or motor control systems, for the same reason that wind in a junkyard does not assemble airplanes and radar.

Brains Are Composed Primarily of Adaptive Problem-Solving Devices

In fact, natural selection is the only known cause of and explanation for complex functional design in organic systems. Hence, all naturally occurring functional organization in organisms should be ascribed to its operation, and hypotheses about function are likely to be correct only if they are the kinds of functionality that natural selection produces.

This leads to the most important point for cognitive neuroscientists to abstract from modern evolutionary biology: Although not everything in the designs of organisms is the product of selection, all complex functional organization is. Indeed, selection can only account for functionality of a very narrow kind: approximately, design features organized to promote the reproduction of an individual and his or her relatives in ancestral environments (Williams, 1966; Dawkins, 1986). Fortunately for the modern theory of evolution, the only naturally occurring complex functionality that ever has been documented in undomesticated plants, animals, or other organisms is functionality of just this kind, along with its derivatives and by-products.

This has several important implications for cognitive neuroscientists:

1. Technical definition of function In explaining or exploring the reliably developing organization of a cognitive device, the *function* of a design refers solely to how it systematically caused its own propagation in ancestral environments. It does not validly refer to any intuitive or folk definitions of function such as "contributing to personal goals," "contributing to one's well-being," or "contributing to society." These other

kinds of usefulness may or may not exist as side effects of a given evolved design, but they can play no role in explaining how such designs came into existence or why they have the organization that they do.

It is important to bear in mind that the evolutionary standard of functionality is entirely independent of any ordinary human standard of desirability, social value, morality, or health (Cosmides and Tooby, in press).

2. Adapted to the past The human brain, to the extent that it is organized to do anything functional at all, is organized to construct information, make decisions, and generate behavior that would have tended to promote inclusive fitness in the ancestral environments and behavioral contexts of Pleistocene hunter-gatherers and before. (The preagricultural world of hunter-gatherers is the appropriate ancestral context because natural selection operates far too slowly to have built complex information-processing adaptations to the post-hunter-gatherer world of the last few thousand years.)

3. No evolved "reading modules" The problems that our cognitive devices are designed to solve do not reflect the problems that our modern life experiences lead us to see as normal, such as reading, driving cars, working for large organizations, reading insurance forms, learning the oboe, or playing Go. Instead, they are the odd and seemingly esoteric problems that our hunter-gatherer ancestors encountered generation after generation over hominid evolution. These include such problems as foraging, kin recognition, "mind reading" (i.e., inferring beliefs, desires, and intentions from behavior), engaging in social exchange, avoiding incest, choosing mates, interpreting threats, recognizing emotions, caring for children, regulating immune function, and so on, as well as the already well-known problems involved in perception, language acquisition, and motor control.

4. Side effects are personally important but scientifically misleading Although our architectures may be capable of performing tasks that are "functional" in the (nonbiological) sense that we may value them (e.g., weaving, playing piano), these are incidental side effects of selection for our Pleistocene competencies—just as a machine built to be a hair-dryer can, incidentally, dehydrate fruit or electrocute. But it will be difficult to make sense of our cognitive mechanisms if one attempts to interpret them as devices designed to perform functions that were not selectively important for our hunter-gatherer ancestors, or if one fails to consider the adaptive functions these abilities are side effects of.

5. Adaptationism provides new techniques and principles Whenever one finds better-than-chance functional organization built into our cognitive or neural architecture, one is looking at adaptations—devices that acquired their distinctive organization from natural selection acting on our hunter-gatherer or more distant primate ancestors. Reciprocally, when one is searching for intelligible functional organization underlying a set of cognitive or neural phenomena, one is far more likely to discover it by

using an adaptationist framework for organizing observations because adaptive organization is the only kind of functional organization that is there to be found.

Because the reliably developing mechanisms (i.e., circuits, modules, functionally isolable units, mental organs, or computational devices) that cognitive neuroscientists study are evolved adaptations, all the biological principles that apply to adaptations apply to cognitive devices. This connects cognitive neuroscience and evolutionary biology in the most direct possible way. This conclusion should be a welcome one because it is the logical doorway through which a very extensive body of new expertise and principles can be made to apply to cognitive neuroscience, stringently constraining the range of valid hypotheses about the functions and structures of cognitive mechanisms. Because cognitive neuroscientists are usually studying adaptations and their effects, they can supplement their present research methods with carefully derived adaptationist analytic tools.

6. **Ruling out and ruling in** Evolutionary biology gives specific and rigorous content to the concept of function, imposing strict rules on its use (Williams, 1966; Dawkins, 1982, 1986). This allows one to rule out certain hypotheses about the proposed function of a given cognitive mechanism. But the problem is not just that cognitive neuroscientists sometimes impute functions that they ought not to. An even larger problem is that many fail to impute functions that they ought to. For example, an otherwise excellent recent talk by a prominent cognitive neuroscientist began with the claim that one would not expect jealousy to be a ''primary'' emotion—that is, a universal, reliably developing part of the human neural architecture (in contrast to others, such as disgust or fear). Yet there is a large body of theory in evolutionary biology—sexual selection theory—that predicts that sexual jealousy will be widespread in species with substantial parental investment in offspring (particularly in males); behavioral ecologists have documented mate-guarding behavior (behavior designed to keep sexual competitors away from one's mate) in a wide variety of species, including various birds, fish, insects, and mammals (Krebs and Davies, 1997; Wilson and Daly, 1992); male sexual jealousy exists in every documented human culture (Daly et al., 1982; Wilson and Daly, 1992); it is the major cause of spousal homicides (Daly and Wilson, 1988), and in experimental settings, the design features of sexual jealousy have been shown to differ between the sexes in ways that reflect the different adaptive problems faced by ancestral men and women (Buss, 1994). From the standpoint of evolutionary biology and behavioral ecology, the hypothesis that sexual jealousy is a primary emotion—more specifically, the hypothesis that the human brain includes neurocognitive mechanisms whose function is to regulate the conditions under which sexual jealousy is expressed and what its cognitive and behavioral manifestations will be like—is virtually inescapable (for an evolutionary/cognitive approach to emotions, see Tooby and Cosmides, 1990a, 1990b). But if cognitive neuroscientists are not aware of this body of theory and evidence, they will not design experiments capable of revealing such mechanisms.

7. Biological parsimony, not physics parsimony The standard of parsimony imported from physics, the traditional philosophy of science, or from habits of economical programming is inappropriate and misleading in biology, and hence, in neuroscience and cognitive science, which study biological systems. The evolutionary process never starts with a clean work board, has no foresight, and incorporates new features solely on the basis of whether they lead to systematically enhanced propagation. Indeed, when one examines the brain, one sees an amazingly heterogeneous physical structure. A correct theory of evolved cognitive functions should be no less complex and heterogeneous than the evolved physical structure itself and should map on to the heterogeneous set of recurring adaptive tasks faced by hominid foragers over evolutionary time. Theories of engineered machinery involve theories of the subcomponents. One would not expect that a general, unified theory of robot or automotive mechanism could be accurate.

8. Many cognitive adaptations Indeed, analyses of the adaptive problems humans and other animals must have regularly solved over evolutionary time suggest that the mind contains a far greater number of functional specializations than is traditionally supposed, even by cognitive scientists sympathetic to "modular" approaches. From an evolutionary perspective, the human cognitive architecture is far more likely to resemble a confederation of hundreds or thousands of functionally dedicated computers, designed to solve problems endemic to the Pleistocene, than it is to resemble a single general purpose computer equipped with a small number of domain-general procedures, such as association formation, categorization, or production rule formation (for discussion, see Cosmides and Tooby, 1987, 1994; Gallistel, 1990; Pinker, 1997; Sperber, 1994; Symons, 1987; Tooby and Cosmides, 1992.).

9. Cognitive descriptions are necessary Understanding the neural organization of the brain depends on understanding the functional organization of its computational relationships or cognitive devices. The brain originally came into existence and accumulated its particular set of design features only because these features functionally contributed to the organism's propagation. This contribution—that is, the evolutionary function of the brain—is obviously the adaptive regulation of behavior and physiology *on the basis of information* derived from the body and from the environment. The brain performs no significant mechanical, metabolic, or chemical service for the organism—its function is purely informational, computational, and regulatory in nature. Because the function of the brain is informational in nature, its precise functional organization can only be accurately described in a language that is capable of expressing its information functions—that is, in cognitive terms, rather than in cellular, anatomical, or chemical terms. Cognitive investigations are not some soft, optional activity that goes on only until the "real" neural analysis can be performed. Instead, the mapping of the computational adaptations of the brain is an unavoidable and indispensable step in the neuroscience research enterprise. It must proceed in tandem

with neural investigations and provides one of the primary frameworks necessary for organizing the body of neuroscience results.

The reason is straightforward. Natural selection retained neural structures on the basis of their ability to create adaptively organized relationships between information and behavior (e.g., the sight of a predator activates inference procedures that cause the organism to hide or flee) or between information and physiology (e.g., the sight of a predator increases the organism's heart rate, in preparation for flight). Thus, it is the information-processing structure of the human psychological architecture that has been functionally organized by natural selection, and the neural structures and processes have been organized insofar as they physically realize this cognitive organization. Brains exist and have the structure that they do because of the computational requirements imposed by selection on our ancestors. The adaptive structure of our computational devices provides a skeleton around which a modern understanding of our neural architecture should be constructed.

Brain Architectures Consist of Adaptations, By-products, and Random Effects

To understand the human (or any living species') computational or neural architecture is a problem in reverse engineering: We have working exemplars of the design in front of us, but we need to organize our observations of these exemplars into a systematic functional and causal description of the design. One can describe and decompose brains into properties according to any of an infinite set of alternative systems, and hence there are an indefinitely large number of cognitive and neural phenomena that could be defined and measured. However, describing and investigating the architecture in terms of its adaptations is a useful place to begin, because (1) the adaptations are the cause of the system's organization (the reason for the system's existence), (2) organisms, properly described, consist largely of collections of adaptations (evolved problem-solvers), (3) an adaptationist frame of reference allows cognitive neuroscientists to apply to their research problems the formidable array of knowledge that evolutionary biologists have accumulated about adaptations, (4) all of the complex functionally organized subsystems in the architecture are adaptations, and (5) such a frame of reference permits the construction of economical and principled models of the important features of the system, in which the wealth of varied phenomena fall into intelligible, functional, and predictable patterns. As Ernst Mayr put it, summarizing the historical record, "the adaptationist question, 'What is the function of a given structure or organ?' has been for centuries the basis for every advance in physiology" (Mayr, 1983, p. 32). It should prove no less productive for cognitive neuroscientists. Indeed, all of the inherited design features of organisms can be partitioned into three categories: (1) adaptations (often, although not always, complex); (2) the by products or concomitants of adaptations; and (3) random effects. Chance and selection, the two components of the evolutionary process, explain different types of design properties in organisms, and all aspects of design must be attributed to one of these two forces. The conspicuously distinctive cumulative impacts of chance and selection allow the development of rigorous standards of evidence for recognizing and establishing the existence of adaptations and distinguishing them from the nonadaptive aspects of organisms caused by the nonselectionist mechanisms of evolutionary change (Williams, 1966, 1985; Pinker and Bloom, 1992; Symons, 1992; Thornhill, 1991; Tooby and Cosmides, 1990a, 1990b, 1992; Dawkins, 1986).

Design Evidence

Adaptations are systems of properties (''mechanisms'') crafted by natural selection to solve the specific problems posed by the regularities of the physical, chemical, developmental, ecological, demographic, social, and informational environments encountered by ancestral populations during the course of a species' or population's evolution (table 9.1). Adaptations are recognizable by ''evidence of special design'' (Williams, 1966)—that is, by recognizing certain features of the evolved species-typical design of an organism ''as components of some special problem-solving machinery'' (Williams, 1985, p. 1). Moreover, they are so well organized and such good engineering solutions to adaptive problems that a chance coordination between problem and solution is effectively ruled out as a counter-hypothesis. Standards for recognizing special design include whether the problem solved by the structure is an evolutionarily longstanding adaptive problem, and such factors as economy, efficiency, complexity, precision, specialization, and reliability, which, like a key fitting a lock, render the design too good a solution to a defined adaptive problem to be coincidence (Williams, 1966). Like most other methods of empirical hypothesis testing, the demonstration that something is an adaptation is always, at core, a probability assessment concerning how likely a set of events is to have arisen by chance alone. Such assessments are made by investigating whether there is a highly nonrandom coordination between the recurring properties of the phenotype and the structured properties of the adaptive problem, in a way that meshed to promote fitness (genetic propagation) in ancestral environments (Tooby and Cosmides, 1990b, 1992). For example, the lens, pupil, iris, retina, visual cortex, and other parts of the eye are too well coordinated, both with each other and with features of the world, such as the properties of light, optics, geometry, and the reflectant properties of surfaces, to have co-occurred by chance. In short, like the functional aspects of any other engineered system, they are recognizable as adaptations for analyzing scenes from reflected light by their organized and functional relationships to the rest of the design and to the structure of the world.

In contrast, concomitants or by products of adaptations are those properties of the phenotype that do not contribute to functional design per se, but that happen to be coupled to properties that are. Consequently, they were dragged along into the

Table 9.1
The formal properties of an adaptation

An adaptation is:

1. A cross-generationally recurring set of characteristics of the phenotype
2. that is reliably manufactured over the developmental life history of the organism,
3. according to instructions contained in its genetic specification,
4. in interaction with stable and recurring features of the environment (i.e., it reliably develops normally when exposed to normal ontogenetic environments),
5. whose genetic basis became established and organized in the species (or population) over evolutionary time, because
6. the set of characteristics systematically interacted with stable and recurring features of the ancestral environment (the ''adaptive problem''),
7. in a way that systematically promoted the propagation of the genetic basis of the set of characteristics better than the alternative designs existing in the population during the period of selection. This promotion virtually always takes place through enhancing the reproduction of the individual bearing the set of characteristics, or the reproduction of the relatives of that individual.

Adaptations. The most fundamental analytic tool for organizing observations about a species' functional architecture is the definition of an adaptation. To function, adaptations must evolve such that their causal properties rely on and exploit these stable and enduring statistical structural regularities in the world, and in other parts of the organism. Things worth noticing include the fact that an adaptation (such as teeth or breasts) can develop at any time during the life cycle, and need not be present at birth; an adaptation can express itself differently in different environments (e.g., speaks English, speaks Tagalog); an adaptation is not just any individually beneficial trait, but one built over evolutionary time and expressed in many individuals; an adaptation may not be producing functional outcomes currently (e.g., agoraphobia), but only needed to function well in ancestral environments; finally, an adaptation (like every other aspect of the phenotype) is the product of gene-environment interaction. Unlike many other phenotypic properties, however, it is the result of the interaction of the species-standard set of genes with those aspects of the environment that were present and relevant during the species' evolution. For a more extensive definition of the concept of adaptation, see Tooby and Cosmides, 1990b, 1992.

species-typical architecture because of selection for the functional design features to which they are linked. For example, bones are adaptations, but the fact that they are white is an incidental by-product. Bones were selected to include calcium because it conferred hardness and rigidity to the structure (and was dietarily available), and it simply happens that alkaline earth metals appear white in many compounds, including the insoluble calcium salts that are a constituent of bone. From the point of view of functional design, by-products are the result of ''chance,'' in the sense that the process that led to their incorporation into the design was blind to their consequences (assuming that they were not negative). Accordingly, such by-products are distinguishable from adaptations by the fact that they are not complexly arranged to have improbably functional consequences (e.g., the whiteness of bone does nothing for the vertebrae).

In general, by-products will be far less informative as a focus of study than adaptations because they are consequences and not causes of the organization of the system (and hence are functionally arbitrary, unregulated, and may, for example, vary capriciously between individuals). Unfortunately, unless researchers actively seek to study organisms in terms of their adaptations, they usually end up measuring and investigating arbitrary and random admixtures of functional and functionless aspects of organisms, a situation that hampers the discovery of the underlying organization of the biological system. We do not yet, for example, even know which exact aspects of the neuron are relevant to its function and which are by-products, so many computational neuroscientists may be using a model of the neuron that is wildly inaccurate.

Finally, entropic effects of many types are always acting to introduce disorder into the design of organisms. Traits introduced by accident or by evolutionary random walks are recognizable by the lack of coordination that they produce within the architecture or between the architecture and the environment, as well as by the fact that they frequently cause uncalibrated variation between individuals. Examples of such entropic processes include genetic mutation, recent change in ancestrally stable environmental features, and developmentally anomalous circumstances.

How Well-Engineered Are Adaptations?

The design of our cognitive and neural mechanisms should only reflect the structure of the adaptive problems that our ancestors faced to the extent that natural selection is an effective process. Is it one? How well or poorly engineered are adaptations? Some researchers have argued that evolution primarily produces inept designs, because selection does not produce perfect optimality (Gould and Lewontin, 1979). In fact, evolutionary biologists since Darwin have been well aware that selection does not produce perfect designs (Darwin, 1859; Williams, 1966; Dawkins, 1976, 1982, 1986; for a recent convert from the position that organisms are optimally designed to the more traditional adaptationist position, see Lewontin, 1967, 1979; see Dawkins, 1982, for an extensive discussion of the many processes that prevent selection from reaching perfect optimality). Still, because natural selection is a hill-climbing process that tends to choose the best of the variant designs that actually appear, and because of the immense numbers of alternatives that appear over the vast expanse of evolutionary time, natural selection tends to cause the accumulation of very well-engineered functional designs.

Empirical confirmation can be gained by comparing how well evolved devices and human engineered devices perform on evolutionarily recurrent adaptive problems (as opposed to arbitrary, artificial modern tasks, such as chess). For example, the claim that language competence is a simple and poorly engineered adaptation cannot be taken seriously, given the total amount of time, engineering, and genius that has gone into the still unsuccessful effort to produce artificial systems that can remotely approach—let alone equal—human speech perception, comprehension, acquisition, and production (Pinker and Bloom, 1992).

Even more strikingly, the visual system is composed of collections of cognitive adaptations that are well-engineered products of the evolutionary process, and although they may not be ''perfect'' or ''optimal''—however these somewhat vague concepts may be interpreted—they are far better at vision than any human-engineered system yet developed.

Wherever the standard of biological functionality can be clearly defined—semantic induction, object recognition, color constancy, echolocation, relevant problem-solving generalization, chemical recognition (olfaction), mimicry, scene analysis, chemical synthesis—evolved adaptations are at least as good as and usually strikingly better than human engineered systems, in those rare situations in which humans can build systems that can accomplish them at all. It seems reasonable to insist that before a system is criticized as being poorly designed, the critic ought to be able to construct a better alternative—a requirement, it need hardly be pointed out, that has never been met by anyone who has argued that adaptations are poorly designed. Thus, although adaptations are certainly suboptimal in some ultimate sense, it is an empirically demonstrable fact that the short-run constraints on selective optimization do not prevent the emergence of superlatively organized computational adaptations in brains. Indeed, aside from the exotic nature of the problems that the brain was designed to solve, it is exactly this sheer functional intricacy that makes our architecture so difficult to reverse-engineer and to understand.

Cognitive Adaptations Reflect the Structure of the Adaptive Problem and the Ancestral World

Four lessons emerge from the study of natural competences, such as vision and language: (1) most adaptive information-processing problems are complex; (2) the evolved solution to these problems is usually machinery that is well engineered for the task; (3) this machinery is usually specialized to fit the particular nature of the problem; and (4) its evolved design often embodies substantial and contentful ''innate knowledge'' about problem-relevant aspects of the world.

Well-studied adaptations overwhelmingly achieve their functional outcomes because they display an intricately engineered coordination between their specialized design features and the detailed structure of the task and task environment. Like a code that has been torn in two and given to separate couriers, the two halves (the structure of the mechanism and the structure of the task) must be put together to be understood. To function, adaptations evolve such that their causal properties rely on and exploit these stable and enduring statistical and structural regularities in the world. Thus, to map the structures of our cognitive devices, we need to understand the structures of the problems that they solve and the problem-relevant parts of the hunter-gatherer world. If studying face recognition mechanisms, one must study the recurrent structure of faces. If studying social cognition, one must study the recurrent structure of hunter-gatherer social life. For vision, the problems are not so very different for a modern scientist and a Pleistocene hunter-gatherer, so the folk notions of function that perception researchers use are not a problem. But the more one strays from low-level perception, the more one needs to know about human behavioral ecology and the structure of the ancestral world.

Experimenting with Ancestrally Valid Tasks and Stimuli

Although bringing cognitive neuroscience current with modern evolutionary biology offers many new research tools (Preuss, 1995; see also chapter 84), we have out of necessity limited discussion to only one: an evolutionary functionalist research strategy (see chapter 87 and Tooby and Cosmides, 1992, for a description; for examples, see chapters in Barkow et al., 1992; Daly and Wilson, 1995; Gaulin, 1995; and chapter 81). The adoption of such an approach will modify research practice in many ways. Perhaps most significantly, researchers will no longer have to operate purely by intuition or guesswork to know which kinds of tasks and stimuli to expose subjects to. Using knowledge from evolutionary biology, behavioral ecology, animal behavior, and hunter-gatherer studies, they can construct ancestrally or adaptively valid stimuli and tasks. These are stimuli that would have had adaptive significance in ancestral environments, and tasks that resemble (at least in some ways) the adaptive problems that our ancestors would have been selected to be able to solve.

The present widespread practice of using arbitrary stimuli of no adaptive significance (e.g., lists of random words, colored geometric shapes) or abstract experimental tasks of unknown relevance to Pleistocene life has sharply limited what researchers have observed and can observe about our evolved computational devices. This is because the adaptive specializations that are expected to constitute the majority of our neural architecture are designed to remain dormant until triggered by cues of the adaptively significant situations that they were designed to handle. The Wundtian and British Empiricist methodological assumption that complex stimuli, behaviors, representations, and competences are compounded out of simple ones has been empirically falsified in scores of cases (see, e.g., Gallistel, 1990), and so, restricting experimentation to such stimuli and tasks simply restricts what researchers can find to a highly impoverished and unrepresentative set of phenomena. In contrast, experimenters who use more biologically meaningful stimuli have had far better luck, as the collapse of behaviorism and its replacement by modern behavioral ecology have shown in the study of animal behavior. To take one example of its applicability to humans, effective mechanisms for Bayesian inference—undetected by 20 years of previous research using "modern" tasks and data formats—were activated by exposing subjects to information formatted in a way that hunter-gatherers would have encountered it (Brase et al., 1998; Cosmides and Tooby, 1996; Gigerenzer and Hoffrage, 1995). Equally, when subjects were given ancestrally valid social inference tasks (cheater detection, threat interpretation), previously unobserved adaptive reasoning specializations were activated, guiding subjects to act in accordance with evolutionarily predicted but otherwise odd patterns (Cosmides, 1989; Cosmides and Tooby, 1992; see also chapter 87).

Everyone accepts that one cannot study human language specializations by exposing subjects to meaningless sounds: the acoustic stimuli must contain the subtle, precise, high level relationships that make sound language. Similarly, to move on to the study of other complex cognitive devices, subjects should be exposed to stimuli that contain the subtle, ancestrally valid relationships relevant to the diverse functions of these devices. In such an expanded research program, experimental stimuli and tasks would involve constituents such as faces, smiles, disgust expressions, foods, the depiction of socially significant situations, sexual attractiveness, habitat quality cues, animals, navigational problems, cues of kinship, rage displays, cues of contagion, motivational cues, distressed children, species-typical "body language," rigid object mechanics, plants, predators, and other functional elements that would have been part of ancestral hunter-gatherer life. Investigations would look for functional subsystems that not only deal with such low-level and broadly functional competences as perception, attention, memory, and motor control, but also with higher-level ancestrally valid competences as well—mechanisms such as eye direction detectors (Baron-Cohen, 1994), face recognizers (e.g., Johnson and Morton, 1991), food memory subsystems (e.g., Hart et al., 1985; Caramazza and Shelton, 1998), person-specific memory, child care motivators (Daly and Wilson, 1995), and sexual jealousy modules.

Although these proposals to look for scores of content-sensitive circuits and domain-specific specializations will strike many as bizarre and even preposterous, they are well grounded in modern biology. We believe that in a decade or so they will look tame. If cognitive neuroscience is anything like investigations in domain-specific cognitive psychology (Hirschfeld and Gelman, 1994) and in modern animal behavior, researchers will be rewarded with the materialization of a rich array of functionally patterned phenomena that have not been observed so far because the mechanisms were never activated in the laboratory by exposure to ecologically appropriate stimuli. Although presently, the functions of most brain structures are largely unknown, pursuing such research directions may begin to populate the empty regions of our maps of the brain with circuit diagrams of discrete, functionally intelligible computational devices.

In short, because theories and principled systems of knowledge can function as organs of perception, the incorporation of a modern evolutionary framework into cognitive neuroscience may allow the community to detect ordered relationships in phenomena that otherwise seem too complex to be understood.

Conclusion

The aforementioned points indicate why cognitive neuroscience is pivotal to the progress of the brain sciences. There are an astronomical number of physical interactions and relationships in the brain, and blind empiricism rapidly drowns itself among the deluge of manic and enigmatic measurements. Through blind empiricism, one can equally drown at the cognitive level in a sea of irrelevant things that our computational devices can generate, from writing theology or dancing the mazurka to calling for the restoration of the Plantagenets to the throne of France. However, evolutionary biology, behavioral ecology, and hunter-gatherer studies can be used to identify and supply descriptions of the recurrent adaptive problems humans faced during their evolution. Supplemented with this knowledge, cognitive research techniques can abstract out of the welter of human cognitive performance a series of maps of the functional information-processing relationships that constitute our computational devices and that evolved to solve this particular set of problems: our cognitive architecture. These computational maps can then help us abstract out of the ocean of physical relationships in the brain that exact and minute subset that implements those information-processing relationships because it is only these relationships that explain the existence and functional organization of the system. The immense number of other physical relationships in the brain are incidental by-products of those narrow aspects that implement the functional computational architecture. Consequently, an adaptationist inventory and functional mapping of our cognitive devices can provide the essential theoretical guidance for neuroscientists that will allow them to home in on these narrow but meaningful aspects of neural organization and to distinguish them from the sea of irrelevant neural phenomena.

Acknowledgments

The authors gratefully acknowledge the financial support of the James S. McDonnell Foundation, the National Science Foundation (NSF grant BNS9157-449 to John Tooby), and a Research Across Disciplines grant (Evolution and the Social Mind) from the UCSB Office of Research.

Note

1. The genes underlying complex adaptations cannot vary substantially between individuals because if they did, the obligatory genetic shuffling that takes place during sexual reproduction

would break apart the complex adaptations that had existed in the parents when these are recombined in the offspring generation. All the genetic subcomponents necessary to build the complex adaptation rarely would reappear together in the same individual if they were not being supplied reliably by both parents in all matings (for a discussion of the genetics of sexual recombination, species-typical adaptive design, and individual differences, see Tooby, 1982; Tooby and Cosmides, 1990b).

References

Barkow, J., L. Cosmides, and J. Tooby, eds., 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Baron-Cohen, S., 1994. The eye-direction detector: A case for evolutionary psychology. In *Joint-Attention: Its Origins and Role in Development*, C. Moore and P. Dunham, eds. Hillsdale, N.J.: Erlbaum.

Brase, G., L. Cosmides, and J. Tooby, 1998. Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *J. Exp. Psychol. Gen.* 127: 3–21.

Buss, D., 1994. *The Evolution of Desire*. New York: Basic Books.

Caramazza, A., and J. Shelton, 1998. Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *J. Cogn. Neurosci.* 10: 1–34.

Cosmides, L., 1989. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition* 31: 187–276.

Cosmides, L., and J. Tooby, 1981. Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol.* 89: 83–129.

Cosmides, L., and J. Tooby, 1987. From evolution to behavior: Evolutionary psychology as the missing link. In *The Latest on the Best: Essays on Evolution and Optimality*, J. Dupre, ed. Cambridge: Mass.: MIT Press, pp. 277–306.

Cosmides, L., and J. Tooby, 1992. Cognitive adaptations for social exchange. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 163–228.

Cosmides, L., and J. Tooby, 1994. Beyond intuition and instinct blindness: The case for an evolutionarily rigorous cognitive science. *Cognition* 50: 41–77.

Cosmides, L., and J. Tooby, 1996. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition* 58: 1–73.

Cosmides, L., and J. Tooby, in press. Toward an evolutionary taxonomy of treatable conditions. *J. Abnorm. Psychol.*

Daly, M., and M. Wilson, 1984. *Sex, Evolution and Behavior*, Second Edition. Boston: Willard Grant.

Daly, M., and M. Wilson, 1988. *Homicide*. New York: Aldine.

Daly, M., and M. Wilson, 1995. Discriminative parental solicitude and the relevance of evolutionary models to the analysis of motivational systems. In *The Cognitive Neurosciences*, M. S. Gazzaniga, ed. Cambridge, Mass.: MIT Press, pp. 1269–1286.

Daly, M., M. Wilson, and S. J. Weghorst, 1982. Male sexual jealousy. *Ethol. Sociobiol.* 3: 11–27.

Darwin, C., 1859. *On the Origin of Species*. London, Murray. New edition: Cambridge, Mass.: Harvard University Press.

Dawkins, R., 1976. *The Selfish Gene*. New York: Oxford University Press.

Dawkins, R., 1982. *The Extended Phenotype*. San Francisco: W. H. Freeman.

Dawkins, R., 1986. *The Blind Watchmaker*. New York: Norton.

Gallistel, C. R., 1990. *The Organization of Learning*. Cambridge, Mass.: MIT Press.

Gaulin, S., 1995. Does evolutionary theory predict sex differences in the brain? In *The Cognitive Neurosciences*, M. S. Gazzaniga, ed. Cambridge, Mass.: MIT Press, pp. 1211–1225.

Gigerenzer, G., and U. Hoffrage, 1995. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychol. Rev.* 102: 684–704.

Gould, S. J., and R. C. Lewontin, 1979. The spandrels of San Marco and the Panglossian program: A critique of the adaptationist programme. *Proc. R. Soc. Lond.* 205: 281–288.

Haig, D., 1993. Genetic conflicts in human pregnancy. *Q. Rev. Biol.* 68: 495–532.

Hamilton, W. D., 1964. The genetical evolution of social behavior. *J. Theor. Biol.* 7: 1–52.

Hart, J. Jr., R. S. Berndt, and A. Caramazza, 1985. Category-specific naming deficit following cerebral infarction. *Nature* 316: 439–440.

Hirschfeld, L., and S. Gelman, eds. 1994. *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York Cambridge University Press.

Johnson, M., and J. Morton, 1991. *Biology and Cognitive Development: The Case of Face Recognition*. Oxford: Blackwell.

Krebs, J. R., and N. B. Davies, 1997. *Behavioural Ecology: An Evolutionary Approach*, 4th edition. London: Blackwell Science.

Lee, R. B., and I. DeVore, 1968. *Man the Hunter*. Chicago: Aldine.

Lewontin, R., 1967. Spoken remark in *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution*, P. Moorhead and M. Kaplan, eds. *Wistar Institute Symposium Monograph* 5: 79.

Lewontin, R., 1979. Sociobiology as an adaptationist program. *Behav. Sci.* 24: 5–14.

Marr, D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

Mayr, E., 1983. How to carry out the adaptationist program. *Am. Naturalist* 121: 324–334.

Pinker, S., 1994. *The Language Instinct.* New York: Morrow.

Pinker, S., 1997. *How the Mind Works.* New York: Norton.

Pinker, S., and P. Bloom, 1992. Natural language and natural selection. Reprinted in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York Oxford University Press, pp. 451–493.

Preuss, T., 1995. The argument from animals to humans in cognitive neuroscience. In *The Cognitive Neurosciences*, M. S. Gazzaniga, ed. Cambridge, Mass.: MIT Press, pp. 1227–1241.

Shepard, R. N., 1984. Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychol. Rev.* 91: 417–447.

Shepard, R. N., 1987a. Evolution of a mesh between principles of the mind and regularities of the world. In *The Latest on the Best: Essays on Evolution and Optimality*, J. Dupre, ed. Cambridge, Mass.: MIT Press, pp. 251–275.

Shepard, R. N., 1987b. Towards a universal law of generalization for psychological science. *Science* 237: 1317–1323.

Sperber, D., 1994. The modularity of thought and the epidemiology of representations. In *Mapping the Mind: Domain Specificity in Cognition and Culture*, L. Hirschfeld and S. Gelman, eds. New York: Cambridge University Press, pp. 39–67.

Symons, D., 1987. If we're all Darwinians, what's the fuss about? In *Sociobiology and Psychology*, C. B. Crawford, M. F. Smith, and D. L. Krebs, eds. Hillsdale, N.J.: Erlbaum, pp. 121–146.

Symons, D., 1992. On the use and misuse of Darwinism in the study of human behavior. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 137–159.

Thornhill, R., 1991. The study of adaptation. In *Interpretation and Explanation in the Study of Behavior*, M. Bekoff and D. Jamieson, eds. Boulder, Colo.: Westview Press.

Tooby, J., 1982. Pathogens, polymorphism, and the evolution of sex. *J. Theor. Biol.* 97: 557–576.

Tooby, J., and L. Cosmides, 1990a. The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethol. Sociobiol.* 11: 375–424.

Tooby, J., and L. Cosmides, 1990b. On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *J. Pers.* 58: 17–67.

Tooby, J., and L. Cosmides, 1992. The psychological foundations of culture. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York Oxford University Press, pp. 19–136.

Williams, G. C., 1966. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought.* Princeton, N.J.: Princeton University Press.

Williams, G. C., 1985. A defense of reductionism in evolutionary biology. *Oxford Surv. Biol.* 2: 1–27.

Williams, G. C., and D. C. Williams, 1957. Natural selection of individually harmful social adaptations among sibs with special reference to social insects. *Evolution* 17: 249–253.

Wilson, M., and M. Daly, 1992. The man who mistook his wife for a chattel. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, J. Barkow, L. Cosmides, and J. Tooby, eds. New York: Oxford University Press, pp. 289–322.

10 Evolutionary Psychology: A Critique

David J. Buller

What Is Evolutionary Psychology?

Some researchers define "evolutionary psychology" as simply "the evolutionary study of mind and behavior" (Caporael 2001, p. 608). So conceived, evolutionary psychology is a *field of inquiry*, defined not by any specific theories about human psychology, but only by a commitment to developing such theories within the framework of evolutionary biology. Other researchers claim that an evolutionary perspective on human behavior and mentality entails a number of specific doctrines regarding the nature and evolution of the human mind (Buss 1995; Cosmides and Tooby 1997; Pinker 1997; Symons 1992; Tooby and Cosmides 1992). For these researchers, evolutionary psychology is a *paradigm*, a tightly interwoven web of theoretical claims, methodological commitments, and empirical results. This paradigm, which I will call "Evolutionary Psychology" (capitalized) to distinguish it from the field of inquiry ("evolutionary psychology"), is the focus of this chapter.

The basic tenet of Evolutionary Psychology is that, just as evolution by natural selection has created human morphological adaptations, so it has created human psychological adaptations. Our psychological adaptations, however, are presumably complex traits, and the construction of complex adaptations typically requires hundreds of thousands of years of cumulative selection. Our ancestors spent the Pleistocene—the epoch spanning 1.8 million to 10,000 years ago—living in small hunter-gatherer groups, but only the past 10,000 years living as agriculturists and the past few hundred years living in industrial societies. Consequently, Evolutionary Psychologists argue, "it is improbable that our species evolved complex adaptations even to agriculture, let alone to postindustrial society" (Cosmides et al. 1992, p. 5). Rather, our psychological adaptations must have been designed during the Pleistocene to solve the adaptive

This chapter is a revised version of "Evolutionary Psychology," which appeared in M. Nani and M. Marraffa (eds.), *A Field Guide to the Philosophy of Mind* (http://host.uniroma3.it/progetti/kant/field/ep.htm), 2000.

problems faced by our hunter-gatherer ancestors. As Cosmides and Tooby say, "our modern skulls house a Stone Age mind" (1997, p. 85).

Adaptive problems are problems whose solutions enhance the ability to survive or reproduce. And the adaptive problems faced by our Pleistocene ancestors ranged from avoiding predators and inedible flora to acquiring mates and forming social alliances. Since these problems required very different behavioral solutions, Evolutionary Psychologists argue, a successful solution to one problem could not have transferred to another. So each adaptive problem would have selected for its own dedicated psychological mechanism. As Symons argues, "it is no more probable that some sort of general-purpose brain/mind mechanism could solve all the behavioral problems an organism faces (find food, choose a mate, select a habitat, etc.) than it is that some sort of general-purpose organ could perform all physiological functions (pump blood, digest food, nourish an embryo, etc.)" (1992, p. 142). Thus, Evolutionary Psychologists conclude, the human mind must be "organized into *modules* or mental organs, each with a specialized design that makes it an expert in one arena of interaction with the world. The modules' basic logic is specified by our genetic program. Their operation was shaped by natural selection to solve the problems of the hunting and gathering life led by our ancestors in most of our evolutionary history" (Pinker 1997, p. 21; emphasis added). Given the enormous number of adaptive problems our Pleistocene ancestors faced, Tooby and Cosmides estimate that the human mind consists of "hundreds or thousands" of such evolved modules (2000, p. 1171).

Modules have the following properties (Buss 1995; Cosmides and Tooby 1997; Tooby and Cosmides 1992). First, they are *domain specific*—that is, each module is dedicated to solving one problem or a restricted range of closely related problems. As such, their information-processing procedures are activated by, and are sensitive to, only information about a particular aspect of the world, in much the way that the ear is responsive only to specific vibratory frequencies. Second, each module comes equipped with substantial innate knowledge about its proprietary problem domain and a set of innate procedures for employing that knowledge to solve problems in its proprietary domain. Third, modules develop reliably and without formal instruction in every "normal" member of the species.

Since "selection usually tends to make complex adaptations universal" (Tooby and Cosmides 1992, p. 38), Evolutionary Psychologists argue that the network of evolved modules in the human mind are "the brain/mind mechanisms that collectively constitute human nature" (Symons 1992, p. 144). This universal human nature can, however, produce individual and cultural differences when modules encounter different developmental and occurrent inputs. As Tooby and Cosmides say, an Evolutionary Psychologist "observes variable manifest psychologies or behaviors between individuals and across cultures and views them as the product of a common, underlying evolved psychology operating under different circumstances" (Tooby and Cosmides 1992,

p. 45). This entails that ''individual differences, including heritable individual differences, are unlikely to represent differences in the presence or absence of complex adaptive mechanisms'' (Buss 1995, p. 11).

To summarize, the fundamental theoretical tenets of Evolutionary Psychology are these. First, the human mind consists of ''hundreds or thousands'' of ''genetically specified'' modules, or special-purpose computational devices, each of which is an adaptation for solving a specific adaptive problem. Second, the information-processing functions of modules are designed to solve the problems of survival and reproduction that were faced by our Pleistocene hunter-gatherer ancestors. And, third, evolved modules collectively constitute a universal human nature. In the sections to follow, I will argue that each of these tenets is mistaken. (See Buller 2005 for a comprehensive and detailed critique of Evolutionary Psychology.)

Modularity and the Adapted Mind

The principal argument for the claim that the mind consists of ''hundreds or thousands'' of ''genetically specified'' modules is this. First, our ancestors encountered a diverse array of adaptive problems, and each adaptive problem ''domain'' required its own ''domain-specific'' solution. Second, no single ''domain-general'' psychological mechanism could have successfully solved widely different adaptive problems. Therefore, a distinct psychological mechanism evolved for each distinct adaptive problem our ancestors faced.

The crucial step in this argument is clearly the second premise—the claim that no single ''domain-general'' mechanism could have generated solutions to multiple and varied adaptive problems. Cosmides and Tooby support this premise with the following argument: ''A woman who used the same taste preference mechanisms in choosing a mate that she used to choose nutritious foods would choose a very strange mate indeed, and such a design would rapidly select itself out'' (1994, p. 90). But this argument conflates *behavioral solutions* to problems (choosing a high-quality mate and choosing a nutritious food item) with the *psychological mechanisms* that produce behavior. And there is no reason to think that the same psychological mechanism couldn't produce diverse behaviors that solved distinct problems, in much the way that the same computer program can produce both spreadsheets and bar charts.

To illustrate this point, consider the domain-general ''mechanism'' of social learning, which involves observation of role models followed by imitation of the observed behavior of those models. Suppose a female employs social learning in figuring out how to select nutritious peaches: She observes her parents selecting plump and juicy peaches, and she does the same. If she now switches problem domains to the selection of a mate, the mechanism of social learning would clearly not guide her to search for a plump and juicy mate. Rather, it would guide her to observe and imitate the mate-selection behavior of female role models, and this would lead to the acquisition of mate-selection criteria that are specific to the problem domain of selecting a mate. So the domain-general mechanism of social learning would generate behavioral solutions specific to each problem domain in which it operated.

The point of this example is not to defend social-learning accounts of behavior, but to illustrate that domain-general learning mechanisms could operate on domain-specific inputs, and make use of information specific to those inputs, in order to generate domain-specific behavioral solutions to the problems they encounter. So the need for highly specific *behavioral solutions* to adaptive problems in our evolutionary history wouldn't necessarily have selected for *distinct mechanisms*. Thus, the principal argument for modularity rests on a false premise.

Although our species doesn't appear to have solved the adaptive problems it faced by evolving a massively modularized brain, it also doesn't appear to have solved those problems by evolving a brain consisting of just a few domain-general learning mechanisms (such as social learning). Indeed, the evidence indicates that the evolution of human intelligence was more complicated than either of these simple alternatives. For environmental complexity of the sort posed by the multiple and varied adaptive problems faced by our ancestors typically selects for *phenotypic plasticity* (Godfrey-Smith 1998; Sterelny 2003). Phenotypic plasticity is the capacity of a single genotype to produce more than one adaptive phenotype—more than one anatomical form, physiological state, or psychological mechanism—in response to environmental conditions. And research in developmental neurobiology has shown that mechanisms of neural development embody a plasticity that produces, through interaction with the local environment, brain structures that perform relatively specialized cognitive functions (Buller and Hardcastle 2000).

According to our best evidence to date, the brain structures that perform specialized cognitive functions—and that would have been involved in generating cognitive solutions to adaptive problems throughout our species' evolutionary history—develop through a process of diffuse proliferation of brain cells and connections followed by a "pruning" that shapes this diffuse connectivity into relatively specialized structures. That is, functionally specialized brain structures are produced by a process consisting of both "additive" events (the formation and migration of brain cells and the formation of neural connections) and "subtractive" events (the pruning of synapses through cell death and axonal retraction) (Elman et al. 1996). In this process, gene-directed protein synthesis is involved in the additive events that build the diffuse connectivity with which brain development begins. The subtractive events, however, are not under genetic control. Rather, the subtractive events occur through cell competition, whereby cells with the strongest patterns of innervation (primarily from sensory inputs) retain their connections and the others die. Thus, genes specify the proteins involved in the additive events during brain development, but the forms and functions of brain structures are then shaped by environmental inputs. So the specialized brain structures we have are primarily environmentally induced, not "genetically specified."

Our species may, nonetheless, have been faced with recurrent adaptive problems throughout its evolutionary history, and human brains may have recurrently produced information-processing solutions to these problems. But contrary to Evolutionary Psychology, distinct "genetically specified" modules were not required to solve these recurrent adaptive problems. In designing the human brain, selection hit upon a different solution: a plasticity that allows particular environmental demands to participate heavily in tailoring the responses to those very demands. This developmental plasticity, which forms functionally specialized circuits in response to demands of the local environment, is a *domain-general* mechanism with respect to behavioral response. But its function is to produce more highly specialized mechanisms, which in turn produce behavioral solutions specific to the problem domains that have been involved in shaping them. This developmental process can produce relatively stable brain structures that specialize primarily in particular information-processing tasks. But, the extent to which "modular" outcomes of human brain development have been regular throughout some of our evolutionary history is due to the fact that developmentally plastic human brains have encountered recurrent environmental demands throughout that history, not to "genetic specification" of the outcomes.

There are two morals to be drawn from this. First, the cognitively specialized brain structures that are the outcome of brain development have not been shaped by natural selection. For natural selection does not retain environmentally induced phenotypic characteristics of organisms; that would be Lamarckian evolution. Instead, natural selection retains only those genes that have fitness-enhancing effects on an organism's morphology. And, as we have seen, genes do not guide the subtractive process that shapes specialized brain structures. Consequently, the specialized structures in an adult human's brain are not the product of hundreds of thousands of years of cumulative selection for incremental, genetically-induced modifications to the human brain; they are, instead, the product of that individual's history of interaction with the local environment.

Second, it is a mistake to seek adaptation among the *products* of brain development—that is, among the relatively special-purpose brain structures that emerge during the course of brain development. Those products are highly plastic responses to environmental inputs. The human cognitive adaptation is, instead, the *process* that generates those special-purpose brain structures (Deacon 1997). That is, the brain's *plasticity* is the adaptation, and the contingently stable brain structures in an adult's brain are by-products of that adaptation's functioning in its local environment. Similarly, the antibody-assembly process in the immune system is an adaptation, but the particular antibodies it produces are not. The specific antibodies in an adult's immune system (many of which way not have been present in our ancestors) are, instead, the result of a plastic system's interaction with the pathogenic environment. In both cases, a particular trait (a psychological mechanism or an antibody) is present in an individual because of how the local environment has acted on a mechanism of plasticity. An adaptation, however, is a trait that is present in an individual because that individual inherited "genes for" that trait from ancestors in whom those genes were selected for. Thus, the relatively "modular," yet developmentally reversible, structures in an adult brain don't have the right etiology to be biological adaptations.

Evolutionary Psychologists frequently support their modular view of the mind by arguing that the only alternative is the view that "all adult mental organization and content is . . . cultural in derivation and substance" (Tooby and Cosmides 1992, p. 115). And they argue persuasively that this alternative is implausible. But there is a middle ground, which is favored by the evidence about brain development.

Rather than consisting of a plethora of "genetically specified" modules, the "innate" structure of the brain consists in a comparatively small number of *learning biases*, which take the form of a heightened responsiveness to certain classes of stimuli (Elman et al. 1996; Karmiloff-Smith 1992). These learning biases increase the probability that interaction with the environment will eventually produce domain-specialized structures, but there is no isomorphism between the "innate" learning biases and the knowledge databases embodied in the eventually resulting structures. Rather, development proceeds by a process of gradually branching domain specificity (or problem specialization), and the initial learning biases pertain only to the first and most general domains in this process (Karmiloff-Smith 1992). For example, a relatively large chunk of an adult human brain is devoted to face recognition, but infants appear to preferentially attend to *any* stimulus consisting of three high-contrast blobs configured like the two eyes and the mouth of a face, and they show no preference for actual faces over blobs. There is a face-recognition learning bias, then, that takes the form of heightened responsiveness to three high-contrast blobs, but full-blown face recognition is the result of a gradual process of progressive specialization (Elman et al. 1996, pp. 115–118). There is no direct mapping from the very minimally specified "innate" learning bias to the complex knowledge structure embodied in a mature face-recognition "module." A brain that develops in this way will mimic one that possesses a plethora of "innate" modules, even though its "innate" structure is relatively minimal.

According to this alternative picture, human psychological adaptation does not consist in "hundreds or thousands" of "genetically specified" modules. Rather, the fundamental adaptation is the brain's developmental plasticity, which is capable of producing a wide variety of problem-specialized information-processing structures that are responsive to local conditions. Additional adaptations lie in the minimal learning biases instantiated in the early stages of brain development.

Detecting Cheating in the Evidence for Modularity

But Evolutionary Psychologists have presented *empirical evidence*, not simply arguments, for their modular view of the mind. In particular, Cosmides (1989) claims that the modular view of the mind predicts the existence of a "cheater-detection module," and she claims to have gathered strong empirical evidence of this module. If the modular view of the mind has, indeed, accurately predicted such a discovery, it has a lot going for it, despite the foregoing arguments. In fact, however, there is no good evidence of a cheater-detection module.

Precisely why reciprocal altruism—the mutual exchange of fitness benefits—has evolved in our lineage (and others) remains a much-debated question. But it is widely agreed that, once individuals evolve altruistic propensities, selection favors the evolution of *cheaters*, nonreciprocators who accept the fitness benefit of another's altruistic act without paying the fitness cost of providing a benefit in return. The evolution of cheaters, in turn, selects for the ability to detect and avoid cheaters. Accordingly, Cosmides (1989) argues that the human mind should be innately equipped with a *cheater-detection module*, a special-purpose psychological adaptation for detecting cheaters in social exchanges, which evolved to save us the fitness costs of being exploited.

Evidence for Cosmides' hypothesis derives from studies with the Wason selection task. In Wason selection tasks, subjects are given a conditional, *if P, then Q*, together with four two-sided cards displaying information of the form *P, not-P, Q*, and *not-Q*, and they are instructed to turn over those cards necessary to determine whether the conditional is true. The logically correct solution is to turn over the cards displaying *P* and *not-Q* in order to see whether their other sides contain *not-Q* and *P* respectively, since these, and only these, cards can falsify the conditional. Two results from studies with the Wason selection task are taken as evidence of a cheater-detection module.

First, there appears to be a *content effect* in the selection task: The frequency with which subjects select the "logically correct" cards appears to vary as a function of what the conditionals *are about*. For example, when presented with the conditional "*If a card has an 'R' on one side, then it has a '2' on the other side*" and cards showing *R, J, 2*, and *8*, an average of only 4 percent of subjects choose the *R* and *8* cards (the *P* and *not-Q* cards), and 79 percent choose either the *R* card alone or the *R* and *2* (*P* and *Q*) cards (Evans 1982, pp. 157–159). In contrast, when presented with the conditional "*If a person is drinking beer, then that person must be over 19 years of age*" and cards showing *drinking beer, drinking Coke, 22 years old*, and *16 years old*, 73 percent of subjects choose the *drinking beer* and *16 years old* cards (the *P* and *not-Q* cards), while only 20 percent choose *drinking beer* alone and none choose both *drinking beer* and *22 years old* (Griggs and Cox 1982). As these examples illustrate, differential performance on Wason selection tasks is due primarily to an increase in the frequency with which the *not-Q* card is selected.

Since all conditionals in selection tasks apparently have the same logical form, the performance differential seems to indicate that subjects are reasoning about the *content*, not the logical form, of the conditionals. In particular, since the above drinking-age problem represents a *social exchange*, a situation in which an obligation is incurred in order to receive a benefit, Cosmides (1989) argues that it activates a cheater-detection module, which looks for violations of the conditional rule (*drinking beer* while being *sixteen years old*), whereas the abstract letter-number problem falls on deaf modules, which didn't evolve to solve abstract non-adaptive problems. Cosmides supported this analysis by presenting subjects with artificial abstract and social-exchange problems and finding the same "content effect" biased in favor of improved performance on the social-exchange problems.

Second, when the logically correct cards differ from those representing cheating, subjects appear to ignore logic and choose the cards that represent cheating. For example, Cosmides (1989) gave two groups the following instructions: "You are a member of an island culture in which men get facial tattoos upon getting married. The island has a native plant called 'cassava root,' an aphrodisiac that makes men who eat it irresistible to women. Since sex between unmarried people is taboo on the island, the island's elders have enacted the following rule...." She then gave one group a "standard" social-contract conditional "*If a man eats cassava root, then he must have a tattoo on his face*" and the other group the "switched" conditional "*If a man has a tattoo on his face, then he eats cassava root.*" Both groups were shown the same four cards: *eats cassava root, eats molo nuts, tattoo*, and *no tattoo*. She found that 75 percent of subjects chose *eats cassava root* and *no tattoo* (the *P* and *not-Q* cards) in response to the "standard" social-contract conditional, but that 67 percent chose the same cards in response to the "switched" conditional, despite their being the logically incorrect *not-P* and *Q* cards in that version of the problem. Thus, Cosmides concluded, subjects don't apply logical principles in solving selection tasks, but simply focus on whether someone has accepted a benefit without fulfilling an obligation. And this, she claims, is evidence of cognitive specialization for detecting cheaters in social exchanges.

But neither result is good evidence of a cheater-detection module. Indeed, both results are compatible with a non-modular mind that applies domain-general logical principles in solving Wason selection tasks.

First, there is no genuine content effect in Wason selection tasks. The idea that there is presupposes that the conditionals in selection tasks have the same logical form and differ only in their contents (Over 2003). But there are distinct kinds of conditional, each with unique logical properties (Edgington 1995). In particular, the conditional in the letter-number problem is an *indicative conditional*, which makes the truth of one proposition conditional upon the truth of another, whereas the conditional in the drinking-age problem is a *deontic conditional*, which makes *an obligation* conditional upon the truth of a proposition. Indeed, all of the problems on which Cosmides found that subjects do best involved deontic conditionals. And deontic conditionals actually impose obligations *categorically* in their Q parts, while indicating on whom the obligations fall in their *P* parts (Fodor 2000). Since deontic conditionals actually require Q (of those of whom *P* is true), attention is immediately drawn to the *not-Q* card, and the frequency with which it's selected increases accordingly. Thus, differential performance on Wason selection tasks actually demonstrates a *logic effect*. Subjects apply different logical principles to indicative and deontic conditionals (as they would apply different logical principles to conditionals and conjunctions), and select *not-Q* with greater frequency in response to deontic conditionals because the correctness of *not-Q* is made more perspicuous by the logic of deontics. Moreover, differential performance on selection tasks is an artifact of pairing *arbitrary* indicative conditionals with deontic conditionals. Several studies have found that, when subjects are given "real-world," rather than arbitrary, indicatives, the frequency of logically correct responses equals that for deontics (Cheng and Holyoak 1989; Manktelow and Over 1990; Sperber et al. 1995).

Second, the idea that subjects ignore the logic of conditionals in order to focus on whether cheating has occurred falsely presupposes that changing the wording of conditionals in selection tasks changes how subjects *mentally represent* their logical forms. In Cosmides' "switched" problem, for example, subjects were presented with a background story that clearly required *a facial tattoo* of those *eating cassava root*, but were then asked to evaluate compliance to the conditional rule "*If a man has a tattoo on his face, then he eats cassava root,*" which not only didn't make sense in the context of the background story, but didn't contain the obligating word "must," which was present in the "standard" social-contract version of the problem (cf. Cosmides 1989, p. 217). Under such circumstances, language-comprehension mechanisms process the conditional *together with the background information* and output a mental representation of the logical form of the conditional that makes sense given the background information (as we all do, for example, when we represent the logic of the expression "all is not lost" as "not all is lost"). Thus, subjects didn't select the "logically incorrect" cards in Cosmides' "switched" problem; they selected the logically correct cards relative to their *representation* of the (deontic) logical form of the conditional. In other words, subjects simply applied logic to a sensible interpretation of the problem. (Similar arguments apply to results obtained by Gigerenzer and Hug (1992) and Fiddick et al. (2000).) So "logically incorrect" results in Wason selection tasks fail to show that subjects don't reason using general logical rules. There is, then, no good evidence of a cheater-detection mechanism.

"Our Modern Skulls House a Stone Age Mind"

Evolutionary Psychologists offer the following single argument in support of the claim that each human psychological adaptation is adapted to Pleistocene conditions. The

10,000 years since the end of the Pleistocene, they argue, "is only a small stretch in evolutionary terms, less than 1% of the two million years our ancestors spent as Pleistocene hunter-gatherers. For this reason, it is unlikely that new complex designs—ones requiring the coordinated assembly of many novel, functionally integrated features—could evolve in so few generations" (Cosmides et al. 1992, p. 5). Thus, Evolutionary Psychologists conclude, our psychological adaptations must be adapted to the Pleistocene conditions under which they evolved.

This argument, however, commits a couple of simple fallacies. First, the issue is not whether "*new* complex designs" that require the "coordinated assembly" of many features could have emerged in the 10,000 years, or 400 generations, since the Pleistocene. Without doubt, selection could not build a human mind from scratch in a mere 400 generations. But, from the fact that a "new complex design" could not have evolved since the Pleistocene, it doesn't follow that *old* complex designs, which evolved during the Pleistocene or even earlier, could not have been significantly *modified* by selection in 400 generations. Since the argument doesn't address this possibility, it fails to show that the psychological adaptations of contemporary humans must be identical to those of our Pleistocene ancestors.

Second, the argument assumes that the 1 percent of human evolutionary history since the Pleistocene is unimportant in comparison to the 99 percent spent as hunter-gatherers. But, as Wilson says, "it makes no sense to express evolutionary time as a proportion of the species' history" (1994, p. 226). It doesn't matter whether a lineage spends only 1 percent of its evolutionary history in a new environment, Wilson argues; what matters is what kinds of change occur during that 1 percent of its evolutionary history. Thus, "rather than marvelling at the antiquity of our species, we should be asking what kinds of evolutionary change can be expected in 10, 100, or 1000 generations" (p. 226).

There is, in fact, ample evidence that Evolutionary Psychologists greatly underestimate the evolutionary change that may have occurred since the end of the Pleistocene. In considering such change, we need to address two questions. First, have the environments inhabited by human populations since the Pleistocene changed in ways that have altered the selection pressures on human psychology? Second, if so, has there been sufficient time for an evolutionary response to these changes?

The answer to the first question is undoubtedly yes, and this is due largely to environmental changes produced by human activity. The agricultural and industrial revolutions, for example, precipitated fundamental changes in the social structures of human populations, which in turn altered the selection pressures on a variety of interpersonal behaviors. For example, while Pleistocene humans lived in groups of 50–150 individuals, post-agricultural humans have lived in increasingly larger groups, which has affected the challenges humans face when mating, forming alliances, or negotiating status hierarchies. In addition, changing social structures have wrought radical changes in the kinds of task that must be performed to acquire the resources necessary for successful child rearing. Even if hunter-gatherers had evolved ''Darwinian algorithms'' to solve the problems involved in acquiring resources necessary for child rearing in savanna environments, such Darwinian algorithms would be useless in the world of Wage-Laborer Man, since the tasks leading to acquisition of food and other resources have changed so drastically. Thus, environmental change since the Pleistocene has assuredly created strong selection pressure for psychological evolution.

But has there been sufficient time since the Pleistocene for an evolutionary response to these environmental changes? The question is *not* whether there has been enough time for human populations to evolve minds that are adapted to twenty-first-century environments. The question, instead, is whether there has been enough time for *modification* of the psychological adaptations possessed by our Pleistocene ancestors. And there are two reasons for answering in the affirmative.

First, there are clear cases of post-Pleistocene adaptive evolution in physiological and morphological traits. For example, ''the persistent domestication of cattle, and the associated dairying activities, did alter the selective environments of some human populations for sufficient generations to select for genes that today confer greater adult lactose tolerance'' (Laland et al. 2000, p. 132). The evolution of lactose tolerance was driven by *niche construction*, a process in which a population actively modifies the niche it inhabits, thereby modifying the selection pressures driving its own evolution. Niche construction typically accelerates the pace of evolution as successive generations of a population continually modify the sources of selection acting on themselves and subsequent generations. And humans have been supreme niche constructors. The development of agriculture and industry greatly altered human niches, and developments in medicine have continually altered the toll of disease on survival and, as a consequence, opportunities to reproduce. Indeed, niche construction has pervaded nearly every aspect of human life in recent centuries, ranging from methods of shelter construction to methods of food preparation and preservation (think of pasteurization, for example), from methods of contraception to organized education. And there is no reason to think that niche construction has driven only physiological and morphological evolution. For, on the psychological side, techniques of teaching, whether skill- or information-based, have altered the cognitive niche in which humans develop, and the recent development of information technologies is radically altering the cognitive niche to which future generations will adapt.

Second, the idea that human psychological adaptations cannot have evolved since the end of the Pleistocene depends on a false assumption about the rate at which selection can alter traits in a population. Recent work has shown that evolution by natural selection can occur very rapidly. Reznick et al. (1997) split populations of guppies living in high-predation environments, leaving a part of each population in its high-predation environment and moving the other part to a low-predation environment.

They found that life-history traits of the transplanted guppies evolved significantly in a mere 18 generations. The descendants of the transplanted guppies matured to a larger size and achieved reproductive viability at a later age than the non-transplanted guppies, they produced fewer litters (with fewer and larger offspring in each litter), and they allocated less of their total resources to reproduction during their early reproductive lives. Moreover, Reznick and his colleagues identified both the genetic basis of this change and the mechanism by which selection drove it (namely, differential mortality by predation). If this much evolution can occur in 18 generations, the 400 human generations since the end of the Pleistocene has certainly been sufficient time for selection-driven evolution in human psychological traits.

Thus, it is overwhelmingly likely that there has been some adaptive psychological evolution since the end of the Pleistocene, which has rendered contemporary humans psychologically different from their Pleistocene ancestors. There is no reason to think that "our modern skulls house a Stone Age mind."

"Human Nature"

One of the most obvious things in the world is that people differ in their attitudes, preferences, and behavioral responses to similar situations. This is true not only of individuals from different cultures, but of individuals within the same culture. According to Evolutionary Psychologists, however, "variable manifest psychologies or behaviors between individuals and across cultures" are "the product of a common, underlying evolved psychology operating under different circumstances" (Tooby and Cosmides 1992, p. 45). This doctrine relies heavily on the claim that variation exists only among the *outputs* of our psychological adaptations as a function of variation in the inputs to invariant psychological adaptations. Manifest psychological variation, according to Evolutionary Psychologists, is never a function of variation in the underlying psychological mechanisms of "normal" human beings. Thus, according to Evolutionary Psychologists, there are no *psychological polymorphisms*—that is, there are no alternative forms of psychological adaptations that are maintained by selection's acting on underlying genetic differences between individuals.

There are two arguments offered in support of this view. The first argument is as follows (see Tooby and Cosmides 1992, pp. 78–79). "Complex adaptations are intricate machines . . . that require coordinated gene expression, involving hundreds or thousands of genes to regulate their development" (p. 78). Since sexual reproduction is a process in which random halves of each parent's genes are "recombined" to form the genome of a zygote, if parents differed in any of their complex adaptations, randomly recombining the genes for those adaptations would make it highly improbable that offspring would receive all the genes necessary to build any of the adaptations. Consequently, if individuals differed in their complex adaptations, no adaptation could be reliably reproduced across generations. ''Therefore, it follows that humans, and other complex, long-lived, outbreeding organisms, must be very nearly uniform in those genes that underlie our complex adaptations'' and that this genetic uniformity ''tends to impose near uniformity at the functional level in complex adaptive designs'' (p. 79). It follows that no psychological differences result from genetic polymorphisms maintained by selection, since such polymorphisms would constitute alternative adaptive designs. Thus, Evolutionary Psychologists conclude, there must be ''a universal and uniform human nature'' (p. 79). To the extent that genetic differences influence psychological differences, Evolutionary Psychologists argue, they can affect only ''quantitative variation'' in qualitatively identical adaptations (in the way that genetic variation can affect the size of one's stomach, for example, but not whether one has a stomach).

Evolutionary Psychologists do, however, recognize an important exception to the universality of human nature. Since mating and reproduction pose different problems for the two sexes, selection has designed certain sex-specific suites of complex adaptations for solving these problems. With respect to phenomena such as mate choice, then, human nature bifurcates along the fault line of sex, with each sex possessing its own ''nature'' (Tooby and Cosmides 1992, pp. 81–82).

This argument, however, is multiply problematic. First, as Wilson (1994) points out, if the argument were sound, there would be no genetic polymorphisms in any sexually reproducing species, but there are many well documented examples of such polymorphisms. For example, males of the marine isopod crustacean species *P. sculpta* come in small, medium, and large, and these sizes perfectly correlate with distinct mating strategies (Shuster and Wade 1991). Large males secure and ''guard'' harems of females in the recesses of sponges, acquiring their copulations with the harem members. Small males acquire copulations by ''sneaking'' past inattentive large males and thereby gaining access to the females in the harem. Medium males morphologically resemble females, so they ''mimic'' the female courtship display and thereby gain entry to a large male's harem, where the medium male then copulates with the females inside. These three ''adaptive designs'' have equal reproductive success, and the genes underlying them have been identified. Yet such a genetic polymorphism constitutes a clear violation of the assumptions at play in Tooby and Cosmides' argument. So the argument is mistaken.

Second, the argument mistakenly assumes that selection acts only on qualitative variation and that, as long as individuals are ''qualitatively identical,'' quantitative differences are selectively irrelevant. But this assumption is false. Indeed, sexual dimorphism, which Evolutionary Psychologists take to be a ''qualitative'' difference, is actually the result of very ancient selection on quantitative differences in gamete size. This form of selection, called *disruptive selection*, favored the two extremes of gamete size (favoring large gametes for the nutrients they could store and small ones for their motility in reaching the larger gametes), while selecting against medium-sized gametes. So observable quantitative variation may actually be acted on and maintained by selection; and, when it is, it is a genetic polymorphism.

Third, the argument mistakenly assumes that, since adaptations require hundreds or thousands of genes for their development, if individuals differ in some adaptation, they must differ with respect to hundreds or thousands of genes (which the argument purports to show is impossible). As we have seen, however, Evolutionary Psychologists consider *male* and *female* to be distinct adaptive designs. Yet this adaptive difference is a product of a *single gene* difference, the *SRY* gene on the *Y* chromosome, which codes for testis-determining factor. Of course, *SRY* produces its effects only against a background of hundreds or thousands of genes shared by males and females; but the *differences in adaptations* result from a single gene difference against that background. As Wilson (1994) argues, there could be many other adaptational differences in humans that are likewise due to single gene differences.

Consequently, the argument fails to show that selection did not and does not produce and maintain some psychological polymorphisms in human populations. And this, in turn, means that the argument fails to show that selection must have created "a universal and uniform human nature."

The other argument for a universal human nature I call "the argument from *Gray's Anatomy*." As Tooby and Cosmides argue, "the fact that any given page out of *Gray's Anatomy* describes in precise anatomical detail individual humans from around the world demonstrates the pronounced monomorphism present in complex human physiological adaptations. Although we cannot yet directly 'see' psychological adaptations (except as described neuroanatomically), no less could be true of them" (1992, p. 38). Selection, in other words, has designed in humans a "universal architecture," in the sense that "everyone has two eyes, two hands, the same sets of organs, and so on" (Tooby and Cosmides 1992, p. 78). Since selection has presumably designed our minds as well as our bodies, we should expect selection to have designed a universal psychological architecture, which constitutes our "universal human nature." But there are problems even with this seemingly commonsensical argument.

First, as Wilson points out, "uniformity at the coarsest scale does not imply uniformity at finer scales" that are still selectively relevant (1994, p. 224). Every human may have a brain with two hemispheres, a cortex, an occipital lobe, and so on, but this doesn't imply universality of more micro-level psychological mechanisms. Since Evolutionary Psychologists claim that our universal psychological adaptations are modules, which are finer-grained brain structures, they need to demonstrate universality at this "finer scale." But the argument from *Gray's Anatomy* fails to do so.

Second, the "coarsest scale" to which Evolutionary Psychologists retreat in their argument from *Gray's Anatomy* is incommensurate with their definition of human nature as consisting of ''qualities that define us as a unique species'' (Buss 1999, p. 47). For the universals appealed to in this argument (two hands, two eyes, a stomach, skin) characterize the whole primate order and sometimes the whole class of mammals and even all vertebrates. So the analogical appeal to this ''coarsest scale'' of uniformity supports no conclusion about universal psychological adaptations that purportedly ''define us as a unique species.''

Third, the basic structural plan that typifies the ''universal architecture'' of a species—and that, at ever coarser scales of description, typifies the body plan of an order, class, and subphylum—consists primarily of features that have *persisted* down lineages and through speciations for tens to hundreds of millions of years. Thus, while selection probably played a role in designing the structural plan of humans, it didn't design that structural plan during human history, but during the history of the common ancestor of humans and other primates or vertebrates. So we can't really infer anything about psychological adaptations, which purportedly resulted from selection during relatively recent human history, from the fact that all humans (except the ''abnormal'') have two eyes, two hands, one nose, and a mouth.

Of course, this doesn't mean that there are *no* psychological universals of the sort that might interest Evolutionary Psychologists. It just means that there are no a priori considerations definitively showing that there are such universals awaiting discovery by Evolutionary Psychologists. It also means that, insofar as Evolutionary Psychology takes psychological adaptation as its object of inquiry, it must be prepared to investigate psychological variation just as studiously as any potential psychological universality. In other words, evolved ''human nature'' isn't constituted solely by psychological universals, but is at least partially constituted by adaptive variation.

This, however, prompts some questions. What if there are psychological universals? What should we make of them? As we saw earlier, Evolutionary Psychologists are mistaken in thinking that the totality of human psychological adaptations reflects adaptation to Pleistocene environments. Selection is undoubtedly continuing to modify trait frequencies in human populations. That means that any psychological universals we might happen to discover are temporally bound. They characterize human populations at a given moment in evolutionary history, and they are subject to change. Today's universals may be possessed by only a fraction of the species, or even extinguished, tomorrow. Thus, as Hull (1989) argues, it is a mistake to think that any universals we might discover reveal to us the ''nature'' of our species, in any interesting sense of ''nature.''

Moreover, if there are psychological universals, at least some of them will be the result of genetic drift, rather than selection (since some portion of all fixated traits are due to drift). Of course, Evolutionary Psychologists argue that drift-fixated traits are not typically incorporated into the ''complex organized design'' of the organism (Tooby and Cosmides 1992, p. 52). For this reason, they don't take drift-fixated psychological traits to be part of "human nature." But nothing in orthodox neo-Darwinian evolutionary theory justifies treating adaptations as somehow more "central" than drift-fixated traits, as somehow a part of the core "nature" of a species in a way that drift-fixated traits are not. To privilege adaptations in this way is to adopt what Godfrey-Smith (1999) calls a form of *natural theology*: It is to replace God with Natural Selection as the Creator, but to still maintain that the Creator's "intention" (as manifested in what is selected-for) represents the "nature" of a species, departure from which is "abnormal." But this particular way of wielding evolutionary theory is not intrinsic to evolutionary theory; it is an unjustified addition to it. Consequently, there is no *evolutionary* justification for the adaptation-centered concept of "human nature" employed by Evolutionary Psychology.

Conclusion

There can be little doubt that evolution has occurred and that *Homo sapiens* is among its products. There can also be little doubt that the evolutionary history of our lineage has left its mark on human psychology just as assuredly as it has left its mark on human morphology. The human mind, unquestionably, is the product of evolution. But what follows from this fact? I have argued not only that *none* of the central tenets of Evolutionary Psychology follow from this fact, but that all of the theoretical tenets of this paradigm are either misguided or unsubstantiated. This does not imply, however, that evolutionary psychology (as a *field of inquiry*) is bankrupt. It implies, rather, that Evolutionary Psychology is not the paradigm that will guide evolutionary psychology toward a more adequate evolutionary understanding of human psychology.

References

Buller, David J. (2005). *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge, MA: MIT Press.

Buller, David J., and Valerie Gray Hardcastle (2000). Evolutionary Psychology, Meet Developmental Neurobiology: Against Promiscuous Modularity. *Brain and Mind* 1: 307–325.

Buss, David M. (1995). Evolutionary Psychology: A New Paradigm for Psychological Science. *Psychological Inquiry* 6: 1–30.

Buss, David M. (1999). *Evolutionary Psychology: The New Science of the Mind*. Boston: Allyn and Bacon.

Caporael, Linnda R. (2001). Evolutionary Psychology: Toward a Unifying Theory and a Hybrid Science. *Annual Review of Psychology* 52: 607–628.

Cheng, Patricia W., and Keith J. Holyoak (1989). On the Natural Selection of Reasoning Theories. *Cognition* 33: 285–313.

Cosmides, Leda (1989). The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task. *Cognition* 31: 187–276.

Cosmides, Leda, and John Tooby (1994). Origins of Domain Specificity: The Evolution of Functional Organization. In L. A. Hirschfeld and S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 85–116). New York: Cambridge University Press.

Cosmides, Leda, and John Tooby (1997). The Modular Nature of Human Intelligence. In A. B. Scheibel and J. W. Schopf (Eds.), *The Origin and Evolution of Intelligence* (pp. 71–101). Sudbury, MA: Jones and Bartlett.

Cosmides, Leda, John Tooby, and Jerome H. Barkow (1992). Evolutionary Psychology and Conceptual Integration. In J. H. Barkow, L. Cosmides and J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 3–15). New York: Oxford University Press.

Deacon, Terrence W. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York: W. W. Norton & Co.

Edgington, Dorothy (1995). On Conditionals. *Mind* 104: 235–329.

Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Evans, Jonathan (1982). *The Psychology of Deductive Reasoning*. London: Routledge & Kegan Paul.

Fiddick, Laurence, Leda Cosmides, and John Tooby (2000). No Interpretation without Representation: The Role of Domain-Specific Representations and Inferences in the Wason Selection Task. *Cognition* 77: 1–79.

Fodor, Jerry (2000). Why We Are So Good at Catching Cheaters. *Cognition* 75: 29–32.

Gigerenzer, Gerd, and Klaus Hug (1992). Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change. *Cognition* 43: 127–171.

Godfrey-Smith, Peter (1998). *Complexity and the Function of Mind in Nature*. New York: Cambridge University Press.

Godfrey-Smith, Peter (1999). Adaptationism and the Power of Selection. *Biology and Philosophy* 14: 181–194.

Griggs, Richard A., and James R. Cox (1982). The Elusive Thematic-Materials Effect in Wason's Selection Task. *British Journal of Psychology* 73: 407–420.

Hull, David L. (1989). On Human Nature. In *The Metaphysics of Evolution* (pp. 11–24). Albany, NY: SUNY Press.

Karmiloff-Smith, Annette (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Laland, Kevin N., John Odling-Smee, and Marcus W. Feldman (2000). Niche Construction, Biological Evolution, and Cultural Change. *Behavioral and Brain Sciences* 23: 131–175.

Manktelow, K. I., and D. E. Over (1990). Deontic Thought and the Selection Task. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, and G. Erdos (Eds.), *Lines of Thinking: Reflections on the Psychology of Thought: Representation, Reasoning, Analogy and Decision Making* (pp. 153–164). New York: Wiley.

Over, David E. (2003). From Massive Modularity to Metarepresentation: The Evolution of Higher Cognition. In D. E. Over (Ed.), *Evolution and the Psychology of Thinking: The Debate* (pp. 121–144). Hove: Taylor and Francis.

Pinker, Steven (1997). *How the Mind Works*. New York: W. W. Norton.

Reznick, David N., Frank H. Shaw, F. Helen Rodd, and Ruth G. Shaw (1997). Evaluation of the Rate of Evolution in Natural Populations of Guppies (*Poecilia Reticulata*). *Science* 275: 1934–1937.

Shuster, Stephen M., and Michael J. Wade (1991). Equal Mating Success among Male Reproductive Strategies in a Marine Isopod. *Nature* 350: 608–610.

Sperber, Dan, Francesco Cara, and Vittorio Girotto (1995). Relevance Theory Explains the Selection Task. *Cognition* 57: 31–95.

Sterelny, Kim (2003). *Thought in a Hostile World: The Evolution of Human Cognition*. Malden, MA: Blackwell.

Symons, Donald (1992). On the Use and Misuse of Darwinism in the Study of Human Behavior. In J. H. Barkow, L. Cosmides and J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 137–159). New York: Oxford University Press.

Tooby, John, and Leda Cosmides (1992). The Psychological Foundations of Culture. In J. H. Barkow, L. Cosmides, and J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 19–136). New York: Oxford University Press.

Tooby, John, and Leda Cosmides (2000). Toward Mapping the Evolved Functional Organization of the Mind and Brain. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (second ed., pp. 1167–1178). Cambridge, MA: MIT Press.

Wilson, David Sloan (1994). Adaptive Genetic Variation and Human Evolutionary Psychology. *Ethology and Sociobiology* 15: 219–235.

VI Laws in Evolutionary Theory

11 The Evolutionary Contingency Thesis

John H. Beatty

1 Introduction

As Stephen Gould recently put it, evolution is like a videotape that, if replayed over and over, would have a different ending every time (Gould 1989, especially pp. 45–52, 277–291).[1] Last time it featured, among other things, duckbilled platypuses and human beings. And while we can be pretty sure that future replays will be every bit as humorous in their own special way, we cannot expect them to be quite as edifying. We are all, humans included, unlikely outcomes.

Gould contrasts such contingent "details" as ourselves with general "laws" that guide the course of evolution: "Invariant laws of nature . . . set the channels in which organic design must evolve." But, Gould emphasizes, "the channels are so broad relative to the details that fascinate us!" (Gould 1989, p. 289). In other words, laws of nature only *loosely* constrain the outcomes of evolution.

In this paper, I will further elaborate Gould's thesis and further defend it. My version of the thesis may appear at first to contradict Gould's, especially inasmuch as he emphasizes contingent "details," while I emphasize contingent *generalities*. Correspondingly, my version may appear stronger, though I believe he intends his to be every bit as strong.[2]

The thesis that I will defend, most briefly put, is this: all distinctively biological generalizations describe evolutionarily contingent states of nature—moreover, "highly" contingent states of nature in a sense that I will explain. This means that there are no laws of biology. For, whatever "laws" are, they are supposed to be more than just contingently true. To anticipate one obvious objection, I will also argue that there are no laws of evolution—the principles of evolutionary genetics are themselves evolutionarily contingent states of nature.

From Gereon Wolters and James G. Lennox, eds., *Concepts, Theories, and Rationality in the Biological Science* (University of Pittsburgh Press, 1995).

What I call the "evolutionary contingency thesis" is interesting in its own right. But it is also bound up with and bears upon a number of other issues in philosophy of biology, over and above the existence of laws of biology. Here, I will discuss its bearing on the explanatory ideals of biology, especially the "theoretical pluralism" so characteristic of biology, and also its bearing on the nature of controversy in biology, specifically the "relative significance" controversies that are so prevalent in the life sciences.

2 The Evolutionary Contingency Thesis and Laws of Biology

The evolutionary contingency thesis, somewhat more elaborately stated, is as follows:

All generalizations about the living world:

a) are just mathematical, physical, or chemical generalizations (or deductive consequences of mathematical, physical, or chemical generalizations plus initial conditions),

or

b) are distinctively biological, in which case they describe contingent outcomes of evolution.

The first part of this claim is meant to acknowledge that there are generalizations about the living world whose truth values are not a matter of evolutionary history. Evolution has not and will not result in any forms of life that are not subject to the laws of probability, or to Newton's laws of motion. Nor will evolution result in any carbon based forms of life that are not subject to the principles of organic chemistry. But while these sorts of principles are true of the living world, we do not call them "biological" principles.[3]

The second part of the evolutionary contingency thesis requires a lot more explanation than the first part. To begin, what is meant by the claim that all distinctively biological generalizations describe *evolutionary outcomes*? (After I discuss the sense in which they describe "evolutionary outcomes," I will then consider the sense in which they describe "contingent" evolutionary outcomes.)

By saying that biological generalizations describe evolutionary outcomes, I mean to refer to the *rule-making* capabilities of the agents of evolutionary change. All distinctively biological generalizations owe the extent of their generality to evolution by one or another, or some combination of, evolutionary agents, like directed and random mutation, hybridization, natural and sexual selection, random drift, etc. In this paper, I will focus on the rule-making (and later, rule-breaking) capabilities of evolution by random mutation and natural selection.

The rule-making capabilities of natural selection were of particular interest to the physicist-turned-biologist Max Delbrück, who characterized natural selection as "the overly faithful assistant of a credulous professor, the assistant being so anxious to please that he discards all those data which conflict with his master's theory" (Delbrück 1952, p. 12). In other words, generalizations emerge as certain traits are selected for, and as other traits—exceptions to the emerging rule—are selected against.

Consider for the purposes of illustration (and for fun) a very modest generalization from physical anthropology: "Humans are relatively hairless." That is, we have a lot less body hair than our closest ancestors. Why should this be the case? This turns out to be a highly contentious issue! But what is not disputed is that an evolutionary answer is called for. A number of different evolutionary accounts have been proposed, most of them based on natural selection.

For example, it has been argued that body hair was disadvantageous for early humans because hair harbors ticks, lice, and other insects that spread disease (Olson 1966). Others have argued that body hair was disadvantageous for early humans because of the warm climates in which they lived. That is, quadrupedal mammals need body hair in part to protect them from the sun' heat, but our early bipedal ancestors did not expose as much of their surface area to the sun; their body hair merely trapped heat (Wheeler 1985). Others argue that body hair was not so much disadvantageous as superfluous once humans evolved more effective means of dissipating heat, through the acquisition of many (2.5 million) sweat glands (Brace 1966; Campbell 1966). Darwin argued in *The Descent of Man* that the loss of body hair among humans was mainly due to sexual selection. In this respect, he urged, it is well to bear in mind the New Zealand proverb, "there is no woman for a hairy man" (and presumably vice-versa; Darwin 1871, vol. 2, p. 378).

On all these accounts, natural selection generates the relative hairlessness rule by eliminating its exceptions. But while "humans are relatively hairless" may be a slightly provocative principle, it is hardly fundamental, and so this is not a very telling case.

Consider another example, which also illustrates how *distinctively* biological generalizations describe evolutionary outcomes. This one has to do with the ubiquity of a particular metabolic pathway—the Krebs cycle—among aerobic organisms:

> In aerobic organisms, carbohydrate metabolism proceeds via a series of chemical reactions, including the eight steps of the Krebs cycle.

This generalization might at first seem so broad as to be just a matter of chemistry. When reformulated or drawn so as to exclude reference to aerobic organisms, and to include not only the substrates of the reactions (e.g., citrate or citric acid), but also the enzyme catalysts involved (shown in parentheses), the required reaction temperatures, the resulting generalization *is* just a chemical generalization (see figure 11.1).

But the ubiquity of the eight reactions of the Krebs cycle among aerobic organisms is not simply a matter of chemistry. It depends on genetically based aspects of those organisms, like their ability to synthesize the enzymes that facilitate the various reactions of the cycle. Genetically based traits are subject to changes in frequency due to

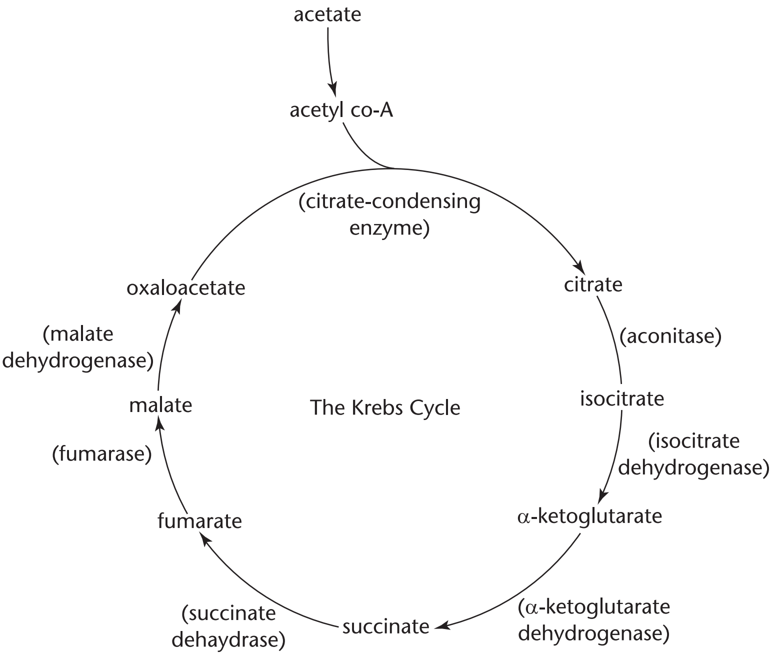


Figure 11.1

the various agents of evolutionary change, and so the prevalence of the Krebs cycle among aerobic organisms—and hence the truth of the generalization above—is a matter of evolutionary history. Having spent the early part of his career establishing the steps of the cycle, Hans Krebs spent the latter years of his life trying to give an account of its ubiquity in terms of evolution by natural selection (e.g., Krebs 1981; Baldwin and Krebs 1981). For instance, he argued that alternative ways of metabolizing acetic acid (the starting substrate of the cycle) are not as energy efficient.[4]

The case of the Krebs cycle is interesting not just because it is so fundamental, but also because, depending on how one generalizes about it, the resulting claim is either a chemical generalization whose truth value is not a matter of evolutionary history, or a distinctively biological generalization describing an evolutionary outcome.[5]

Consider one more example of the rule-making capabilities of evolution by natural selection. This example is intended to anticipate questions as to whether the generalizations of evolutionary biology themselves describe evolutionary outcomes. The example concerns Mendel's first ''law'' of inheritance, from which one of the central principles of evolutionary biology, the Hardy–Weinberg ''law'' is derived.

Mendel's first law concerns the way in which the genes of a sexual organism are partitioned (''segregate'') among the gametes it produces. The law states that, With respect to each pair of genes of a sexual organism, 50% of the organism's gametes will carry one representative of that pair, and 50% will carry the other representative of that pair.

Consider now that the process of gamete formation is itself a genetically based trait. There are genes that contribute to 50:50 segregation ratios, and genes that contribute to biased ratios (see, e.g., Sandler et al. 1968 and White 1973; see Magee 1987 for a review of the genetics of the process). Consider also that the degree of prevalence of any genetically based trait—50:50 segregation of genes included—is subject to agents of evolutionary change, like mutation and natural selection. The fact that Mendelian gamete production is so prevalent thus requires an evolutionary explanation—for example, some account of why 50:50 segregation ratios are adaptive. Indeed, this is currently a matter of considerable interest among evolutionary biologists (see the literature review by Bell 1982, pp. 438–442).

The fact that Mendel's law describes an evolutionary outcome is especially interesting because one of the central principles of evolutionary biology, the so-called Hardy–Weinberg "law" of gene frequency change, is a straightforward deductive consequence of Mendel's "law." Hence, the Hardy–Weinberg "law" of evolution itself describes an evolutionary outcome (see also Beatty 1981; 1982; Rosenberg 1985, pp. 132–136). As population geneticist Marcy Uyenoyama has so plainly put it, "Just as the meiotic mechanism [of gamete formation] directs evolution through its effects on the pattern of inheritance, the process of genetic transmission itself evolves by natural selection" (Uyenoyama 1987, p. 21).[6]

So much for the sense in which distinctively biological generalizations describe "evolutionary outcomes." What does it mean to say that such generalizations describe evolutionarily *contingent* states of affairs? This has to do with the *rule-breaking* capabilities of the agents of evolutionary change: the agents of evolution not only make, but also break the rules that pertain to the living world. More formally, to say that biological generalizations are evolutionarily contingent is to say that they are not laws of nature—they do not express any *natural necessity*; they may be true, but nothing in nature necessitates their truth.[7]

Admittedly, as Bas van Fraassen complains, philosophers of science are better at illustrating this distinction than explaining it (van Fraassen 1989, pp. 1–129).[8] A common sort of illustration (van Fraassen's own, p. 27; see also Hempel 1966, pp. 54–58) contrasts the following two generalizations:

1. All solid spheres of enriched uranium (U_{235}) have a diameter of less than one hundred meters.
2. All solid spheres of gold (Au) have a diameter of less than one hundred meters.

Suppose both claims are true. Still, there seems to be more to the former than the latter, by virtue of which we might accord the former but not the latter the status of law (or perhaps in this case it would be better to suggest that the first generalization describes an "instance" of a more general law). The critical mass of uranium excludes the possibility of such a large sphere of the substance. But nothing that we know about gold excludes the possibility of such a large sphere of that material. Now if on these grounds we accord to the former claim but not to the latter the status of law (or "instance" of a law), then we acknowledge that there is more to being a law than just being true. That something more has to do with what nature necessitates or precludes.[9]

There are at least two senses in which nature *fails to necessitate* the truth of biological generalizations. These correspond to two senses of evolutionary "contingency," one stronger than the other, although the weaker one is not absolutely weak. There are, in other words, at least two senses in which the agents of evolution can break rules as well as make them.

The weaker sense has to do with the fact that the conditions that lead to the evolutionary predominance of a particular trait within a particular group may change, so that the predominance of the trait declines. Somewhat more colloquially: what the agents of evolution render general, they may later render rare. Two sources of this kind of contingency are mutation, and natural selection in changing environments. Suppose that relative hairlessness owes its prevalence to the fact that it was favored under particular circumstances by natural selection—relative hairiness being selected against—selection acting, as Delbrück suggested, "like the overly faithful assistant" who "discards all those data which conflict with his master's theory." Is there anything *naturally necessary* about the circumstances under which relative hairlessness was favored—something that could not change? Is the professor really so single minded? And are the loyalties of the professor's assistant really so unswerving?[10]

In the case of generalizations about the Krebs cycle, and also in the case of Mendel's "law" and the derivative Hardy–Weinberg "law," we know that the assistant is *not* so loyal, or that the professor has not settled on one description of nature. There are so many exceptions. There are species of aerobic organisms that do not proceed through all the steps of the Krebs cycle, and that take different chemical routes from the same starting point to the same endpoint (e.g., organisms whose metabolism proceeds via the "glyoxylate shunt," which bypasses the two steps of the Krebs cycle leading from isocitrate to succinate; Gottschalk 1986, pp. 120–121; Moat and Foster 1988, pp. 27, 30, 136–139; Brock et al. 1984, pp. 142–143). As biochemist P. D. J. Weitzman reasons, one should *expect* to find such variations since,

> different organisms would be expected to make different demands on the several functions of the cycle and thus control the cycle in different ways in accordance with their individual metabolic "life-styles." It seems reasonable to assume that the evolutionary paths to different organisms have been accomplished by the evolution of distinctive regulatory and other individual functional features in the CAC [the Krebs cycle] (Weitzman 1985, pp. 253–254).

And there are many pairs of genes, in many species, that do not segregate in a 50:50 fashion; in these cases there is instead a marked bias in the production of gametes containing one rather than the other representative of the pair (e.g., Crow 1979). As Graham Bell (1982, p. 439) recently acknowledged, many of his fellow evolutionary biologists who contemplate such issues actually find it easier to imagine circumstances in which unequal segregation of alleles among gametes would prevail than to imagine the circumstances which would favor evolution by natural selection of 50:50 segregation ratios à la Mendel.[11]

My line of reasoning up to this point—especially as it applies to Mendel's ''law''—has been criticized by Elliott Sober (1987; 1989) and Marc Ereshefsky (1991, 1992). Their objection is that the argument rests on a confusion between the *falsity* of a generalization, and its *inapplicability*. As they rightly point out, Mendel's ''law'' can be recast in the conditional form, ''if *A*, then *B*.'' ''*B*'' is presumably 50:50 segregation of genes among gametes. ''*A*'' presumably describes some appropriately specified type of organism. This generalization would only be false if there were, in the past, present or future, organisms of type *A* that did not make gametes in the specified way. But just because some organisms do not make gametes in the specified way does not mean that the generalization is false, for those organisms may not be instances of *A*. The generalization is not falsified by these organisms, rather, the generalization does not apply to them. Of course, whether the generalization is false or just inapplicable depends on the formulation of *A*. The usual formulation of *A* has to do with being a sexual organism. If early sexual organisms did not all form gametes in the specified way, or if present sexual organisms do not all form gametes in this way, or if species of sexual organisms could evolve to the point where they do not all form gametes in this way, then Mendel's ''law'' *as usually formulated* is either straightforwardly false, or describes a contingent outcome of evolution and hardly a law of nature, in which case Sober's and Ereshefsky's objection fails.

Sober's and Ereshefsky's criticism is well taken only in connection with nonstandard formulations of Mendel's ''law'' (which they do not specify). But perhaps there are ways to modify the antecedent of the usual version so that it no longer just refers to sexual organisms, and so that it would be inapplicable rather than false in the past, present or future when 50:50 segregation of genes among gametes fails to obtain. It might then express a naturally necessary state of affairs. Perhaps all biological generalizations can be so formulated (or reformulated).

Of course, we have to be careful not to build into the antecedent *A* the conditions from which the conclusion *B* is logically and mathematically derivable. Laws *of nature* are supposed to be true as a matter of *empirical fact*, and not just as a matter of logic and mathematics.

Consider to this end a general strategy for rescuing the law-like status of biological generalizations, by qualifying them appropriately. Suppose we were to conditionalize generalizations about the prevalence of a trait upon the presence of the circumstances in which the trait would be expected to be favored by natural selection. These sorts of generalizations would correspond to what Sober calls ''source laws'' of natural selection, in the sense that such generalizations specify the sources of the selection pressures for the traits in question (Sober 1984, pp. 50–51, 58–59).[12]

So Mendel's ''law'' as usually formulated may not really be a law, but there may be a law to the effect that *under particular circumstances* sexual organisms will produce gametes in the way Mendel described, because under those circumstances sexual organisms that produced biased ratios of gametes will be selected against. (Similarly, the Hardy–Weinberg ''law'' could be reformulated so as to be conditional upon the occurrence of the same circumstances.) The new version of Mendel's ''law'' (and the new Hardy–Weinberg ''law'') will then be inapplicable, rather than false, when those circumstances do not obtain. To be sure, evolutionary biologists who work on the evolution of 50:50 segregation ratios have hypothesized circumstances that they think would favor that arrangement.[13]

The generalization about the ubiquity of the Krebs cycle already comes close to conditionalizing the prevalence of a trait upon the presence of the circumstances that favor the trait. By referring to aerobic organisms, the generalization at least implicitly refers to the oxygen environments in which the cycle plays its important respiratory role (in addition to its various biosynthetic roles), and in response to which important steps of the cycle evolved.[14]

In fact, there are in biology many such correlations of traits with the sorts of circumstances that favor their predominance. There is, for instance, the category of so called ''ecological rules'' like Bergmann's and Allen's rules. According to Bergmann's rule, given a species of warm-blooded vertebrates, those races of the species that live in cooler climates tend to be larger than those races of the species living in warmer climates. According to Allen's rule, given a species of warm-blooded vertebrates, those races of the species that live in cooler climates have shorter protruding body parts like bills, tails, and ears than those races of the species that live in warmer climates. One common understanding of the correlation between body size and cooler climate is that increased body size results (other things being equal) in decreased surface area, which slows heat dissipation, which becomes more and more advantageous as the climate becomes colder and colder. Basically, the same reasoning applies to Allen's rule (Mayr 1963, pp. 320–324).

The problem with such rules is that they are so riddled with exceptions, and for interesting reasons that have to do with the second, stronger sense in which all distinctively biological generalizations describe ''contingent'' states of affairs.[15] That stronger sense, which I intended earlier when I suggested that biological generalizations are *highly* contingent, has to do with the fact that evolution can lead to different outcomes from the same starting point, even when the same selection pressures are operating.

There are many sources of this second form of contingency. One is so-called ''chance'' or ''random'' mutation, meaning that the probability of occurrence of a mutation is in no way proportional to the advantage it confers. A second source is ''functional equivalence,'' meaning that there are very different ways of adapting to any one environment.

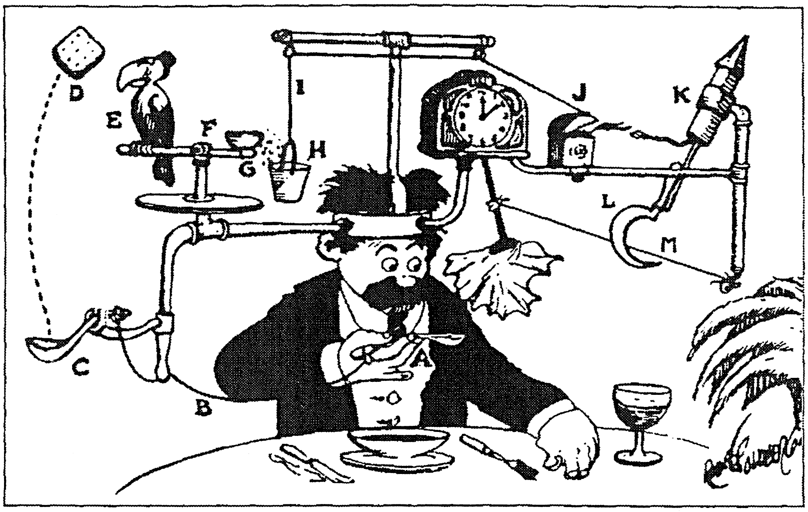
Darwin employed both sources of this stronger form of contingency in his account of *The Various Contrivances by which Orchids are Fertilized by Insects* (Darwin 1862, esp. pp. 282–293; see also Darwin 1872, vol. 1, pp. 241–244).[16] The different reproductive contraptions of orchids had evolved, Darwin believed, from a common form (the original orchid species), in response to a common problem (the need for cross fertilization), and at least originally under virtually the same circumstances (e.g., the same range of available insects). Sometimes this part of the flower had been modified to entice or trap insects, sometimes another part had been modified to do the job. Even when the same parts had been modified to do that job, they did it in very different ways. Among the various orchid species, presumably derived from one, Darwin thus conceived the evolution of reproductive mechanisms occurring over and over again with no generally determined outcome except cross-fertilization. And this was to be expected on the basis of chance variations and the possibility of functional equivalence. Selection acts on whatever opportunities present themselves, with never the same order of useful modifications arising, and with equally functional results.

Somewhat more colloquially, natural selection is like a Rube Goldberg ''tinkerer'' (Jacob 1982, pp. 25–46; see also Grene 1988). Suppose that Mr. Goldberg were faced more than once with exactly the same problem (see figure 11.2). Who would ever expect him to solve it in the same way, even starting with the same materials?

The same line of reasoning can be used to throw doubt upon the status of any supposed source ''law'' of natural selection. Consider again Bergmann's rule. Is decreased surface area the only way to limit heat dissipation? No. Heavier layers of fur or feathers will also do the job, as will behavioral innovations like burrowing. And it has long been argued that exceptions to Bergmann's rule are best explained in these and still other ways. The exceptional groups evolved different, functionally equivalent solutions to the problem of heat dissipation (see, e.g., Mayr 1956; 1963, p. 321).

Similarly, supposing that we can agree on the circumstances that favored the evolution of 50:50 segregation ratios, can we rule out the possibility of a functionally equivalent solution? And once we agree on the circumstances that favored the evolution of the Krebs cycle, will we be able to rule out other, functionally equivalent solutions? These are the sorts of questions that we will have to answer in order to know whether we can formulate anything like biological laws via the source-law strategy.[17]

But those looking for biological laws are not going to give up so easily! Another strategy has been proposed for qualifying biological generalizations in order to render them laws. This strategy has recently been suggested by Ereshefsky (1991), and is currently



...if a man were to make a machine for some special purpose, but were to use old wheels, springs, and pulleys, only slightly altered, the whole machine, with all its parts, might be said to be specially contrived for its present purpose. Thus throughout nature almost every part of each living being has probably served, in a slightly modified condition, for diverse purposes, and has acted in the living machinery of many ancient and distinct specific forms. (Darwin 1862, pp. 383–284)

Figure 11.2

... if a man were to make a machine for some special purpose, but were to use old wheels, springs, and pulleys, only slightly altered, the whole machine, with all its parts, might be said to be specially contrived for its present purpose. Thus throughout nature almost every part of each living being has probably served, in a slightly modified condition, for diverse purposes, and has acted in the living machinery of many ancient and distinct specific forms. (Darwin 1862, pp. 383–384) © 1979 by Rube Goldberg (used with permission)

being developed in detail by Kenneth Waters (ms.), again with special reference to Mendel's "law."[18] Consider that 50:50 segregation of genes among gametes in sexual organisms might be guaranteed by the presence of the appropriate genes—the genes supposedly responsible for the Mendelian mechanism of gamete formation—together with the appropriate environments for the expression of those genes. A generalization correlating the presence of those genes (and the appropriate environments) with the presence of the Mendelian mechanism might really be a law. Similarly, we might formulate a law by conditionalizing the presence of the Krebs cycle on the presence of the appropriate genes (and environments). Note that such generalizations are claims about the expression of particular genes, and as such are very different from the standard generalizations about Mendelian inheritance and the Krebs cycle.

Actually, David Hull once suggested that the best candidates for biological laws were precisely such generalizations about gene expression: as he schematized such laws, "Any organism with the genetic makeup G in any environment ranging from E_1 to E_n undergoing biochemical reactions R_1 through R_n will come to have phenotypic characters $C_1, C_2, \ldots, C_n$" (Hull 1974, p. 80).

For Waters, this gene-expression strategy is an instance of a more general strategy for locating laws within biology. This involves distinguishing between two different kinds of generalizations in biology: 1) evolutionarily contingent "trends," about the *distributions* of biological entities or properties (e.g., the distribution of 50:50 segregation among sexual organisms), and 2) non-evolutionarily contingent "laws," about the *dispositions* of biological entities (e.g., genes) to behave in particular ways (e.g., to result in particular phenotypes). The general strategy, then, is to seek generalizations of type 2, of which generalizations about gene expression are an instance. Waters focuses on gene expression examples.

I do not dispute the possibility of formulating laws by the gene-expression strategy, although that will be very difficult. I do question the possibility of formulating *distinctively biological* laws in this way. Consider first the difficulties. Not surprisingly, generalizations correlating genotypes and phenotypes *may* describe evolutionarily contingent outcomes. That is because, first of all, genes interact in the production of phenotypes; one and the same gene may have very different effects depending on what other genes it occurs in combination with. And second, evolutionary changes with respect to those "other" genes may affect the phenotypic expression of the gene or genes in question. For instance, many evolutionary biologists entertain the idea that the extent of a beneficial gene's dominance—the degree of phenotypic expression of that gene when it occurs in a single dose—evolves over time as natural selection favors the accumulation of other genes that enhance the expression of the gene in question (see e.g., Futuyma 1979, 374–376).

The problem of formulating any particular *law* of gene expression is therefore the problem of stating it in such a way that it would not be rendered false by further evolutionary change with respect to interacting genes. And that requires that we state a sufficiently inclusive set of genetic and environmental conditions. An appropriately conditionalized law of gene expression might then be rendered inapplicable by evolution—the conditions of the law no longer being met—but it might not be rendered false by evolutionary change.

What would such a law look like? It seems that one would be in a better and better position to know whether the set of conditions was sufficiently inclusive the more and more one knew about the chemical pathways leading from the sequence of nucleotide bases that make up the genes in question, to a physical-chemical specification of the phenotype in question. I suspect this is why Hull included the biochemical reactions linking genes to phenotypes in his schema of a law of gene expression.

Suppose that complete chemical pathways from genotypes to phenotypes could be formulated; and suppose that generalizations about these were indeed laws; still, I see no reason to regard them as *distinctively biological* laws. That is, I see no compelling reason to regard a description of a chain of chemical reactions—no matter that the reactants and products include DNA, RNA, and lots of enzymes—as "biological" generalizations. So to summarize this part of the discussion, the closer one's generalizations about gene expression come to describing sequences of chemical reactions, the more certain one can be that they are laws, because one can be more certain that no evolutionary outcomes can contradict them. But at the same time, the generalizations will become less and less distinctively biological.[19]

There are two other issues concerning the meaning and justification of the evolutionary contingency thesis that I would like to address at least briefly. First, by focusing on the contingency of *distinctively biological* generalizations, I do not mean to suggest that the situation is entirely different in physics and chemistry; in particular I do not mean to suggest that there are generalizations in physics and chemistry that are not contingent. I really do not know. Perhaps all true generalizations in physics and chemistry will also prove to be "evolutionarily" contingent, at least in the sense of "cosmologically" contingent. It is worth noting, however, that some physicists, like the particle physicist and cosmologist Steven Weinberg (1992), do indeed have faith that there will be discovered "a final theory," a most fundamental theory that explains everything else, all of cosmology included. Weinberg traces this ideal to Isaac Newton. I will have more to say shortly about this Newtonian ideal.

Second, by emphasizing *whether or not* biological generalizations are contingent, I do not mean to dismiss the possibility of different *degrees* of contingency manifested by different biological generalizations. The idea of degrees of contingency figures more prominently in the work of others. For example, Kenneth Schaffner recommends distinguishing between two degrees of contingency in biology: "essential" and "historical" accidents. "Historical accidentality," he explains, is not so very accidental:

> though initiating from a chance event [like random mutation, it] is augmented by additional nomic circumstances, such as strong natural selection. . . . Historical accidentality thus represents accidentality "frozen into" nomic universality. (Schaffner 1980, p. 90; 1993, p. 121.)

Schaffner follows Francis Crick (1968) in referring to the near (!) universality of the genetic code as a "frozen accident." Crick argued that once the code was established in a particular lineage, any change in the code would have had enormous, cascading effects, resulting in changes in the amino acid sequences of many, many previously adaptive proteins. Such a change could not conceivably be beneficial overall. Thus, as Schaffner explains, "though the origin of specific coding relations may have (likely) been due to a chance mutation . . . , at present the code is sufficiently entrenched by natural selection that it is only historically accidental (Schaffner 1980, p. 90; 1993, p. 121).[20]

Stuart Kauffman (1993) further multiplies the degrees of contingency by stressing the *extent* of evolutionary contingency throughout the biological world. For instance, he is concerned to determine 1) the *extent* to which generalizable features of biological systems are due to self-organizing properties of their constituent parts, 2) the *extent* to which selection may displace biological systems from the states they would occupy on the basis of the self-organizing properties of their parts alone, and 3) the *extent* to which the self-organizing properties of biological systems not only constrain evolution by natural selection but also in certain respects enable evolution by natural selection to take place.[21]

Schaffner and Kauffman and others are surely right to stress that there are *more or less* contingent generalities in biology. The present formulation of the evolutionary contingency thesis may be misleadingly simplistic in this regard.

3 Theoretical Pluralism and Relative Significance Controversies

I want to switch now from articulating the evolutionary contingency thesis, to applying it. In particular I want to consider its bearing on the explanatory ideals of biology, especially on the "theoretical pluralism" so characteristic of biology, and also on the nature of controversy in biology, specifically the "relative significance" controversies that are so prevalent in the life sciences.

"Theoretical pluralism" has to do with the number of theories or mechanisms that are believed to be required to account for a domain of phenomena (see also Beatty 1994). A proponent of theoretical pluralism with respect to a particular domain believes that the domain is essentially heterogeneous, in the sense that a plurality of theories or mechanisms is required to account for it, *different items in the domain requiring explanations in terms of different theories or mechanisms*. There is no single theory or mechanism—not even a single synthetic, multi-causal theory or mechanism—that will account for every item of the domain. This is not merely a matter of insufficient evidence for a single theory; rather, it is a matter of the evidence indicating that multiple accounts are required.[22]

Theoretical pluralism contributes to, and is reflected by, a certain kind of controversy—the so-called "relative significance" dispute. What is at issue in a relative significance dispute is the *extent of applicability* of a theory or mechanism within a domain—roughly, the proportion of items of the domain governed by the theory or mechanism—*not* whether the theory or mechanism in question is *the correct* account of the domain.

Examples of theoretical pluralism and relative significance controversies occur at every level of investigation in biology. As we have already discussed, biochemists raise questions about the extent of applicability of the Krebs cycle among aerobic organisms, suggesting that alternatives to the Krebs cycle are more prevalent than commonly acknowledged (Weitzman 1985, pp. 253–254). Geneticists raise questions about the ubiquity of the Mendelian mechanism of inheritance, arguing that non-Mendelian mechanisms are possibly very common (e.g., Crow 1979). Molecular geneticists wonder and argue about the extent of applicability of the lac operon (negative induction) model of gene regulation, as opposed to alternative negative repression, positive induction and repression, attenuation, etc. models (e.g., Lewin 1990, pp. 240–299; Yanofsky 1981, 1988). Immunologists and geneticists argue about the extent of applicability of alternative accounts of the generation of antibody diversity: e.g., germ-line vs. somatic cell theories (e.g., Kindt and Capra 1984). Physiologists, biophysicists, geneticists and evolutionary biologists working in the area of gerontology argue about the relative applicability of different mechanisms of aging: e.g., somatic mutations in dividing cells vs. "wear and tear" of post-mitotic cells (e.g., Comfort 1979; Maynard Smith 1966; Rose 1985; Finch 1990).

Evolutionary biology, ecology, biogeography, and systematics are rife with relative significance controversies. For instance, evolutionary biologists argue about whether selectionist theories have greater applicability to microevolutionary changes than neutralist theories (e.g., Lewontin 1974; Kimura 1983; Endler 1986; Gillespie 1991). They argue about whether gradualist, adaptationist theories of macroevolution have greater applicability than the punctuated equilibrium theory (e.g., Gould 1980; Lande 1980). They argue about the extent of applicability of the various mechanisms of the evolution of sex, from the "red-queen" hypothesis to the "tangled-bank" hypothesis to the "genetic-load" model to the "DNA-repair" model (e.g., Michod and Levin 1988; Kondrashov 1988). Evolutionary biologists and systematists argue about the extent of applicability of each of the multitude of theories of speciation, from each of the various forms of sympatric speciation, to parapatric speciation, to each of the various forms of allopatric speciation (e.g., Bush 1975; White 1978; Otte and Endler 1989). They also argue about the relative significance of different accounts of the relationship between ontogeny and phylogeny (e.g., the relative applicability of "von Baer's law"; see, e.g., Gould 1977). They argue about whether vicariance accounts of biogeographic patterns have greater applicability than dispersalist/center-of-origin accounts (e.g., Nelson and Platnick 1981a; 1981b).

Ecologists debate the extent of applicability of alternative theories of community structure, from competition theory, to predation and abiotic factor theories, to random colonization models (e.g., Schoener 1982; 1983; Connell 1983; Sih et al. 1989). Again, these are all disputes about the *extent* of applicability of alternative theories or mechanisms within a particular domain, not whether this or that account is *the* universally true one within that range.

I believe there are reasons to be a theoretical pluralist with respect to *every* domain of distinctively biological phenomena, and reasons to anticipate relative significance controversies within every domain. The main reason is that the contingencies of evolutionary history preclude the existence of laws of biology. It is not surprising that a biologist should be more interested in the *extent of applicability* of a theory within its intended domain than in its possible *universality* within that domain. Not expecting universal generalizations to hold within a domain, biologists expect instead to have recourse to a plurality of theories to cover it.[23]

The theoretical pluralism so prevalent in biology contrasts strikingly, I believe, with a traditional ideal, namely, to explain a domain of phenomena in terms of as few as possible different mechanisms, and best of all one single mechanism. This ideal was expressed particularly well by Newton, and so I will call it the Newtonian tradition. Newton elaborated it most succinctly in the first two of his three "rules of reasoning in philosophy" (Newton [1686] 1934, p. 398). According to the first rule, "We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances." As Newton clarified the rule, "To this purpose the philosophers say that Nature does nothing in vain, and more is in vain when less will serve; for Nature is pleased with simplicity and affects not the pomp of superfluous causes." Newton's second rule states my point more clearly: "Therefore [i.e., it follows from the first rule that] to the same natural effects we must, as far as possible, assign the same causes." As he proceeds to illustrate the rule: "As to respiration in a man and in a beast, the descent of stones in Europe and America, the light of our culinary fire and of the sun, the reflection of light in the earth and in the planets."

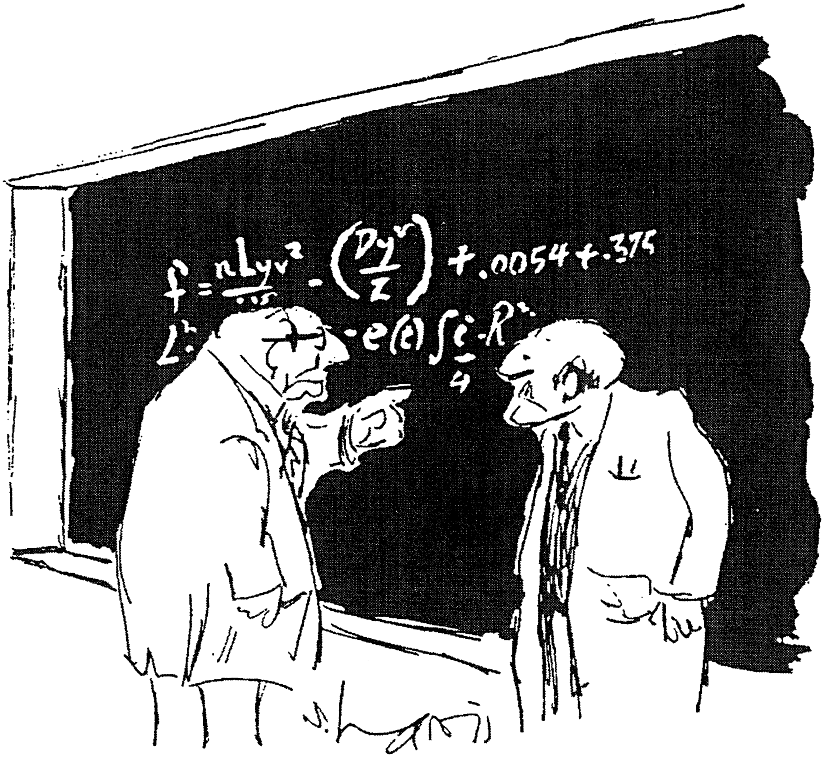
Judging by their acceptance of theoretical pluralism, and by their waging of relative significance disputes, many biologists seem not overly impressed by this rule of reasoning. Indeed, by their *promotion* of theoretical pluralism they seem to *repudiate* the Newtonian ideal.

For example, in their reviews of the modes of speciation, Bush (1975) and White (1978) staunchly defend a pluralistic approach against assumptions or attempts to show that there is a single correct account of the domain. As White insists,

> however much evolutionists of the future may synthesize in the field of speciation, we can be confident that the diversity of living organisms is such that their evolutionary mechanisms cannot be forced into the straightjacket of any narrow, universal dogma. (White 1978, p. 349)

This pluralism is also characteristic of the recent anthology and state-of-the-art summary, *Speciation and its Consequences*, edited by Otte and Endler (1989).

McIntosh (1987) recently summarized the trend toward theoretical pluralism in ecology, away from the ideals of the sixties and early seventies when ecologists like Robert MacArthur envisioned that all of ecology would ultimately be "embodied in a small number of simple laws." Recent anthologies, for example the anthology on community ecology edited by Diamond and Case (1986), proclaim pluralism in the preface and throughout. The editors explicitly distance themselves from the ideals of Newtonian mechanics:



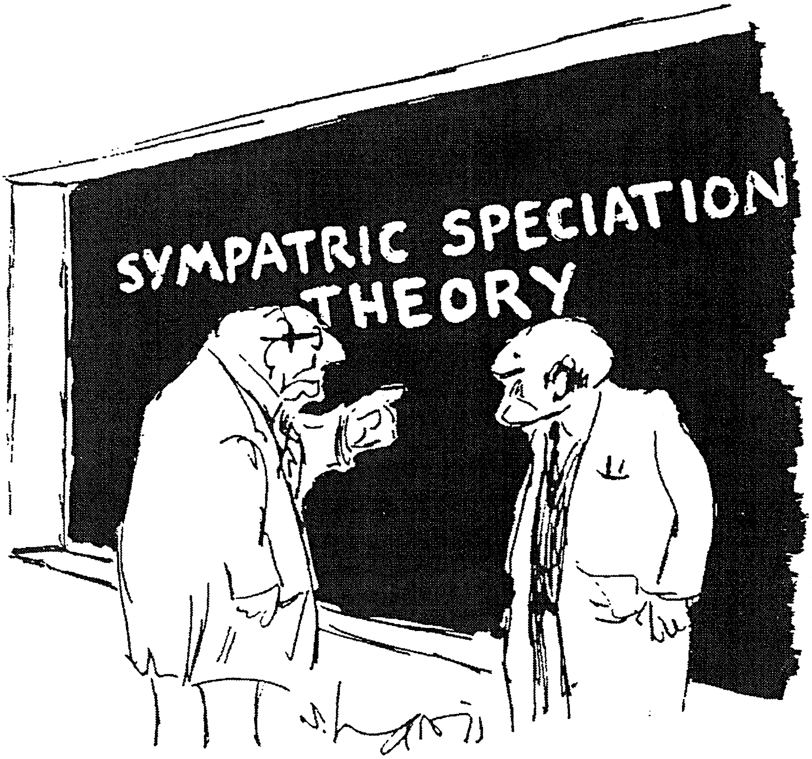
"Does this apply always, sometimes, or never?"

Figure 11.3
© 1977 by Sidney Harris (used with permission)

> Until recently, philosophy of science focused on relatively homogeneous fields such as classical mechanics. As a result, many scientists have been trained to regard pluralistic approaches as soft, unrigorous, unscientific, and indicative of a retarded field. Even scientists who work in pluralistic fields tend to view how science "should" be pursued in ways that are mismatched to their field's special needs. (Diamond and Case 1986, p. x)

The Newtonian tradition may prevail more in the physical sciences (at least in the non-historical—e.g., non-geological, non-cosmological—physical sciences). The difference between that tradition, and the tradition of relative significance controversies that prevails in biology, is well illustrated by the following Sidney Harris cartoon (see figure 11.3) of two physicists (they're not mathematicians—mathematicians don't wear white coats).

The assumption behind the cartoon—what makes it funny—is that physicists are not supposed to argue about such matters. But what makes us think these are physi-



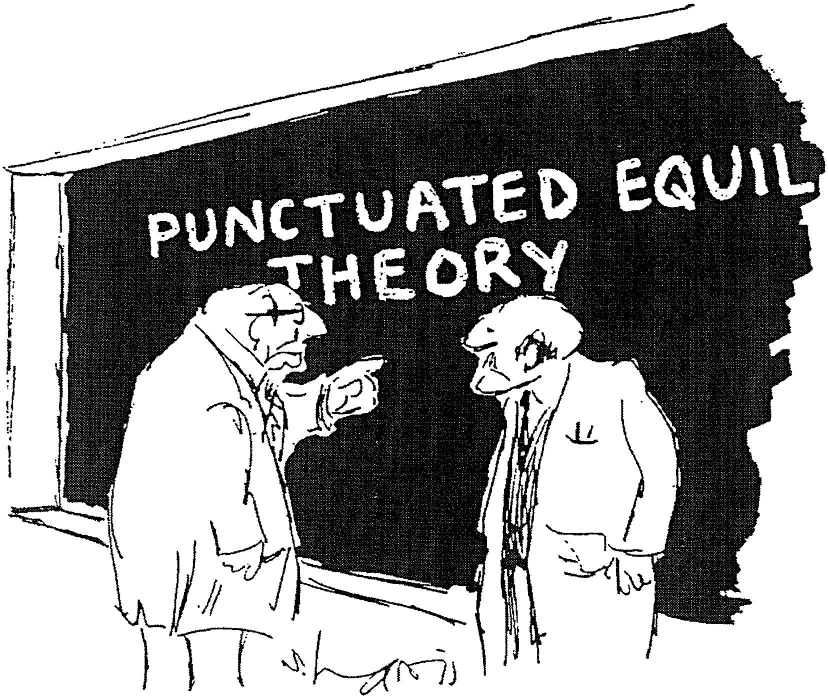
"Does this apply always, sometimes, or never?"

Figure 11.4

cists? Well, if they were not, it would not be funny. Imagine that they are evolutionary biologists arguing about theories of speciation, or theories of the rate of evolutionary change (figures 11.4 and 11.5). Now this is not a joke. It is rather the fact of the matter. To some it is the sad fact of the matter. Which leads me to temper my remarks about theoretical pluralism in biology.

It is important not to exaggerate the differences between the biological and physical sciences. The Newtonian tradition has considerable appeal in biology as well, and not only in the more ahistorical branches of biology, like molecular biology. One also finds it in the more historical areas like systematics, ecology and evolutionary biology. For example, Carson chides his pluralistic peers in the area of speciation for giving up too easily in this regard:

> Despite much modern work in plant and animal population biology, there has been a regrettable lack of unification of theory relating to the modes or processes involved in the origin of new species. I find two reasons for this. First, there is a tendency not to be reductionistic, that is, to accept



"Does this apply always, sometimes, or never?"

Figure 11.5

> many disparate theoretical notions about the way in which species may arise (e.g., White 1978). In the face of this, long and complex classifications of various conceivable modes of speciation (e.g., allopatric, sympatric, parapatric, stasipatric, etc.) have been constructed, discouraging those who seek unifying principles. (Carson 1985, p. 380)

Numerous other biologists are pluralists, but only reluctantly, as if pluralism falls short of the Newtonian ideal. For example, Ghiselin concludes a review of mechanisms of the evolution of sex by admitting that, "Perhaps we shall *have* to accept a pluralistic assemblage of explanations." And again, "From the point of view of having an adequate explanation for all the data, we may *have* to accept more than one hypothesis" (Ghiselin 1988, p. 23, my emphases). And as James and Carol Gould admit, "We have, it seems, an embarrassment of plausible hypotheses to account for the evolution and maintenance of [sex]" (Gould and Gould 1989, p. 65, my emphasis).

Many natural historians reveal the limits of their tolerance for theoretical pluralism by conducting their relative significance arguments in the manner described by Gould and Lewontin:

> In natural history, all possible things happen sometimes; you generally do not support your favoured phenomenon by declaring rivals impossible in theory. Rather, you acknowledge the rival, but circumscribe its domain of action so narrowly that it cannot have any importance in the affairs of nature. Then, you often congratulate yourself for being such an ecumenical chap. (Gould and Lewontin 1978, p. 585)

To a certain extent, as Gould and Lewontin suggest, systematists, ecologists and evolutionary biologists often acknowledge the need for theoretical pluralism, but try to keep it under control by minimizing the significance of all but a couple, or even one, possible account of a domain of phenomena. For example, James and Carol Gould predict with regard to theories of the evolution of sex that one account will turn out to be "*the major factor*," although "there may well be *special cases* in which the [ecological] pressures posited by other models will be critical" (Gould and Gould 1989, p. 65; my emphases).

There are other means by which biologists try to eliminate or contain theoretical pluralism. One important means is by splitting a heterogeneous domain, governed by multiple theories, into two or more homogeneous sub-domains, each governed by only a couple of theories or perhaps even a single theory. For example, the evolution of sex can be partitioned into two sub-domains, the *origin* of sex and the *maintenance* of sex. Some biologists believe that the DNA-repair hypothesis will be *the* account of the origin of sex, while multiple theories may be needed to explain the maintenance of sex. Similarly, attempts are often made to distinguish the domain of microevolutionary changes within which selectionist theories are most significant, from the domain in which neutralist theories are most important (e.g., Endler 1986, Chap. 8; Gillespie 1991, pp. 289–290). Interestingly, while Diamond and Case (quoted above) promote theoretical pluralism in community ecology, they nonetheless seriously entertain the possibility that "one can at least partition communities among a modest number of types and devise a model for each type" (Diamond and Case 1986, p. xi). Douglas Allchin (1991) devotes considerable attention to the strategy of domain partitioning.

But the fact that tactics like these are employed to contain theoretical pluralism indicates that theoretical pluralism is indeed widespread, however much some biologists with Newtonian inclinations may regret it. As Michael Gilpin acknowledges in objecting to a fellow ecologist's pluralistic approach:

> I must confess that I am saddened by [his book's] honest realism, its unabashed pluralism. Something of a romantic, I long for the heady days of an earlier decade when the [alternatives considered by him] vied one against the other to be the organizing principles of our science. (Gilpin 1986, pp. 200–201)

I can imagine an argument for adhering to the Newtonian ideal, independently of its romantic qualities. That is, one might suggest that theoretical pluralism reflects more about the state of our ignorance than about the state of nature: there may actually be laws of biology, and a unitary or unifying theory for each domain of biological phenomena, but we have yet to discover these important generalizations. Whether theoretical pluralism reflects the nature of the biological world, or the state of our ignorance, we cannot at present know. Nonetheless, we should aim for unitary or unifying theories. Then, if the biological world is really so contingent, and each domain inescapably heterogeneous, we will ultimately be forced to deal with theoretical pluralism. But if we begin by advocating theoretical pluralism, then we may never find the unitary or unifying theories that might actually be true. We might rest happy with multiple accounts when a unitary account is possible and could be discovered with just a little more effort. I am not just imagining this argument, it is a lot like David Hull's argument in favor of a single species concept. According to Hull, defenders of a unitary vs. a pluralistic species concept,

> are carrying on in the best scientific tradition of opting for one perspective and pushing it for all its worth. Perhaps species as genealogical [individuals]... may prove ultimately inadequate. Science does march on. If so, then this monism will have proved to have been only temporary, but the only way to find out how adequate a particular conception happens to be is to give it a run for its money. Remaining content with a variety of slightly or radically different species concepts might be admirably open-minded and liberal, but it would be destructive of science ... (Hull 1987, p. 178)

This is a difficult argument to counter.[24] The best I can do is to offer an alternative argument (or rather, sketch of an argument), which rests on the following premise: scientific methodology, including injunctions to seek unified accounts of each and every domain, should be scientifically (in this case evolutionarily) informed. This is, for example, the assumption that structures Elliott Sober's analysis of the evolutionary grounds underlying the parsimony criterion in phylogenetic systematics. As Sober puts it,

> The idea of a presuppositionless "scientific method" implies that methodology is static and insensitive to what we learn about the world. But with theory and method linked by a subtle nexus of interdependence, progress on theories can be expected to improve our methods of inquiry (Sober 1988, p. 239).

Similarly, why should we adhere to a methodology that dictates the search for unitary accounts of each domain of biological phenomena—e.g., a unitary account of inheritance, or a unitary account of carbohydrate metabolism, or a unitary account of gene regulation, or a unitary account of speciation—unless we have reason to believe that the outcomes of evolution are *highly* constrained?

If I have stressed the factors that broaden the range of evolutionary possibilities, that is not because I think they are intrinsically more interesting or important than the factors that limit the range. But unless we believe that the outcomes of evolution are always severely constrained, then perhaps we should be *on the lookout* for multiple accounts of each domain. Only a naive Newtonian would rest satisfied with a unitary account, when, with a little more effort, a multiplicity of accounts might be found!

4 Conclusion

I will summarize very briefly. Gould contrasts the ''laws in the background'' with the ''contingent details'' in the foreground of biology. What this means to me is that there *may* be genuine laws that are relevant to biology (e.g., laws of physics and chemistry), but those laws are not distinctively biological. What is distinctively biological are the contingent details, allowed but not necessitated by the presumed laws. The details can have most any degree of generality—and the degrees of generality of those details may change over time. And all the while evolution is making new rules and breaking old rules, the rules of evolution are themselves changing.

This evolutionary contingency thesis bears upon a number of other issues in philosophy of biology, including the explanatory ideals manifested in ''theoretical pluralism,'' and the nature of ''relative significance'' controversies. Just as a scientific hypothesis derives support from the phenomena it successfully explains, so, too, the evolutionary contingency thesis derives support from these other general features of biology that it makes sense of.

Notes

1. This essay is dedicated to Lorenz Krüger, who made me think hard about chance, and who made it fun.

I am very grateful for improvements that I owe to Peter Abrams, Douglas Allchin, Robert Brandon, Richard Burian, Joseph Cain, James Curtsinger, Daniel Dennett, Marc Ereshefsky, Ronald Giere, Stephen Gould, Ernst Mayr, Robert Richardson, Peter Sloep, Kelly Smith, Kenneth Waters, and also to audiences of philosophers and biologists at Duke, Johns Hopkins, Michigan, Tufts, and Virginia Polytechnic Institute. I am also appreciative of the very thoughtful suggestions made by my commentators, Martin Carrier and Kenneth Schaffner.

2. Richard Lewontin (1990) defends a similarly strong version in his review of Gould.

3. I should probably emphasize early on a point to which I will return later. That is, the evolutionary contingency thesis has nothing at all to say about whether there are laws of physics and chemistry. The physical and chemical generalizations that are true of the living and non-living worlds may also be contingent—presumably not evolutionarily contingent, but perhaps cosmologically contingent. I should probably also allude to one other point to which I will return later. That is, the evolutionary contingency thesis is compatible with there being *degrees* of evolutionary contingency. Distinctively biological generalizations may be *more or less* contingent.

4. Acetic acid can be oxidized and CO_2 produced by a simpler series of reactions, which does not involve attaching the acetyl unit to the oxaloacetate carrier. But this reaction does not lead to

the production of as many energy rich ATP molecules. The use of the oxaloacetate carrier—qua carrier—also makes sense of the cyclical nature of the Krebs reactions. In this way the carrier is regenerated, so that organisms do not have to ingest massive amounts of it in order to carry out acetic acid metabolism. See also Stryer (1988, pp. 392–393).

5. Of course, to say that a particular generalization about the Krebs cycle is not merely a chemical generalization, but also describes an evolutionary outcome, is not to say that the principles of chemistry are irrelevant for understanding it. It is just to say that evolutionary reasoning is also relevant. The evolutionary contingency thesis does not divide generalizations about the living world into those that can be explained entirely in terms of mathematics, physics and chemistry, and those that can be explained entirely in evolutionary terms without reference to principles of mathematics, physics and chemistry. The distinction is rather between those generalizations that can be explained entirely in terms of mathematics, physics and chemistry, and those that require an evolutionary perspective *in addition*.

6. Fisher's "fundamental theorem of natural selection" also depends on Mendel's first "law." That is, it also fails in some cases of abnormal meiosis. Indeed, population geneticists often "prove" the fundamental theorem (or perhaps, given the notorious vagueness of the theorem, it is better to say that population geneticists "interpret" it) by deriving it from the Hardy–Weinberg "law" (e.g., Spiess 1977, pp. 402–436).

Shimony (1989) invokes a line of reasoning very much like the evolutionary contingency thesis to argue that there is no fundamental (independent of genetics) "principle" of evolution by natural selection.

7. My discussion of the existence of laws of biology owes much to the classic (but otherwise very different!) analyses by J. J. C. Smart (there are no laws of biology—1963, pp. 53–59), Michael Ruse (yes there are—1973, pp. 24–31), and David Hull (it is a very difficult issue!—1974, pp. 70–100).

My discussion also owes much to Ernst Mayr's analysis of laws of biology. Mayr does not explicitly state the evolutionary contingency thesis, but I believe that it is central to his denial of the existence of laws of biology, and also central to his broad conception of the nature and identity of biology (Mayr 1982, pp. 32–76). However, it would take a separate paper to defend this point properly.

Philip Kitcher (1985) and John Dupré (1993) have also recently questioned the existence of laws of biology (see also note 23).

The most rigorous defense of laws of biology to date is by Kenneth Waters (ms.), which I discuss later in the text.

8. Van Fraassen actually argues that there are *no* unproblematic criteria for laws of nature; he proposes that we dispense with the category. His criticisms of the most commonly ventured criteria are indeed persuasive. Perhaps there are no laws in *any* science, at least no "laws" in any sense that we can successfully explicate. My concern here, however, is just to show that there are in any case no laws *of biology*.

9. Many philosophers of science argue that the ultimate criterion for determining whether a generalization is a law is its ability to support "counterfactual conditionals." Thus the first claim

above is a law because we could infer on the basis of that generalization that if any object *X*—say a 100 meter sphere of some substance *besides* uranium—*were* (contrary to fact) a solid sphere of enriched uranium, *then* it would cease to exist. Whereas we would not infer on the basis of the second generalization that if any object *X*—say a 100 meter sphere of some substance *besides* gold—*were* gold then it would instantaneously shrink or cease to exist. But surely it is on the basis of what we believe is precluded or allowed by nature that we feel the first but not the second inference is justified. So the counterfactual conditional criterion is actually derivative upon the criterion of natural necessity.

10. Hull (1978, pp. 353–355) raises a similar objection to the lawlike status of many so-called "laws" of biology on behalf of his "taxa-as-individuals" thesis. That is, if taxa are spatiotemporally individuated lineages, then they are the sorts of entities that can evolve (the best argument for construing them as lineages!), in which case generalizations about taxa are only ever temporarily (in evolutionary time) true. However, my argument is broader than Hull's, because it pertains not only to generalizations about particular taxa (e.g., humans), but also to generalizations *across taxa* (e.g., generalizations about sexual organisms). Ereshefsky's most recent criticism of my position overlooks this point (Ereshefsky 1992, pp. 93–96).

11. There are many other sources of this weaker form of contingency, like random drift of gene frequencies in small populations and/or among selectively equivalent genotypes, which could also render temporary any genetically based generality. However, I prefer to emphasize the respects in which evolution *by natural selection* (vs. by mutation alone, or by random drift) can have contingent outcomes. This is because selection has traditionally been conceived and labeled as a "determinate" or "deterministic" agent of evolution (in contrast to mutation and drift—see, e.g., Wright 1949, p. 369; Sober 1984, p. 110). Thus, for instance, Richard Dawkins (1986, p. 41) exclaims, "Mutation is random; natural selection is the very opposite of random." This might leave one with the unfortunate impression (not that any of the authors I just cited ever explicitly sanctioned this inference) that outcomes of evolution are contingent only to the extent that they are due to mutation alone or to random drift, and non-contingent to the extent that they are due to selection. The second sense of contingency to be discussed shortly makes clear why this would be an especially unfortunate conclusion to draw.

12. Sober does not actually invoke any particular criteria characteristic of laws of nature; while it is clear what he means by "*source* law," it is not clear what he means by "source *law*." He may not intend as strong a sense of "law" as is employed here.

13. On one such scenario, equal segregation of genes among gametes is a good way to maintain, and hence take advantage of, widespread overdominance (a form of advantage of heterozygotes over homozygotes—see Liberman and Feldman 1980; see also the discussion by Bell 1982, p. 439). An appropriate conditional might therefore be, "whenever overdominance is widespread . . ."

14. Some researchers have proposed that biosynthetic fragments of the cycle were already present in very primitive anaerobic ancestors of aerobes, before oxygen was present in the atmosphere; the cycle was then completed during the evolution of early aerobes in an increasingly oxygen-rich atmosphere (e.g., Gest 1987; Weitzman 1985).

15. Ruse (1973, pp. 59–62) and Hull (1974, p. 79) long ago raised basically this same objection to the lawlike status of such ecological rules.

16. See Michael Ghiselin's excellent discussion of Darwin on orchids (Ghiselin 1969, pp. 131–159; see also Gigerenzer et al. 1989, pp. 132–140).

17. There are still other sources of this stronger sense of "contingency," that is, besides random mutation and natural selection leading to functional equivalence. For instance, random drift of gene frequencies in small populations can lead to different evolutionary outcomes from the same starting point, even under the same selection regimes, and even when there is no difference in the order and timing of the introduction of mutations into the process. For reasons discussed in note 11, though, I continue to emphasize the contingent aspects of evolution *by natural selection*.

18. I want to make it clear that I am referring and responding to a *draft* of an essay by Waters. I know I will have my hands full dealing with the completed version!

19. There is another strategy for formulating laws that should be mentioned. This "ceteris paribus" or "disturbing condition" strategy is much more general than the source-law and gene-expression strategies; indeed, the latter two strategies can be viewed as instances of the former.

It is sometimes suggested that generalizations like Mendel's "law" can be saved from their exceptions, and rendered true laws, by conditionalizing them upon the occurrence of all the appropriate "ceteris paribus" conditions, and/or the *non* occurrence of all the relevant "disturbing conditions" (e.g., Ereshefsky 1991). Thus, appended to the end of every biological law would be one or both of the following two clauses: "except when any of the following ceteris paribus conditions fails to obtain [ceteris paribus conditions listed], and except when any of the following disturbing conditions occur [disturbing conditions listed]."

As Nancy Cartwright notes, we most often translate "ceteris paribus" as "other things being equal," but what we really mean is "other things being just *right*" (Cartwright 1983, p. 45). But the conditions that are just right for the evolution of a biological generality are exceedingly difficult (impossible?) to enumerate for reasons that we have been discussing. Similarly, it would be exceedingly difficult to enumerate all the relevant disturbing conditions—all the evolutionary scenarios that would result in exceptions to—a previously evolved generality.

Instead of actually listing the ceteris paribus and disturbing conditions, we could just refer to those two general categories. For example, in the case of Mendel's "law," we might generalize, "among sexual organisms, there is 50:50 segregation, except when any of the appropriate ceteris paribus conditions fails to obtain, or when any of the relevant disturbing conditions occurs."

But there is an often noted problem with this strategy of lawmaking. Namely, it seems to purchase the truth of the supposed "law" at the cost of its empirical status. Presumably, laws *of nature* are supposed to be true as a matter of *empirical fact*, and not simply logically true. But how are we to interpret "the appropriate ceteris paribus conditions" and "the relevant disturbing conditions" so as to make the Mendelian generalization empirically and not merely logically true? In other words, how do we avoid the interpretation, "among diploids, there is 50:50 segregation, except when there obtain conditions that lead to non-50:50 segregation?" The problem is not ceteris paribus and disturbing conditions per se, but rather our inability to enumerate them (see, e.g., Giere forthcoming).

20. I appreciate Schaffner's point about degrees of contingency, but his example is interestingly problematic. However well "entrenched by natural selection" the genetic code may be, it has well known exceptions. The code is not universal. There is considerable variation, and not simply as a result of recent mutations that have yet to be eliminated by natural selection. There are "predictable" differences (not just due to recent mutations) in codes between mitochondrial and non-mitochondrial nucleic acids, and also among non-mitochondrial nucleic acids of different taxa. There are even predictable, site-specific code differences in the mitochondrial nucleic acid of a single taxon, and also site-specific code differences in the non-mitochondrial nucleic acids of a single taxon (see the general review by Fox 1987). Given the possibility that these differences are due to evolution by natural selection (which is at present unknown), then in what sense could selection properly be said to have "frozen" the code into "nomic universality?" Schaffner is surely right that there are degrees of contingency in the biological, as in the physical world. But just as surely, the degrees are more continuous than he suggests. Moreover, between his "essentially" accidental and merely "historically" accidental generalizations fall most (all?) distinctively biological generalizations.

21. Kauffman's own position is that self-organizing properties of biological systems place considerable constraints upon the outcomes of evolution, so that there may be distinctively biological generalizations that are contingent, but they are not *highly* contingent. I am not proud to say this, but I admit that I cannot judge the merits of Kauffman's arguments regarding the degree of contingency that actually obtains. It seems to me possible that the greatest merit of his work lies in his *focus* on the extent to which the outcomes of evolution are constrained. As he persuasively argues, it has been the tradition to focus instead on the extent to which the outcomes of evolution are not constrained—e.g., the extent to which evolution by natural selection is a Rube Goldberg tinkerer. Unfortunately, that puts me in the old-fashioned camp. In which case, I can only hope that I have elaborated the old-fashioned position clearly!

22. "Theoretical pluralism" has multiple meanings in the literature. One that is quite defensible, but different from the sense that I will be discussing, has to do with the idea that there are multiple causes for any particular biological phenomenon. For example, 1) a particular phenotypic trait is the result of the interaction of genotype *and* phenotype, 2) the presence of a trait may be viewed from an "ultimate," evolutionary perspective, *and/or* a "proximate," developmental perspective, 3) no evolutionary change is the result of natural selection *alone*—in any finite population random drift plays *some* role, etc. See Mitchell (1992) for a careful discussion of pluralism in this sense. In contrast, by "theoretical pluralism" I mean to refer to the way in which biologists explain a *domain* of phenomena, rather than any *individual* phenomenon. Another form of pluralism, perhaps more closely related to theoretical pluralism, is the form explored in John Dupré's recent book, *The Disunity of Science* (1993). Dupré defends a brand of ontological pluralism based on the rejection of essentialism. Thus he argues that there are no natural kinds in any strong sense in biology. He articulates his anti-essentialist position in several different ways, one of which is related to the non-existence of biological laws. I will return to Dupré's notion of pluralism in note 23.

23. This seems to bear upon what Dupré means by pluralism. The sense of pluralism that he defends involves a denial of the existence of natural kinds, by which he means in part that there

are no true laws of nature (Dupré 1993, pp. 63, 65). One important difference between Dupré's treatment of pluralism and mine is that I offer a *causal explanation* of why pluralism prevails in biology—in terms of the evolutionary contingency thesis—whereas it seems to me that Dupré is most concerned to document or *establish* pluralism.

But Dupré might not agree with my explanation, because in a sense it gives evolutionary thinking a unifying role in biology, and Dupré is concerned to defend the *disunity* of biology. My approach does not unify biology in terms of any particular evolutionary generalization, because evolutionary generalizations are highly contingent just like other biological generalizations. Nevertheless, my approach does suggest a unifying role for evolutionary thinking more broadly speaking: we can make biological sense of pluralism by thinking evolutionarily.

24. My position is similar to but also different from Dupré's (1993, pp. 52–53). Dupré argues that Hull begs the question—that Hull's viewpoint "is plausible only if one is already committed to the view that science requires, in the end, a unified biology with a wholly univocal concept of the species" (p. 53). I think Hull's argument is more sophisticated. To elaborate once more on the argument as it applies to theoretical pluralism, *if* nature is inescapably heterogeneous, the Newtonian would not forever overlook that fact, but would be faced with nature's heterogeneity over and over again. The Newtonian would ultimately be forced to acknowledge theoretical pluralism in that case. According to this argument, theoretical pluralism is possibly misleading, whereas the Newtonian tradition is, at worst, inefficient.

References

Allchin, D. (1991), *Resolving Disagreement in Science: The OxPhos Controversy, 1961–1977*. Ph.D. Dissertation, University of Chicago.

Beatty, J. (1981), "What's Wrong with the Received View of Evolutionary Theory?" in P. D. Asquith and R. N. Giere (eds.), *PSA 1980*, vol. 2. East Lansing, Michigan: Philosophy of Science Association, pp. 397–426.

———. (1982), "The Insights and Oversights of Molecular Genetics: The Place of the Evolutionary Perspective," in P. D. Asquith and T. Nickles (eds.), *PSA 1982*, vol. 2. East Lansing, Michigan: Philosophy of Science Association, pp. 341–355.

———. (1994), "Theoretical Pluralism in Biology, Including Systematics," in L. Grande and O. Rieppel (eds.), *Interpreting the Hierarchy of Nature*. London: Academic Press, pp. 33–60.

Bell, G. (1982), *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*. Berkeley: University of California Press.

Brace, C. L. (1966), "Letter to the Editor," *Science 153*: 362.

Brock, T. D., D. W. Smith, and M. T. Madigan (1984), *Biology of Microorganisms*. Englewood Cliffs, New Jersey: Prentice-Hall.

Bush, G. (1975), "Modes of Animal Speciation," *Annual Review of Ecology and Systematics 6*: 339–364.

Campbell, W. (1966), *Human Evolution*. Chicago: Aldine.

Carson, H. L. (1985), "Unification of Speciation Theory in Plants and Animals," *Systematic Botany 10*: 380–390.

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

Comfort, A. (1979), *The Biology of Senescence*. New York: Elsevier.

Connell, J. H. (1983), "On the Prevalence and Relative Importance of Interspecific Competition: Evidence from Field Experiments," *American Naturalist 122*: 661–696.

Crick, F. H. C. (1968), "The Origin of the Genetic Code," *Journal of Molecular Biology 19*: 367–397.

Crow, J. (1979), "Genes that Violate Mendel's Rules," *Scientific American 240 (2)*: 134–146.

Darwin, C. (1862), *On the Various Contrivances by which British and Foreign Orchids are Fertilised by Insects, and on the Good Effects of Intercrossing*. London: John Murray.

———. (1868), *Variation of Animals and Plants under Domestication*. 2 vols. London: John Murray.

———. (1871), *The Descent of Man, and Selection in Relation to Sex*. 2 vols. London: John Murray.

———. (1872), *On the Origin of Species by Means of Natural Selection*, 6th ed., 2 vols. London: John Murray.

Dawkins, R. (1986), *The Blind Watchmaker*. New York: Norton.

Delbrück, M. (1966), "A Physicist Looks at Biology," in J. Cairns et al. (eds.), *Phage and the Origins of Molecular Biology*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory. Originally published in *Transactions of the Connecticut Academy of Arts and Sciences 38*: 173–190.

Diamond, J., and T. Case, (eds.) (1986), *Community Ecology*. New York: Harper and Row.

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

Endler, J. A. (1986), *Natural Selection in the Wild*. Princeton: Princeton University Press.

Ereshefsky, M. (1991), "The Semantic Approach to Evolutionary Theory," *Biology and Philosophy 5*: 7–28.

———. (1992), "The Historical Nature of Evolutionary Theory," in M. H. Nitecki and D. V. Nitecki (eds.), *History and Evolution*. Albany: State University of New York Press, pp. 81–99.

Finch, C. E. (1990), *Longevity, Senescence, and the Genome*. Chicago: University of Chicago Press.

Fox, T. D. (1987), "Natural Variation in the Genetic Code," *Annual Review of Ecology and Systematics 21*: 67–91.

Futuyma, D. (1979), *Evolutionary Biology*. Sunderland, Mass.: Sinauer.

Gest, H. (1987), "Evolutionary Roots of the Citric Acid Cycle in Prokaryotes," in J. Kay and P. D. J. Weitzman (eds.), *Krebs' Citric Acid Cycle—Half a Century and Still Turning*. London: The Biochemical Society, pp. 3–16.

Ghiselin, M. T. (1969), *The Triumph of the Darwinian Method*. Berkeley: University of California Press.

———. (1988), "The Evolution of Sex: A History of Comparative Points of View," in R. E. Michod and B. R. Levin (eds.), *The Evolution of Sex*. Sunderland, Mass.: Sinauer, pp. 7–23.

Giere, R. (forthcoming), "Science without Laws of Nature," in F. Weinert (ed.), *Laws of Nature*. The Hague: de Gruyter.

Gigerenzer, G. et al. (1989), *The Empire of Chance: How Probability Changed Science and Life*. Cambridge: Cambridge University Press.

Gillespie, J. H. (1991), *The Causes of Molecular Evolution*. Oxford: Oxford University Press.

Gilpin, M. E. (1986), "Review of R. J. Taylor, Predation," *American Scientist 74 (2)*: 200–201.

Goldberg, R. (1979), *The Best of Rube Goldberg*. Engelwood Cliffs, New Jersey: Prentice-Hall.

Gottschalk, G. (1986), *Bacterial Metabolism*. New York: Springer.

Gould, J. L., and C. G. Gould (1989), *Sexual Selection*. New York: Scientific American Library.

Gould, S. J. (1977), *Ontogeny and Phylogeny*. Cambridge, Mass.: Harvard University Press.

———. (1980), "Is a New and General Theory of Evolution Emerging?" *Paleobiology 6*: 119–130.

———. (1989), *Wonderful Life*. New York: Norton.

Gould, S. J., and R. C. Lewontin (1979), "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme," *Proceedings of the Royal Society of London B205*: 581–598.

Grene, M. (1988), "Our Place in Nature," manuscript.

Harris, S. (1977), *What's So Funny About Science?* Los Altos, California: Kaufmann.

Hempel, C. G. (1966), *Philosophy of Natural Science*. Englewood Cliffs, New Jersey: Prentice-Hall.

Hull, D. L. (1974), *Philosophy of Biological Science*. Englewood Cliffs, New Jersey: Prentice-Hall.

———. (1978), "A Matter of Individuality," *Philosophy of Science 45*: 335–360.

———. (1987), "Genealogical Actors in Ecological Roles," *Biology and Philosophy 2*: 168–184.

Jacob, F. (1982), *The Possible and the Actual*. New York: Pantheon.

Kauffman, S. (1993), *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford: Oxford University Press.

Kimura, M. (1983), *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Kindt, T. J., and J. D. Capra (1984), *The Antibody Enigma*. New York: Plenum.

Kitcher, P. (1985), "Darwin's Achievement," in N. Rescher (ed.), *Reason and Rationality in Science*. Washington, D.C.: University Press of America, pp. 127–189.

Kondrashov, A. S. (1988), "Deleterious Mutations and the Evolution of Sexual Reproduction," *Nature 336*: 435–440.

Krebs, H. A. (1981), "The Evolution of Metabolic Pathways," in J. F. Carlile et al. (eds.), *Molecular and Cellular Aspects of Microbial Evolution*. Cambridge: Cambridge University Press, pp. 215–288.

Lande, R. (1980), "Microevolution in Relation to Macroevolution," *Paleobiology 6*: 235–238.

Lewin, B. (1990), *Genes IV*. Oxford: Oxford University Press.

Lewontin, R. C. (1990), "Fallen Angels," *New York Review of Books 37 (6)*: 3–7.

———. (1974), *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Liberman, U., and M. W. Feldman (1980), "On the Evolutionary Significance of Mendel's Ratios," *Theoretical Population Biology 17*: 1–15.

McIntosh, R. P. (1987), "Pluralism in Ecology," *Annual Review of Ecology and Systematics 18*: 321–341.

Magee, P. T. (1987), "Transcription During Meiosis," in P. B. Moens (ed.), *Meiosis*. London: Academic Press, pp. 355–382.

Maynard Smith, J. (1966), "Theories of Aging," in P. L. Krohn (ed.), *Topics in the Biology of Aging*. New York: Interscience.

Mayr, E. (1956), "Geographic Character Gradients and Climatic Adaptation," *Evolution 10*: 105–108.

———. (1963), Animal Species and Evolution. Cambridge: Harvard University Press.

———. (1982), *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Cambridge, Mass.: Harvard University Press.

Michod, R. E., and B. R. Levin (eds.), *The Evolution of Sex*. Sunderland, Mass.: Sinauer.

Mitchell, S. (1992), "On Pluralism and Competition in Evolutionary Explanations," *American Zoologist 32*: 135–144.

Nelson, G., and N. Platnick, (eds.) (1981a), *Systematics and Biogeography: Cladistics and Vicariance*. New York: Columbia University Press.

———. (1981b), *Vicariance Biogeography: A Critique*. New York: Columbia University Press.

Moat, A. G., and J. W. Foster (1988), *Microbial Physiology*. New York: Wiley.

Newton, I. ([1686] 1934), *Mathematical Principles of Natural Philosophy* (A. Motte, transl.; F. Cajori, ed.). Berkeley: University of California Press.

Olson, W. S. (1966), "Letter to the Editor," *Science 153*: 364.

Otte, D., and J. A. Endler, (eds.) (1989), *Speciation and its Consequences*. Sunderland, Mass.: Sinauer.

Rose, M. R. (1985), "The Evolution of Senescence," in P. J. Greenwood et al. (eds.), *Evolution: Essays in Honor of John Maynard Smith*. Cambridge: Cambridge University Press.

Rosenberg, A. (1985), *The Structure of Biological Science*. Cambridge: Cambridge University Press.

Ruse, M. (1973), *The Philosophy of Biology*. London: Hutchinson.

Sandler, L. et al. (1968), "Mutants Affecting Meiosis in Natural Populations of Drosophila Melanogaster," *Genetics 60*: 525–558.

Schaffner, K. F. (1980), "Theory Structure in the Biomedical Sciences," *The Journal of Medicine and Philosophy 5*: 57–97.

———. (1993), *Discovery and Explanation in the Biomedical Sciences*. Chicago: University of Chicago Press.

Schoener, T. (1982), "The Controversy over Interspecific Competition," *American Scientist 70*: 586–595.

———. (1983), "Field Experiments on Interspecific Competition," *American Naturalist 122*: 240–285.

Shimony, A. (1989), "The Non-Existence of a Principle of Natural Selection," *Biology and Philosophy 4*: 255–273.

Sih, A., et al. (1986), "Predation, Competition, and Prey Communities: A Review of Field Experiments," *Annual Review of Ecology and Systematics 16*: 269–311.

Smart, J. J. C. (1963), *Philosophy and Scientific Realism*. London: Routledge and Kegan Paul.

Sober, E. (1984), *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, Mass.: MIT Press.

———. (1987), "Does 'Fitness' Fit the Facts?" *Journal of Philosophy 84*: 220–223.

———. (1988), *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, Mass.: MIT Press.

———. (1989), "Is the Theory of Natural Selection Unprincipled?" *Biology and Philosophy 4*: 275–279.

Spiess, E. B. (1977), *Genes in Populations*. New York: Wiley.

Stryer, L. (1988), *Biochemistry*. New York: Freeman.

Uyenoyama, M. K. (1987), "Genetic Transmission and the Evolution of Reproduction: The Significance of Parent-Offspring Relatedness to the 'Cost of Meiosis,'" in P. B. Moens (ed.), *Meiosis*. London: Academic Press.

van Fraassen, B. (1989), *Laws and Symmetry*. Oxford: Clarendon Press.

Waters, C. K. (ms. in preparation), "Trends and Laws in Biology".

Weinberg, S. (1992), *Dreams of a Final Theory*. New York: Pantheon.

Weitzman, P. D. J. (1985), "Evolution in the Citric Acid Cycle," in K. H. Schleifer, *Evolution of Prokaryotes*. London: Academic Press, pp. 253–275.

Wheeler, P. (1985), "The Loss of Functional Body Hair in Man: The Influence of Themal Environment, Body Form, and Bipedality," *Journal of Human Evolution 42*: 12–27.

White, M. J. D. (1973), *Animal Cytology and Evolution*. London: Clowes.

———. (1978), *Modes of Speciation*. San Francisco: Freeman.

Wright, S. (1949), "Adaptation and Selection," in G. L. Jepsen, E. Mayr, and G. G. Simpson (eds.), *Genetics, Paleontology, and Evolution*. Princeton: Princeton University Press, pp. 365–389.

Yanofsky, C. (1981), "Attenuation in the Control of Expression of Bacterial Operons," *Nature 289*: 751–758.

Yanofsky, C. (1988), "Transcription Attenuation," *Journal of Biological Chemistry 263*: 609–612.

12 Two Outbreaks of Lawlessness in Recent Philosophy of Biology

Elliott Sober

John Beatty (1995) and Alexander Rosenberg (1994) have argued against the claim that there are laws in biology. Beatty's main reason is that evolution is a process full of contingency, but he also takes the existence of relative significance controversies in biology and the popularity of pluralistic approaches to a variety of evolutionary questions to be evidence for biology's lawlessness. Rosenberg's main argument appeals to the idea that biological properties supervene on large numbers of physical properties, but he also develops case studies of biological controversies to defend his thesis that biology is best understood as an instrumental discipline. The present chapter assesses their arguments.

1 Introduction

Are there laws in biology? John Beatty (1995) says there are none and Alexander Rosenberg (1994) says there is just one. Have they got their numbers wrong? That's a question I will want to address. However, my first concern is the arguments they give. Do the considerations they adduce support the lawlessness they advocate?

Beatty and Rosenberg rely on a standard logical empiricist conception of law. Laws are true generalizations that are ''purely qualitative,'' meaning that they do not refer to any place, time, or individual. They have counterfactual force. And finally, Beatty and Rosenberg require that laws be empirical. My main disagreement with this traditional picture is that I want to leave open whether a law is empirical or *a priori*. I have argued elsewhere that the process of evolution is governed by models that can be known to be true *a priori* (Sober 1984, 1993). For example, Fisher's (1930) fundamental theorem of natural selection says that the rate of increase in fitness in a population at a time equals the additive genetic variance in fitness at that time. When appropriately spelled out, it turns out to be a mathematical truth—in populations of a certain sort, fitness increases at the rate that Fisher identified. Fisher's theorem governs the

Philosophy of Science 64 (Proceedings), pp. S458–S467. © 1997 by the Philosophy of Science Association. All rights reserved. Reprinted by permission of The University of Chicago Press.

trajectories of populations just as Newton's laws govern the trajectories of particles. Fisher's theorem, and statements like it, are purely qualitative, support counterfactuals, and describe causal and explanatory relations. Because evolutionary processes are governed by such propositions, I want to say that evolution is lawful. How we are able to *know* the laws of evolution is a separate question. Whether a natural process is lawful is not an epistemological issue (Dretske 1977).

This revised notion of law does not entail that every *a priori* statement is a law. The concept of a *process law* allows us to avoid this result. A process law is a counterfactual-supporting, qualitative generalization, which describes how systems of specified type develop through time. Typically, such laws are time-translationally invariant (Sober 1994); given a system that occupies a particular state at one time, a process law describes the probability distribution of the different states the system may occupy some fixed amount of time later; the date of the starting time is irrelevant. "Bachelors are unmarried," is not a process law, but not because it is *a priori*.

When I claim that there are laws of evolution, and Beatty and Rosenberg demur, there is no disagreement. My use of the term "law" leaves open whether a law is empirical; Beatty's and Rosenberg's does not. Beatty and I agree on the bottom line—there are no *empirical* laws of evolution; Rosenberg and I also agree, save for the one exception he has in mind. So what is there to argue about? It is their *reasons* for denying the existence of empirical laws. This is the bone I want to pick.

2 The Evolutionary Contingency Thesis

Beatty (pp. 46–47) articulates his claim about lawlessness by describing what he calls *the evolutionary contingency thesis* (the ECT): "All generalizations about the living world are just mathematical, physical, or chemical generalizations (or deductive consequences of mathematical, physical, or chemical generalizations plus initial conditions), or are distinctively biological, in which case they describe contingent outcomes of evolution." Beatty agrees that organisms obey the laws of physics. However, there is no additional layer of autonomous biological law that living things also obey.

Beatty illustrates what he means by the evolutionary origins of biological regularities by discussing several examples. One of them is Mendel's first "law," which says that diploid sexual organisms form haploid gametes by a "fair" 50/50 meiotic division. Beatty cites two reasons for thinking that this is not a law. First, it is sometimes false; genes that cause segregation distortion are counterexamples. And more importantly, fair meiosis, when it exists, is a contingent outcome of evolutionary processes; another set of initial conditions would have produced a different segregation ratio.

Below is a schematic version of the ECT. A set (I) of contingent initial conditions obtains at one time (t_0); this causes a generalization to hold true during some later temporal period (from t_1 to t_2):

$$\begin{array}{lll} \mathrm{I} \;\rightarrow & [\text{if P} & \text{then Q}] \\ \hline t_0 & t_1 & t_2 \end{array}$$

Since the generalization is true only because I obtained, we may conclude that the generalization is contingent. However, there is *another* generalization that this scenario suggests, and it is far from clear that *this* generalization is contingents. This generalization will have the following logical form:

(L) If I obtains at one time, then the generalization [if P then Q] will hold thereafter.

The fact that the generalization [if P then Q] is contingent on I does not show that proposition (L) is contingent on anything. This point also holds if (L) is given a probabilistic formulation.

Is proposition (L) contingent? Recall that the ECT is a claim about *causality*; as applied to Mendel's first law, it claims that the evolutionary process *caused* segregation ratios to take the values they did. If causality requires the existence of laws, then there *must* be laws in the background—the evolutionary contingency of [if P then Q] *demands* the existence of laws. Anscombe (1975) argues that causal claims about token events do not entail the existence of general laws. Her point concerns the meaning of the word "cause," and she may be right. Still, it is part of the practice of science to expect noncontingent generalizations when one event causes another, and her observations do nothing to discredit this expectation.

If (L) is noncontingent, is it "distinctively biological?" In one sense, it is. The generalization that helps explain a given segregation ratio describes the variation found in ancestral populations, the fitnesses that attached to those variants, the background biology present in the population, etc. The generalization is biological because of its distinctive vocabulary. However, there is another way to interpret "distinctively biological," which has the opposite result. If a distinctively biological proposition cannot be *a priori*, then (L) is not distinctively biological if it is a mathematical truth. On this interpretation, the fact that (L) is not contingent is no threat to the ECT. This proposal has the curious result that biologists are not doing biology when they construct mathematical models of biological processes; rather, they are doing mathematics. There probably is no point in disputing how the phrase "distinctively biological" should be understood. The idea I want to emphasize is that the contingency of Mendel's "laws" on a set of prior evolutionary events should lead us to expect that there are *other* general propositions that are *not* contingent on that set.

3 Relative Significance Controversies and Theoretical Pluralism

Beatty has two further arguments for the ECT. The fact that biologists engage in "relative significance controversies" and find attractive "the explanatory ideals manifested in 'theoretical pluralism'" are said to "support" the ECT (p. 76). The question may be asked as to what the behavior of *scientists* shows about the existence of *laws*. Beatty's idea seems to be that pluralism and an interest in relative significance controversies are responses to lawlessness; biologists comport themselves as they do because they see that there are no laws. To defend this argument, Beatty must show, not just that biologists act this way, but that this fails to be true in sciences where laws are thought to exist.

Carrier (1995, 88) puts his finger on what is wrong with this claim when he considers the (derived) law of free fall in physics; this says that a body near the earth's surface falls with constant acceleration, provided that no force other than gravity acts upon it. Carrier points out that "every parachutist constitutes an exception" to this law, not because parachutists show that the statement is false, but because they violate the condition specified in the law's antecedent. A bowling ball and a feather exhibit different trajectories when released above the earth's surface because air resistance is an important influence on one, but not on the other. Physicists and biologists both investigate which forces are significant influences on what happens (Sober 1996). And when it comes to feathers, physics teaches us to be pluralists—to see both gravity and air resistance as important influences on the resulting trajectory. When scientists entertain questions about relative significance, and when they claim that a phenomenon has a plurality of causes, this does not show that their subject lacks laws.

This point becomes clearer when one examines what relative significance controversies in biology are about. One example that Beatty mentions—neutrality versus selection as a theory of molecular evolution—is representative. The issue here is whether $Ns \ll 1$ (Kimura 1983). If the product of the effective population size and the selection coefficient attaching to a gene is much less than unity, the gene is said to be "effectively neutral." This question concerns the contingent values that parameters happen to have. The problem is not which general model is true. Judged as a set of if/then statements, Kimura's model of neutral evolution is not in dispute. The model's truth does not depend on any evolutionary contingency.

Beatty thinks that theoretical pluralism in biology is strongly at odds with what he calls "the Newtonian tradition" (p. 68), whose guiding ideas are summarized in Newton's rules of reasoning in philosophy. Newtonians believe the maxim "to the same natural effects we must, as far as possible, assign the same causes." Pluralists, on the other hand, maintain that effects frequently have many causes. My reaction to this point is that pluralists can be good Newtonians and that this contrast does not represent a methodological rift between biology and physics. In both sciences, a defeasible preference for monism is perfectly compatible with a *de facto* embracing of pluralism. Newton said we should, *as far as possible*, prefer more monistic theories over more pluralistic ones. Pluralism in biology involves no rejection of this advice. We prefer monistic theories unless the data force us to embrace pluralism. But if the data *do* have this character, we *should* be pluralists (Forster and Sober 1994, Sober 1996).

One example of Newtonianism in biology may be found in the use of parsimony as a criterion in phylogenetic inference (Sober 1988). Why do the mammalian species we presently observe have hair? It is conceivable that hair evolved independently in every extant species, but this would be dreadfully unparsimonious. It is far more plausible to see the trait as a homology—an inheritance from a common ancestor (Nelson and Platnick 1981, 39). This does not mean that all similarities must be explained in this way. Rather, we should try to interpret similarities as homologies *as far as possible*. When we cannot, we embrace the hypothesis that some traits originated more than once. It is a mistake to think that parsimony is relevant to the search for laws, whereas pluralism is appropriate when one inquires into the character of historical particulars. In both types of science, parsimony is desirable, but defeasible.

4 Supervenience

Rosenberg's brief for lawlessness rests on an entirely different set of arguments than Beatty's. Rosenberg (1994) uses the idea of supervenience to argue that, with one exception, there are no laws in biology. The one genuine law is what Rosenberg calls "the theory of natural selection," by which he means Mary Williams' (1970) axiomatization. Rosenberg represents this axiomatization as saying that (i) there is an upper bound on the number of organisms in a generation, (ii) each organism has a fitness value, (iii) fitter traits increase in frequency and less fit traits decline, and (iv) populations show variation in fitness unless they are on the brink of extinction (p. 106).

I want to raise two questions about this axiomatization. Proposition (iv) is probably true, but I do not see why the existence of variation in fitness should be regarded as a law. Statement (iii) is false if fitness means expected number of offspring; and if fitness means actual number of offspring, it also is false, since traits with higher reproductive outputs can fail to increase if they are not heritable or if there is a counterbalancing mutation or migration pressure. Williams and Rosenberg do not spell out what they mean by "fitness" because they think that philosophical problems (e.g., what to do about the claim that the theory of evolution is tautologous) can be solved by regarding "fitness" as an undefined primitive. However, it needs to be said that interpretive problems about the fitness concept are not solved by refusing to say what the term means. If the term is primitive in an axiomatization, then it is not a defined term *in that system*; this does not remove the need to clarify what the term means in the mouths of biologists (Mills and Beatty 1979).

Anyway, Rosenberg's argument about the rest of biology is the main subject I want to discuss. Rosenberg argues that the supervenience of biology on physics shows that there are no biological laws (aside from the law he thinks is captured in Williams' axiomatization), or that we will never be able to discover any laws, should they exist. Consider the accompanying figure, adapted from Fodor 1975. Suppose P and Q are

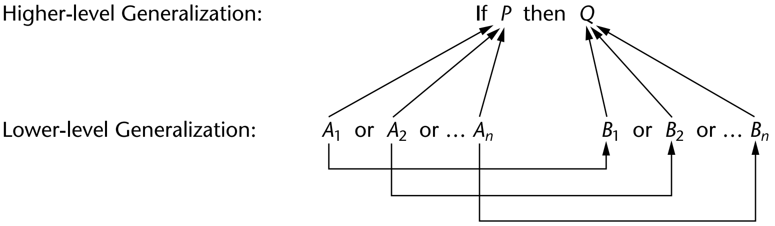


Figure 12.1

predicates in a higher-level science, such as biology or psychology; P supervenes on properties $A_1, A_2, \ldots, A_n$, while Q supervenes on properties $B_1, B_2, \ldots, B_n$. The A and B properties are studied in some lower-level science, physics perhaps. Roughly speaking, supervenience means determination; if one of the A_i's is present, then so is P, and if one of the B_i's is present, so is Q. P and Q are said to be "multiply realizable," two objects may both have P and still be different from the point of view of the lower-level theory, in that one has A_i while the other has A_j $(i \neq j)$. The higher-level predicate describes what these objects have in common, something the lower level-theory cannot do.

This diagram suggests an argument for the lawfulness of [if P then Q]. If each A_i necessitates its counterpart B_i, and if P entails that one of the A_i's must be present, then P necessitates Q. Higher-level generalizations are laws in virtue of the lawfulness of the lower-level generalizations on which they supervene. This does not show that we will be able to discover that [if P then Q] is true and lawful. Rather, the argument suggests that the law exists. My point in describing this argument is not to endorse it, but to raise the question of how Rosenberg manages to use supervenience in defense of biological lawlessness. Rosenberg thinks that chemistry supervenes on physics, but that chemical laws exist and can be discovered. Why does he think that biology is different?

Rosenberg's answer is that the process of natural selection has made the world especially complicated. There are many, many physical structures that perform the same function. Since natural selection selects for traits that perform a given function, and is indifferent as to which structure evolves to do the job, we should expect an immense proliferation of supervenience bases in biology. Selection has led prey organisms to be able to evade their predators; however, the physical properties that permit prey organisms to do this are enormously varied. The evolutionary process has made life so complicated that biology will never be able to arrive at laws. As a result, biology is and will remain an "instrumental" discipline.

One gap in Rosenberg's argument is that he does not tell us *how* complicated the living world is, or *how* complicated it has to be to elude our search for laws. I am not asking for a precise measure of complexity, but for a reason to think that the complexity of nature puts biological laws beyond our ken. Consider, for example, what we know about fitness. Fitness is the supervenient biological property *par excellens*. What do a fit zebra, a fit dandelion, and a fit bacterium have in common? Presumably, nothing much at the level of their physical properties. However, this has not prevented evolutionists from theorizing about fitness. I have already mentioned Fisher's theorem and there are lots of other lawful generalizations that describe the sources and consequences of fitness differences (Sober 1984). It might be objected that these generalizations are *a priori*, and so are not laws, properly speaking. This raises the question of whether laws must be empirical, but let us put that issue aside. If the multiple realizability of a property makes it ''complicated,'' then fitness is complicated. And if the complexity of a property makes it impossible for us to discover qualitative, counterfactual supporting, and explanatory generalizations about the property, then we should have none available about fitness. But we do, as Rosenberg concedes. The human mind does not slam shut in the face of radical multiple realizability. Understanding the sources and consequences of fitness differences is not rendered impossible by the fact that fitness is multiply realizable. It is therefore puzzling why the multiple realizability of other biological properties should mean that we will never know any laws about them.

The above diagram suggests a diagnosis of why Rosenberg thinks that multiple realizability makes supervenient laws unknowable. Perhaps Rosenberg assumes that a supervenient law can be known only by knowing the laws on which it supervenes. If there are 10,000 lower-level generalizations of the form [if A_i then B_i], then there is a lot to know, perhaps more than our frail minds can absorb. However, this argument involves a misinterpretation of the diagram. The diagram does not depict what one must do to discover that [if P then Q] is true and lawful. Rather, it represents the metaphysics of how higher-level and lower-level generalizations might be related. It seems to me that higher-level facts can be known without exhaustively examining their lower-level bases. If so, Rosenberg's ''argument from supervenience'' fails.

5 Three Biological Examples

Rosenberg has another argument for lawlessness in biology. He examines three biological areas and in each case defends an instrumentalist interpretation. The areas are classical genetics, the theory of neutral evolution, and the units of selection problem. It turns out that Rosenberg uses the term ''instrumentalism'' ambiguously. In discussing classical genetics, he claims that Mendel's ''laws'' are *false*, and so are not laws at all. However, Rosenberg does not similarly argue that Kimura's theory of neutral evolution is false. Rather, he claims that the theory's use of probability concepts reflects its *observer-relativity*. The reason that probability is used to describe drift is not that this process is objectively chancy; rather, we talk of chance only because we are ignorant of physical details.

Rosenberg's observer-relativity argument for instrumentalism confuses semantics and pragmatics. What a statement means should not be confused with how and why it is used. Perhaps we use a probability statement to make a prediction only because we are ignorant of finer-grained details. However, this does not mean that the statement is observer-relative in what it says. Consider, for example, the interpretation of probability that equates probability with actual frequency. The actual frequency of an event in a containing population is not observer-dependent, whatever our reasons may be for using such probabilities to make predictions. This undercuts Rosenberg's argument for an instrumental interpretation of the theory of neutral evolution.

Rosenberg's last example concerns the units of selection controversy. He uses a strengthened version of an idea about causality that Sober and Lewontin (1982) defended. This is the idea that C is a positive causal factor for bringing about E precisely when C raises the probability of E in at least one background context, and does not lower it in any. For example, smoking is said to be a positive causal factor for lung cancer, if smoking increases some people's chances of getting cancer and does not lower anyone else's. Lewontin and I intended the range of background contexts to be the ones that are actually exemplified in the population. However, Rosenberg expands this set to include background contexts that are merely conceivable. It is no surprise that causal claims that seem to be true turn out to be false under his strengthened criterion. Just imagine a science fiction circumstance in which smoking actually reduces the chance of lung cancer, e.g., by causing physicians to supply a preventative drug.

This leads Rosenberg to conclude that organisms and groups are never units of selection, but that "properties of the genetic material required for gene expression and replication stand a chance of satisfying [the criterion for being a unit of selection]" (p. 99). Rosenberg then recognizes that biologists do not in fact impose the stringent criterion he describes. Rather, they evaluate claims about units of selection by "identifying the particular factors of the local environment that make the trait conducive to survival of the organism and its reproduction" (p. 101). Rosenberg concludes that scientists adopt a weakened criterion because it suits their instrumental goals and finite cognitive abilities. However, another diagnosis is possible. If a strong criterion never judges organisms or groups to be units of selection, perhaps this is because the criterion is wrong. Alternative conceptions of the units of selection problem exist (Sober and Wilson 1994); rarely do they have the effect of making the subject conform to Rosenberg's picture of instrumental biology.

6 Conclusion

The supervenience of biological properties—even the radical level of multiple realizability wrought by natural selection—does not show that biology is lawless or that laws cannot be known. Moreover, when biologists engage in relative significance controversies and sometimes embrace theoretical pluralism, this is not evidence that biology lacks laws. And the fact that the biological generalizations that hold at one time trace back to earlier evolutionary contingencies does not show that there are no laws of evolution. These negative remarks hold true, regardless of whether one adopts the logical empiricist notion of law or the modified idea of process law that I have suggested.

Still, the oddity remains that when one tries to state an evolutionary law precisely, the result seems always to be an *a priori* model in mathematical biology. Why has biology developed in this way, whereas physical processes seem to obey laws that are empirical? Beatty and Rosenberg try to explain this peculiar state of affairs by describing properties of the evolutionary process. Perhaps it is time to investigate the possibility that biology has no empirical laws of evolution because of the strategies of model building that biologists have adopted.

Acknowledgments

My thanks to Robin Andreasen, André Ariew, John Beatty, Robert Brandon, Wyman Brantley, Martin Carrier, Terry Horgan, Harold Kincaid, Alexander Rosenberg, Chris Stephens, and Ann Wolfe for useful discussion.

References

Anscombe, G. (1975), "Causality and Determination," in E. Sosa (ed.), *Causation and Conditionals*. Oxford: Oxford University Press, pp. 63–81.

Beatty, J. (1995), "The Evolutionary Contingency Thesis," in G. Wolters and J. Lennox (eds.), *Concepts, Theories, and Rationality in the Biological Sciences: The Second Pittsburgh-Konstanz Colloquium in the Philosophy of Science*. Pittsburgh: University of Pittsburgh Press, pp. 45–81.

Carrier, M. (1995), "Evolutionary Change and Lawlikeness," in G. Wolters and J. Lennox (eds.), *Concepts, Theories, and Rationality in the Biological Sciences: The Second Pittsburgh-Konstanz Colloquium in the Philosophy of Science*. Pittsburgh: University of Pittsburgh Press, pp. 83–97.

Dretske, F. (1977), "Laws of Nature," *Philosophy of Science* 44: 248–268.

Fisher, R. (1930), *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press.

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

Forster, M. and E. Sober (1994), "How to Tell When Simpler, More Unified, or Less *Ad Hoc* Theories will Provide More Accurate Predictions," *British Journal for the Philosophy of Science* 45: 1–35.

Kimura, M. (1983), *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Mills, S. and J. Beatty (1979), "The Propensity Interpretation of Fitness," *Philosophy of Science* 46: 263–286.

Nelson, G. and N. Platnick (1981), *Systematics and Biogeography*. New York: Columbia University Press.

Rosenberg, A. (1994), *Instrumental Biology or the Disunity of Science*. Chicago: University of Chicago Press.

Sober, E. (1984), *The Nature of Selection*. Cambridge, MA.: MIT Press.

———. (1988), *Reconstructing the Past*. Cambridge, MA.: MIT Press.

———. (1993), *Philosophy of Biology*. Boulder: Westview Press.

———. (1994), "Temporally Oriented Laws," in *From a Biological Point of View*. Cambridge: Cambridge University Press, pp. 233–252.

———. (1996), "Evolution and Optimality—Feathers, Bowling Balls, and the Thesis of Adaptationism." *1995–96 Philosophic Exchange* 26: 41–57.

Sober, E. and R. Lewontin (1982), "Artifact, Cause, and Genic Selection," *Philosophy of Science* 47: 157–180.

Sober, E. and D. Wilson (1994), "A Critical Review of Philosophical Work on the Units of Selection Problem," *Philosophy of Science* 61: 534–555.

Williams, M. (1970), "Deducing the Consequences of Evolution," *Journal of Theoretical Biology* 29: 343–385.

VII Reductionism

13 1953 and All That: A Tale of Two Sciences

Philip Kitcher

Must we geneticists become bacteriologists, physiological chemists and physicists, simultaneously with being zoologists and botanists? Let us hope so.
—H. J. Muller, 1922[1]

1 The Problem

Toward the end of their paper announcing the molecular structure of DNA, James Watson and Francis Crick remark, somewhat laconically, that their proposed structure might illuminate some central questions of genetics.[2] Thirty years have passed since Watson and Crick published their famous discovery. Molecular biology has indeed transformed our understanding of heredity. The recognition of the structure of DNA, the understanding of gene replication, transcription and translation, the cracking of the genetic code, the study of gene regulation, these and other breakthroughs have combined to answer many of the questions that baffled classical geneticists. Muller's hope—expressed in the early days of classical genetics—has been amply fulfilled.

Yet the success of molecular biology and the transformation of classical genetics into molecular genetics bequeath a philosophical problem. There are two recent theories which have addressed the phenomena of heredity. One, *classical genetics*, stemming from the studies of T. H. Morgan, his colleagues and students, is the successful outgrowth of the Mendelian theory of heredity rediscovered at the beginning of this century. The other, *molecular genetics*, descends from the work of Watson and Crick. What is the relationship between these two theories? How does the molecular theory illuminate the classical theory? How exactly has Muller's hope been fulfilled?

There used to be a popular philosophical answer to the problem posed in these three connected questions: classical genetics has been reduced to molecular genetics. Philosophers of biology inherited the notion of reduction from general discussions in

From *Philosophical Review* 93 (1984): 335–373. Copyright © 1984 Cornell University. Reprinted by permission of the publisher and author.

philosophy of science, discussions which usually center on examples from physics. Unfortunately attempts to apply this notion in the case of genetics have been vulnerable to cogent criticism. Even after considerable tinkering with the concept of reduction, one cannot claim that classical genetics has been (or is being) reduced to molecular genetics.[3] However, the antireductionist point is typically negative. It denies the adequacy of a particular solution to the problem of characterizing the relation between classical genetics and molecular genetics. It does not offer an alternative solution.

My aim in this chapter is to offer a different perspective on intertheoretic relations. The plan is to invert the usual strategy. Instead of trying to force the case of genetics into a mold, which is alleged to capture important features of examples in physics, or resting content with denying that the material can be forced, I shall try to arrive at a view of the theories involved and the relations between them that will account for the almost universal idea that molecular biology has done something important for classical genetics. In so doing, I hope to shed some light on the general questions of the structure of scientific theories and the relations which may hold between successive theories. Since my positive account presupposes that something is wrong with the reductionist treatment of the case of genetics, I shall begin with a diagnosis of the foibles of reductionism.

2 What's Wrong with Reductionism?

Ernest Nagel's classic treatment of reduction[4] can be simplified for our purposes. Scientific theories are regarded as sets of statements.[5] To reduce a theory T_2 to a theory T_1, is to deduce the statements of T_2 from the statements of T_1. If there are nonlogical expressions which appear in the statements of T_2, but do not appear in the statements of T_1, then we are allowed to supplement the statements of T_1 with some extra premises connecting the vocabulary of T_1 with the distinctive vocabulary of T_2 (so-called *bridge principles*). Intertheoretic reduction is taken to be important because the statements which are deduced from the reducing theory are supposed to be explained by this deduction.

Yet, as everyone who has struggled with the paradigm cases from physics knows all too well, the reductions of Galileo's law to Newtonian mechanics and of the ideal gas laws to the kinetic theory do not exactly fit Nagel's model. Study of these examples suggests that, to reduce a theory T_2 to a theory T_1, it suffices to deduce the laws of T_2 from a suitably modified version of T_1, possibly augmented with appropriate extra premises. Plainly, this sufficient condition is dangerously vague. I shall tolerate its vagueness, proposing that we understand the issue of reduction in genetics by using the examples from physics as paradigms of what "suitable modifications" and "appropriate extra premises" are like. Reductionists claim that the relation between classical genetics and molecular biology is sufficiently similar to the intertheoretical relations exemplified in the examples from physics to count as the same type of thing: to wit, as intertheoretical reduction.

It may seem that the reductionist thesis has now become so amorphous that it will be immune to refutation. But this is incorrect. Even when we have amended the classical model of reduction so that it can accommodate the examples that originally motivated it, the reductionist claim about genetics requires us to accept three theses:

R1: Classical genetics contains general laws about the transmission of genes which can serve as the conclusions of reductive derivations.

R2: The distinctive vocabulary of classical genetics (predicates like "① is a gene," "① is dominant with respect to ②") can linked to the vocabulary of molecular biology by bridge principles.

R3: A derivation of general principles about the transmission of genes from principles of molecular biology would explain why the laws of gene transmission hold (to the extent that they do).

I shall argue that each of the theses is false, offering this as my diagnosis of the ills of reductionism. . . .

Philosophers often identify theories as small sets of general laws. However, in the case of classical genetics, the identification is difficult and those who debate the reducibility of classical genetics to molecular biology often proceed differently. David Hull uses a characterization drawn from Dobzhansky: classical genetics is "concerned with gene differences; the operation employed to discover a gene is hybridization: parents differing in some trait are crossed and the distribution of the trait in hybrid progeny is observed."[6] This is not unusual in discussions of reduction in genetics. It is much easier to identify classical genetics by referring to the subject matter and to the methods of investigation, than it is to provide a few sentences that encapsulate the content of the theory.

Why is this? Because when we read the major papers of the great classical geneticists or when we read the textbooks in which their work is summarized, we find it hard to pick out *any* laws about genes. These documents are full of informative statements. Together, they tell us an enormous amount about the chromosomal arrangement of particular genes in particular organisms, about the effect on the phenotype of various mutations, about frequencies of recombination, and so forth. In some cases, we might explain the absence of formulations of general laws about genes (and even of reference to such laws) by suggesting that these things are common knowledge. Yet that hardly accounts for the nature of the textbooks or of the papers that forged the tools of classical genetics.

If we look back to the pre-Morgan era, we do find two general statements about genes, namely Mendel's Laws (or "Rules"). Mendel's second law states that, in a diploid organism which produces haploid gametes, genes at different loci will be transmitted independently; so, for example, if *A*, *a*, and *B*, *b* are pairs of alleles at different loci, and if an organism is heterozygous at both loci, then the probabilities that a gamete will receive any of the four possible genetic combinations, *AB*, *Ab*, *aB*, *ab*, are all equal.[7] Once it was recognized that genes are (mostly) chromosomal segments (as biologists discovered soon after the rediscovery of Mendel's laws), we understand that the law will not hold in general: alleles which are on the same chromosome (or, more exactly, close together on the same chromosome) will tend to be transmitted together because (ignoring recombination) one member of each homologous pair is distributed to a gamete.

Now it might seem that this is not very important. We could surely find a correct substitute for Mendel's second law by restricting the law so that it only talks about genes on nonhomologous chromosomes. Unfortunately, this will not quite do. There can be interference with normal cytological processes so that segregation of nonhomologous chromosomes need not be independent. However, my complaint about Mendel's second law is not that it is incorrect: many sciences use laws that are clearly recognized as approximations. Mendel's second law, amended or unamended, simply becomes irrelevant to subsequent research in classical genetics.

We envisaged amending Mendel's second law by using elementary principles of cytology, together with the identification of genes as chromosomal segments, to correct what was faulty in the unamended law. It is the fact that the application is so easy and that it can be carried out far more generally that makes the "law" it generates irrelevant. We can understand the transmission of genes by analyzing the cases that interest us from a cytological perspective—by proceeding from "first principles," as it were. Moreover, we can adopt this approach whether the organism is haploid, diploid, or polyploid, whether it reproduces sexually or asexually, whether the genes with which we are concerned are or are not on homologous chromosomes, whether or not there is distortion of independent chromosomal segregation at meiosis. Cytology not only teaches us that the second law is false; is also tells us how to tackle the problem at which the second law was directed (the problem of determining frequencies for pairs of genes in gametes). The amended second law is a restricted statement of results obtainable using a general technique. What figures largely in genetics after Morgan is the technique, and this is hardly surprising when we realize that one of the major research problems of classical genetics has been the problem of discovering the distribution of genes *on the same chromosome*, a problem which is beyond the scope of the amended law.

Let us now turn from R1 to R2, assuming, contrary to what has just been argued, that we can identify the content of classical genetics with general principles about gene transmission. (Let us even suppose, for the sake of concreteness, that the principles in question are Mendel's laws—amended in whatever way the reductionist prefers.) To derive these principles from molecular biology, we need a bridge principle. I shall consider first statements of the form

(*) (x) (x is a gene $\leftrightarrow$ Mx),

where Mx is an open sentence (possibly complex) in the language of molecular biology. Molecular biologists do not offer any appropriate statement. Nor do they seem interested in providing one. I claim that no appropriate bridge principle can be found.

Most genes are segments of DNA. (There are some organisms—viruses—whose genetic material is RNA; I shall henceforth ignore them.) Thanks to Watson and Crick, we know the molecular structure of DNA. Hence the problem of providing a statement of the above form becomes that of saying, in molecular terms, which segments of DNA count as genes.

Genes come in different sizes, and, for any given size, we can find segments of DNA of that size that are not genes. Therefore genes cannot be identified as segments of DNA containing a particular number of nucleotide pairs. Nor will it do to give a molecular characterization of those codons (triplets of nucleotides) that initiate and terminate transcription, and take a gene to be a segment of DNA between successive initiating and terminating codons. In the first place, mutation might produce a *single* allele containing within it codons for stopping and restarting transcription. Second, and much more important, the criterion is not general since not every gene is transcribed on mRNA.

The latter point is worth developing. Molecular geneticists recognize regulatory genes as well as structural genes. To cite a classic example, the operator region in the *lac* operon of *E. coli* serves as a site for the attachment of protein molecules, thereby inhibiting transcription of mRNA and regulating enzyme production. Moreover, it is becoming increasingly obvious that genes are not always transcribed, but play a variety of roles in the economy of the cell.

At this point, the reductionist may try to produce a bridge principle by brute force. Trivially, there are only a finite number of terrestrial organisms (past, present, and future) and only a finite number of genes. Each gene is a segment of DNA with a particular structure and it would be possible, in principle, to provide a detailed molecular description of that structure. We can now give a molecular specification of the gene by enumerating the genes and disjoining the molecular descriptions. The point made above, that the segments which we count as genes do not share any structural property, can now be put more precisely: any instantiation of (*) which replaces M by a structural predicate from the language of molecular biology will insert a predicate that is essentially disjunctive.

Why does this matter? Let us imagine a reductionist using the enumerative strategy to deduce a general principle about gene transmission. After great labor, it is revealed that all actual genes satisfy the principle. I claim that more than this is needed to reduce a *law* about gene transmission. We envisage laws as sustaining counterfactuals, as applying to examples that might have been but which did not actually arise. To reduce the law it is necessary to show how possible but nonactual genes would have satisfied it. Nor can we achieve the reductionist's goal by adding further disjuncts to the envisaged bridge principle. For although there are only finitely many *actual* genes, there are indefinitely many genes which *might* have arisen.

At this point, the reductionist may protest that the deck has been stacked. There is no need to produce a bridge principle of the form (*). Recall that we are trying to derive a general law about the transmission of genes, whose paradigm is Mendel's second law. Now the gross logical form of Mendel's second law is:

$$(x)\ (y)\ ((Gx\ \&\ Gy) \rightarrow Axy). \tag{1}$$

We might hope to obtain this from statements of the forms

$$(x)\ (Gx \rightarrow Mx) \tag{2}$$

$$(x)\ (y)\ ((Mx\ \&\ My) \rightarrow Axy) \tag{3}$$

where Mx is an open sentence in the language of molecular biology. Now there will certainly be true statements of the form (2): for example, we can take Mx as "x is composed of DNA v x is composed of RNA." The question is whether we can combine some such statement with other appropriate premises—for example, some instance of (3)—so as to derive, and thereby explain (1). No geneticist or molecular biologist has advanced any suitable premises, and with good reason. We discover true statements of the form (2) by hunting for weak necessary conditions on genes, conditions which have to be met by genes but which are met by hordes of other biological entities as well. We can only hope to obtain *weak* necessary conditions because of the phenomenon that occupied us previously: from the molecular standpoint, genes are not distinguished by any common structure. Trouble will now arise when we try to show that the weak necessary condition is jointly sufficient for the satisfaction of the property (independent assortment at meiosis) that we ascribe to genes. The difficulty is illustrated by the example given above. If we take Mx to be "x is composed of DNA v x is composed of RNA" then the challenge will be to find a general law governing the distribution of all segments of DNA and RNA!

I conclude that R2 is false. Reductionists cannot find the bridge principles they need, and the tactic of abandoning the form (*) for something weaker is of no avail. I shall now consider R3. Let us concede both of the points that I have denied, allowing that there are general laws about the transmission of genes and that bridge principles are forthcoming. I claim that exhibiting derivations of the transmission laws from principles of molecular biology and bridge principles would not explain the laws, and, therefore, would not fulfill the major goal of reduction.

As an illustration, I shall use the envisaged amended version of Mendel's second law. Why do genes on nonhomologous chromosomes assort independently? Cytology provides the answer. At meiosis, chromosomes line up with their homologues. It is then possible for homologous chromosomes to exchange some genetic material, producing pairs of recombinant chromosomes. In the meiotic division, one member of each recombinant pair goes to each gamete, and the assignment of one member of one pair to a gamete is probabilistically independent of the assignment of a member of another pair to that gamete. Genes which occur close on the same chromosome are likely to be transmitted together (recombination is not likely to occur between them), but genes on nonhomologous chromosomes will assort independently.

This account is a perfectly satisfactory explanation of why our envisaged law is true to the extent that it is. (We recognize how the law could fail if there were some unusual mechanism linking particular nonhomologous chromosomes.) To emphasize the adequacy of the explanation is not to deny that it could be extended in certain ways. For example, we might want to know more about the mechanics of the process by which the chromosomes are passed on to the gametes. In fact, cytology provides such information. However, appeal to molecular biology would not deepen our understanding of the transmission law. Imagine a successful derivation of the law from principles of chemistry and a bridge principle of the form (*). In charting the details of the molecular rearrangements the derivation would only blur the outline of a simple cytological story, adding a welter of irrelevant detail. Genes on nonhomologous chromosomes assort independently because nonhomologous chromosomes are transmitted independently at meiosis, and, so long as we recognize this, we do not need to know what the chromosomes are made of.

In explaining a scientific law, L, one often provides a deduction of L from other principles. Sometimes it is possible to explain some of the principles used in the deduction by deducing them, in turn, from further laws. Recognizing the possibility of a sequence of deductions tempts us to suppose that we could produce a better explanation of L by combining them, producing a more elaborate derivation in the language of our ultimate premises. But this is incorrect. What is relevant for the purposes of giving one explanation may be quite different from what is relevant for the purposes of explaining a law used in giving that original explanation. This general point is illustrated by the case at hand. . . .

There is a natural reductionist response. The considerations of the last paragraphs presuppose far too subjective a view of scientific explanation. After all, even if *we* become lost in the molecular details, beings who are cognitively more powerful than we could surely recognize the explanatory force of the envisaged molecular derivation. However, this response misses a crucial point. The molecular derivation forfeits something important.

Recall the original cytological explanation. It accounted for the transmission of genes by identifying meiosis as a process of a particular kind: a process in which paired entities (in this case, homologous chromosomes) are separated by a force so that one member of each pair is assigned to a descendant entity (in this case, a gamete). Let us call processes of this kind *PS-processes*. I claim first that explaining the transmission law requires identifying PS-processes as forming a natural kind to which processes of meiosis belong, and second that PS-processes cannot be identified as a kind from the molecular point of view.

If we adopt the familiar covering law account of explanation, then we shall view the cytological narrative as invoking a law to the effect that processes of meiosis are PS-processes and as applying elementary principles of probability to compute the distribution of genes to gametes from the laws that govern PS-processes. If the illumination provided by the narrative is to be preserved in a molecular derivation, then we shall have to be able to express the relevant laws as laws in the language of molecular biology, and this will require that we be able to characterize PS-processes as a natural kind from the molecular point of view. The same conclusion, to wit that the explanatory power of the cytological account can be preserved only if we can identify PS-processes as a natural kind in molecular terms, can be reached in analogous ways if we adopt quite different approaches to scientific explanation—for example, if we conceive of explanation as specifying causally relevant properties or as fitting phenomena into a unified account of nature.

However, PS-processes are heterogeneous from the molecular point of view. There are no constraints on the molecular structures of the entities which are paired or on the ways in which the fundamental forces combine to pair them and to separate them. The bonds can be forged and broken in innumerable ways: all that matters is that there be bonds that initially pair the entities in question and that are subsequently (somehow) broken. In some cases, bonds may be formed directly between constituent molecules of the entities in question; in others, hordes of accessory molecules may be involved. In some cases, the separation may occur because of the action of electromagnetic forces or even of nuclear forces; but it is easy to think of examples in which the separation is effected by the action of gravity. I claim, therefore that PS-processes are realized in a motley of molecular ways. (I should note explicitly that this conclusion is independent of the issue of whether the reductionist can find bridge principles for the concepts of classical genetics.)

We thus obtain a reply to the reductionist charge that we reject the explanatory power of the molecular derivation simply because we anticipate that our brains will prove too feeble to cope with its complexities. The molecular account objectively fails to explain because it cannot bring out that feature of the situation which is highlighted in the cytological story. It cannot show us that genes are transmitted in the ways that we find them to be because meiosis is a PS-process and because any

PS-process would give rise to analogous distributions. Thus R3—like R1 and R2—is false.

3 The Root of the Trouble

Where did we go wrong? Here is a natural suggestion. The most fundamental failure of reductionism is the falsity of R1. Lacking an account of theories which would readily be applied to the cases of classical genetics and molecular genetics, the attempt to chart the relations between these theories was doomed from the start. If we are to do better, we must begin by asking a preliminary question: what is the structure of classical genetics?

I shall follow this natural suggestion, endeavoring to present a picture of the structure of classical genetics which can be used to understand the intertheoretic relations between classical and molecular genetics. As we have seen, the main difficulty in trying to axiomatize classical genetics is to decide what body of statements one is attempting to axiomatize. The history of genetics makes it clear that Morgan, Muller, Sturtevant, Beadle, McClintock, and others have made important contributions to genetic theory. But the statements occurring in the writings of these workers seem to be far too specific to serve as parts of a general theory. They concern the genes of particular kinds of organisms—primarily paradigm organisms, like fruit flied, bread molds, and maize. The idea that classical genetics is simply a heterogeneous set of statements about dominance, recessiveness, position effect, nondisjunction, and so forth, in *Drosophila, Zea mays, E. coli, Neurospora*, etc. flies in the face of our intuitions. The statements advanced by the great classical geneticists seem more like *illustrations* of the theory than *components* of it. (To know classical genetics it is not necessary to know the genetics of any particular organism, not even *Drosophila melanogaster*.) But the only alternative seems to be to suppose that there are general laws in genetics, never enunciated by geneticists but reconstructible by philosophers. At the very least, this supposition should induce the worry that the founders of the field, and those who write the textbooks of today, do a singularly bad job.

Our predicament provokes two main questions. First, if we focus on a particular time in the history of classical genetics, it appears that there will be a set of statements about inheritance in particular organisms, which constitutes the corpus which geneticists of that time accept: what is the relationship between this corpus and the version of classical genetic theory in force at the time? (In posing this question, I assume, contrary to fact, that the community of geneticists was always distinguished by unusual harmony of opinion; it is not hard to relax this simplifying assumption.) Second, we think of genetic theory as something that persisted through various versions: what is the relation among the versions of classical genetic theory accepted at different times (the versions of 1910, 1930, and 1950, for example) which makes us want to count them as versions of the same theory?

We can answer these questions by amending a prevalent conception of the way in which we should characterize the state of a science at a time. The corpus of statements about the inheritance of characteristics accepted at a given time is only one component of a much more complicated entity that I shall call the *practice* of classical genetics at that time. There is a common language used to talk about hereditary phenomena, a set of accepted statements in that language (the corpus of beliefs about inheritance mentioned above), a set of questions taken to be the appropriate questions to ask about hereditary phenomena, and a set of patterns of reasoning which are instantiated in answering some of the accepted questions; (also: sets of experimental procedures and methodological rules, both designed for use in evaluating proposed answers; these may be ignored for present purposes). The practice of classical genetics at a time is completely specified by identifying each of the components just listed.[8]

A pattern of reasoning is a sequence of *schematic sentences*, that is sentences in which certain items of nonlogical vocabulary have been replaced by dummy letters, together with a set of *filling instructions* which specify how substitutions are to be made in the schemata to produce reasoning which instantiates the pattern. This notion of pattern is intended to explicate the idea of the common structure that underlies a group of problem-solutions.

The foregoing definitions enable us to answer the two main questions I posed above. Beliefs about the particular genetic features of particular organisms illustrate or exemplify the version of genetic theory in force at the time in the sense that these beliefs figure in particular problem-solutions generated by the current practice. Certain patterns of reasoning are applied to give the answers to accepted questions, and, in making the application, one puts forward claims about inheritance in particular organisms. Classical genetics persists as a single theory with different versions at different times in the sense that different practices are linked by a chain of practices along which there are relatively small modifications in language, in accepted questions, and in the patterns for answering questions. In addition to this condition of historical connection, versions of classical genetic theory are bound by a common structure: each version uses certain expressions to characterize hereditary phenomena, accepts as important questions of a particular form, and offers a general style of reasoning for answering those questions. Specifically, throughout the career of classical genetics, the theory is directed toward answering questions about the distribution of characteristics in successive generations of a genealogy, and it proposes to answer those questions by using the probabilities of chromosome distribution to compute the probabilities of descendant genotypes.

The approach to classical genetics embodied in these answers is supported by reflection on what beginning students learn. Neophytes are not taught (and never have been taught) a few fundamental theoretical laws from which genetic "theorems" are to be deduced. They are introduced to some technical terminology, which is used to advance a large amount of information about special organisms. Certain questions about heredity in these organisms are posed and answered. Those who understand the theory are those who know what questions are to be asked about hitherto unstudied examples, who know how to apply the technical language to the organisms involved in these examples, and who can apply the patterns of reasoning which are to be instantiated in constructing answers. More simply, successful students grasp general patterns of reasoning which they can use to resolve new cases.

I shall now add some detail to my sketch of the structure of classical genetics, and thereby prepare the way for an investigation of the relations between classical genetics and molecular genetics. The initial family of problems in classical genetics, the family from which the field began, is the family of *pedigree problems*. Such problems arise when we confront several generations of organisms, related by specified connections of descent, with a given distribution of one or more characteristics. The question that arises may be to understand the given distribution of phenotypes, or to predict the distribution of phenotypes in the next generation, or to specify the probability that a particular phenotype will result from a particular mating. In general, classical genetic theory answers such questions by making hypotheses about the relevant genes, their phenotypic effects, and their distribution among the individuals in the pedigree. Each version of classical genetic theory contains one or more problem-solving patterns exemplifying this general idea, but the detailed character of the pattern is refined in later versions, so that previously recalcitrant cases of the problem can be accommodated.

Each case of a pedigree problem can be characterized by a set of *data*, a set of *constraints*, and a question. In any example, the data are statements describing the distribution of phenotypes among the organisms in a particular pedigree, or a diagram conveying the same information. The level of detail in the data may vary widely: at one extreme we may be given a full description of the interrelationships among all individuals and the sexes of all those involved; or the data may only provide the numbers of individuals with specific phenotypes in each generation; or, with minimal detail, we may simply be told that from crosses among individuals with specified phenotypes a certain range of phenotypes is found.

The constraints on the problem consist of general cytological information and descriptions of the chromosomal constitution of members of the species. The former will include the thesis that genes are (almost always) chromosomal segments and the principles that govern meiosis. The latter may contain a variety of statements. It may be pertinent to know how the species under study reproduces, how sexual dimorphism is reflected at the chromosomal level, the chromosome number typical of the species, what loci are linked, what the recombination frequencies are, and so forth. As in the case of the data, the level of detail (and thus of stringency) in the constraints can very widely.

Lastly, each problem contains a question that refers to the organisms described in the data. The question may take several forms: ''What is the expected distribution of phenotypes from a cross between *a* and *b*?'' (where *a*, *b* are specified individuals belonging to the pedigree described by the data), ''What is the probability that a cross between *a* and *b* will produce an individual having *P*?'' (where *a*, *b* are specified individuals of the pedigree described by the data and *P* is a phenotypic property manifested in this pedigree), ''Why do we find the distribution of phenotypes described in the data?'' and others.

Pedigree problems are solved by advancing pieces of reasoning that instantiate a small number of related patterns. In all cases the reasoning begins from a *genetic hypothesis*. The function of a genetic hypothesis is to specify the alleles that are relevant, their phenotypic expression, and their transmission through the pedigree. From that part of the genetic hypothesis that specifies the genotypes of the parents in any mating that occurs in the pedigree, together with the constraints on the problem, one computes the expected distribution of genotypes among the offspring. Finally, for any mating occurring in the pedigree, one shows that the expected distribution of genotypes among the offspring is consistent with the assignment of genotypes given by the genetic hypothesis.

The form of the reasoning can easily be recognized in examples—examples that are familiar to anyone who has ever looked at a textbook or a research report in genetics. What interests me is the style of reasoning itself. The reasoning begins with a genetic hypothesis that offers four kinds of information: (1) Specification of the number of relevant loci and the number of alleles at each locus; (2) specification of the relationships between genotypes and phenotypes; (3) specification of the relations between genes and chromosomes, of facts about the transmission of chromosomes to gametes (for example, resolution of the question whether there is disruption of normal segregation) and about the details of zygote formation; (4) assignment of genotypes to individuals in the pedigree. After showing that the genetic hypothesis is consistent with the data and constraints of the problem, the principles of cytology and the laws of probability are used to compute expected distributions of genotypes from crosses. the expected distributions are then compared with those assigned in part (4) of the genetic hypothesis.

Throughout the career of classical genetics, pedigree problems are addressed and solved by carrying out reasoning of the general type just indicated. Each version of classical genetic theory contains a pattern for solving pedigree problems with a method for computing expected genotypes which is adjusted to reflect the particular form of the genetic hypotheses that it sanctions. Thus one way to focus the differences among successive versions of classical genetic theory is to compare their conceptions of the possibilities for genetic hypotheses. As genetic theory develops, there is a changing set of conditions on admissible genetic hypotheses. Prior to the discovery of polygeny and pleiotropy (for example), part (1) of any adequate genetic hypothesis was viewed as governed by the requirement that there would be a one-one correspondence between loci and phenotypic traits.[9] After the discovery of incomplete dominance and epistasis, it was recognized that part (2) of an adequate hypothesis might take a form that had not previously been allowed: one is not compelled to assign to the heterozygote a phenotype assigned to one of the homozygotes, and one is also permitted to relativize the phenotypic effect of a gene to its genetic environment.[10] Similarly, the appreciation of phenomena of linkage, recombination, nondisjunction, segregation distortion, meiotic drive, unequal crossing over, and crossover suppression, modify conditions previously imposed on part (3) of any genetic hypothesis. In general, we can take each version of classical genetic theory to be associated with a set of conditions (usually not formulated explicitly) which govern admissible genetic hypotheses. While a general form of reasoning persists through the development of classical genetics, the patterns of reasoning used to resolve cases of the pedigree problem are constantly fine-tuned as geneticists modify their views about what forms of genetic hypothesis are allowable.

So far I have concentrated exclusively on classical genetic theory as a family of related patterns of reasoning for solving the pedigree problem. It is natural to ask if versions of the theory contain patterns of reasoning for addressing other questions. I believe that they do. The heart of the theory is the theory of *gene transmission*, the family of reasoning patterns directed at the pedigree problem. Out of this theory grow other subtheories. The theory of *gene mapping* offers a pattern of reasoning which addresses questions about the relative positions of loci on chromosomes. It is a direct result of Sturtevant's insight that one can systematically investigate the set of pedigree problems associated with a particular species. In turn, the theory of gene mapping raises the question of how to identify mutations, issues which are to be tackled by the *theory of mutation*. Thus we can think of classical genetics as having a central theory, the theory of gene transmission, which develops in the ways I have described above, surrounded by a number of satellite theories that are directed at questions arising from the pursuit of the central theory. Some of these satellite theories (for example, the theory of gene mapping) develop in the same continuous fashion. Others, like the theory of mutation, are subject to rather dramatic shifts in approach.

4 Molecular Genetics and Classical Genetics

Armed with some understanding of the structure and evolution of classical genetics, we can finally return to the question with which we began. What is the relation between classical genetics and molecular genetics? When we look at textbook presentations and the pioneering research articles that they cite, it is not hard to discern major ways in which molecular biology has advanced our understanding of hereditary phenomena. We can readily identify particular molecular explanations which illuminate issues that were treated incompletely, if at all, from the classical perspective.

What proves puzzling is the connection of these explanations to the theory of classical genetics. I hope that the account of the last section will enable us to make the connection.

I shall consider three of the most celebrated achievements of molecular genetics. Consider first the question of *replication*. Classical geneticists believed that genes can replicate themselves. Even before the experimental demonstration that all genes are transmitted to all the somatic cells of a developing embryo, geneticists agreed that normal processes of mitosis and meiosis must involve gene replication. Muller's suggestion that the central problem of genetics is to understand how mutant alleles, incapable of performing wild-type functions in producing the phenotype, are nonetheless able to replicate themselves, embodies this consensus. Yet classical genetics had no account of gene replication. A molecular account was an almost immediate dividend of the Watson-Crick model of DNA.

Watson and Crick suggested that the two strands of the double helix unwind and each strand serves as the template for the formation of a complementary strand. Because of the specificity of the pairing of nucleotides, reconstruction of DNA can be unambiguously directed by a single strand. This suggestion has been confirmed and articulated by subsequent research in molecular biology.[11] The details are more intricate than Watson and Crick may originally have believed, but the outline of their story stands.

A second major illumination produced by molecular genetics concerns the characterization of mutation. When we understand the gene as a segment of DNA we recognize the ways in which mutant alleles can be produced. "Copying errors" during replication can cause nucleotides to be added, deleted, or substituted. These changes will often lead to alleles that code for different proteins, and which are readily recognizable as mutants through their production of deviant phenotypes. However, molecular biology makes it clear that there can be *hidden* mutations, mutations that arise through nucleotide substitutions that do not change the protein produced by a structural gene (the genetic code is redundant) or through substitutions that alter the form of the protein in trivial ways. The molecular perspective provides us with a general answer to the question, "What is a mutation?" namely, that a mutation is the modification of a gene through insertion, deletion, or substitution of nucleotides. This general answer yields a basic method for tackling (in principle) questions of form, "Is *a* a mutant allele?" namely, a demonstration that *a* arose through nucleotide changes from alleles that persist in the present population. The method is frequently used in studies of the genetics of bacteria and bacteriophage, and can sometimes be employed even in inquiries about more complicated organisms. So, for example, there is good biochemical evidence for believing that some alleles which produce resistance to pesticides in various species of insects arose through nucleotide changes in the alleles naturally predominating in the population.[12]

I have indicated two general ways in which molecular biology answers questions that were not adequately resolved by classical genetics. Equally obvious are a large number of more specific achievements. Identification of the molecular structures of particular genes in particular organisms has enabled us to understand why those genes combine to produce the phenotypes they do. . . .

The claim that genes can replicate does not have the status of a central law of classical genetic theory. It is not something that figures prominently in the explanations provided by the theory (as, for example, the Boyle–Charles law is a prominent premise in some of the explanations yielded by phenomenological thermodynamics). Rather, it is a claim that classical geneticists took for granted, a claim presupposed by explanations, rather than an explicit part of them. Prior to the development of molecular genetics that claim had come to seem increasingly problematic. If genes can replicate, how do they manage to do it? Molecular genetics answered the worrying question. It provided a theoretical demonstration of the possibility of an antecedently problematic presupposition of classical genetics.

We can say that a theory presupposes a statement *p* if there is some problem-solving pattern of the theory, such that every instantiation of the pattern contains statements that jointly imply the truth of *p*. Suppose that, at a given stage in the development of a theory, scientists recognize an argument from otherwise acceptable premises which concludes that it is impossible that *p*. Then the presupposition *p* is problematic for those scientists. What they would like would be an argument showing that it is possible that *p* and explaining what is wrong with the line of reasoning which appears to threaten the possibility of *p*. If a new theory generates an argument of this sort, then we can say that the new theory gives a theoretical demonstration of the possibility of an antecedently problematic presupposition of the old theory. . . .

Because theoretical demonstrations of the possibility of antecedently problematic presuppositions involve derivation of conclusions of one theory from the premises supplied by a background theory, it is easy to assimilate them to the classical notion of reduction. However, on the account I have offered, there are two important differences. First, there is no commitment to the thesis that genetic theory can be formulated as (the deductive closure of) a conjunction of laws. Second, it is not assumed that all general statements about genes are equally in need of molecular derivation. Instead, one particular thesis, a thesis that underlies all the explanations provided by classical genetic theory, is seen as especially problematic, and the molecular derivation is viewed as addressing a specific problem that classical geneticists had already perceived. Where the reductionist identifies a general benefit in deriving all the axioms of the reduced theory, I focus on a particular derivation of a claim that has no title as an axiom of classical genetics, a derivation which responds to a particular explanatory difficulty of which classical geneticists were acutely aware. The reductionist's global relation between theories does not obtain between classical and molecular genetics, but something akin to it does hold between special fragments of these theories.

The second principal achievement of molecular genetics, the account of mutation, involves a conceptual refinement of prior theory. Later theories can be said to provide conceptual refinements of earlier theories when the later theory yields a specification of entities that belong to the extensions of predicates in the language of the earlier theory, with the result that the ways in which the referents of these predicates are fixed are altered in accordance with the new specifications. Conceptual refinement may occur in a number of ways. A new theory may supply a descriptive characterization of the extension of a predicate for which no descriptive characterization was previously available; or it may offer a new description which makes it reasonable to amend characterizations that had previously been accepted. . . .[13]

Finally, let us consider the use of molecular genetics to illuminate the action of particular genes. Here we again seem to find a relationship that initially appears close to the reductionist's ideal. Statements that are invoked as premises in particular problem-solutions—statements that ascribe particular phenotypes to particular genotypes—are derived from molecular characterizations of the alleles involved. On the account of classical genetics offered in section 3, each version of classical genetic theory includes in its schema for genetic hypotheses a clause which relates genotypes to phenotypes. . . . [W]e might hope to discover a pattern of reasoning within molecular genetics that would generate as its conclusion the schema for assigning phenotypes to genotypes.

It is not hard to characterize the relation just envisioned. Let us say that a theory T′ provides an *explanatory extension* of a theory T just in case there is some problem-solving pattern of T one of whose schematic premises can be generated as the conclusion of a problem-solving pattern of T′. When a new theory provides an explanatory extension of an old theory, then particular premises occurring in explanatory derivations given by the old theory can themselves be explained by using arguments furnished by the new theory. However, it does not follow that the explanations provided by the old theory can be improved by replacing the premises in question with the pertinent derivations. What is relevant for the purposes of explaining some statement S may not be relevant for the purposes of explaining a statement S′ which figures in an explanatory derivation of S.

Even though reductionism fails, it may appear that we can capture part of the spirit of reductionism by deploying the notion of explanatory extension. The thesis that molecular genetics provides an explanatory extension of classical genetics embodies the idea of a global relationship between the two theories, while avoiding two of the three troubles that were found to beset reductionism. That thesis does not simply assert that some specific presupposition of classical genetics (for example, the claim that genes are able to replicate) can be derived as the conclusion of a molecular argument, but offers a general connection between premises of explanatory derivations in classical genetics and explanatory arguments from molecular genetics. It is formulated so as to accommodate the failure of R1 and to honor the picture of classical genetics developed in section 3. Moreover, the failure of R2 does not affect it. . . .

Nevertheless, even born-again reductionism is doomed to fall short of salvation. Although it is true that molecular genetics belongs to a cluster of theories which, taken together, provide an explanatory extension of classical genetics, molecular genetics, on its own, cannot deliver the goods. There are some cases in which the ancillary theories do not contribute to the explanation of a classical claim about gene action. In such cases, the classical claim can be derived and explained by instantiating a pattern drawn from molecular genetics. The example of human hemoglobin provides one such case. [Individuals who are homozygous for a mutant allele for the synthesis of human hemoglobin develop sickle-cell anemia, a phenomenon that can be explained at the molecular level (ed.).] But this example is atypical.

Consider the way in which the hemoglobin example works. Specification of the molecular structures of the normal and mutant alleles, together with a description of the genetic code, enables us to derive the composition of normal and mutant hemoglobin. Application of chemistry then yields descriptions of the interactions of the proteins. With the aid of some facts about human blood cells, one can then deduce that the sickling effect will occur in abnormal cells, and, given some facts about human physiology, it is possible to derive the descriptions of the phenotypes. There is a clear analogy here with some cases from physics. The assumptions about blood cells and physiological needs seem to play the same role as the boundary conditions about shapes, relative positions, and velocities of planets that occur in Newtonian derivations of Kepler's laws. In the Newtonian explanation we can see the application of a general pattern of reasoning—the derivation of explicit equations of motion from specifications of the forces acting—which yields the general result that a body under the influence of a centrally directed inverse square force will travel in a conic section; the general result is then applied to the motions of the planets by incorporating pieces of astronomical information. Similarly, the derivation of the classical claims about the action of the normal and mutant hemoglobin genes can be seen as a purely chemical derivation of the generation of certain molecular structures and of the interactions among them. The chemical conclusions are then applied to the biological system under consideration by introducing three "boundary conditions"; first, the claim that the altered molecular structures only affect development to the extent of substituting a different molecule in the erythrocytes (the blood cells that transport hemoglobin); second, a description of the chemical conditions in the capillaries; and third, a description of the effects upon the organism of capillary blockage.

The example is able to lend comfort to reductionism precisely because of an atypical feature. In effect, one concentrates on the *differences* among the phenotypes, takes for granted the fact that in all cases development will proceed normally to the extent of manufacturing erythrocytes—which are, to all intents and purposes, simply sacks for containing hemoglobin molecules—and compares the difference in chemical effect of the cases in which the erythrocytes contain different molecules. *The details of the process of development can be ignored.* However, it is rare for the effect of a mutation to be so simple. Most structural genes code for molecules whose presence or absence make subtle differences. Thus, typically, a mutation will affect the distribution of chemicals in the cells of a developing embryo. A likely result is a change in the timing of intracellular reactions, a change that may, in turn, alter the shape of the cell. Because of the change of shape, the geometry of the embryonic cells may be modified. Cells that usually come into contact may fail to touch. Because of this, some cells may not receive the molecules necessary to switch on certain batteries of genes. Hence the chemical composition of these cells will be altered. And so it goes.

Quite evidently, in examples like this (which include most of the cases in which molecular considerations can be introduced into embryology) the reasoning that leads us to a description of the phenotype associated with a genotype will be much more complicated than that found in the hemoglobin case. It will not simply consist in a chemical derivation adapted with the help of a few boundary conditions furnished by biology. Instead, we shall encounter a sequence of subarguments: molecular descriptions lead to specifications of cellular properties, from these specifications we draw conclusions about cellular interactions, and from these conclusions we arrive at further molecular descriptions. There is clearly a pattern of reasoning here which involves molecular biology and which extends the explanations furnished by classical genetics by showing how phenotypes depend upon genotypes—but I think it would be folly to suggest that the extension is provided by molecular genetics alone.

In section 2, we discovered that the traditional answer to the philosophical question of understanding the relation that holds between molecular genetics and classical genetics, the reductionist's answer, will not do. Section 3 attempted to build on the diagnosis of the ills of reductionism, offering an account of the structure and evolution of classical genetics that would improve on the picture offered by those who favor traditional approaches to the nature of scientific theories. In the present section, I have tried to use the framework of section 3 to understand the relations between molecular genetics and classical genetics. Molecular genetics has done something important for classical genetics, and its achievements can be recognized by seeing them as instances of the intertheoretic relations that I have characterized. Thus I claim that the problem from which we began is solved.

So what? Do we have here simply a study of a particular case—a case which has, to be sure, proved puzzling for the usual accounts of scientific theories and scientific change? I hope not. Although the traditional approaches may have proved helpful in understanding some of the well-worn examples that have been the stock-in-trade of twentieth-century philosophy of science, I believe that the notion of scientific practice sketched in section 3 and the intertheoretic relations briefly characterized here will both prove helpful in analyzing the structure of science and the growth of scientific knowledge *even in those areas of science where traditional views have seemed most successful*. Hence the tale of two sciences which I have been telling is not merely intended as a piece of local history that fills a small but troublesome gap in the orthodox chronicles. I hope that it introduces concepts of general significance in the project of understanding the growth of science.

5 Antireductionism and the Organization of Nature

One loose thread remains. The history of biology is marked by continuing opposition between reductionists and antireductionists. Reductionism thrives on exploiting the charge that it provides the only alternative to the mushy incomprehensibility of vitalism. Antireductionists reply that their opponents have ignored the organismic complexity of nature. Given the picture painted above, where does this traditional dispute now stand?

I suggest that the account of genetics which I have offered will enable reductionists to provide a more exact account of what they claim, and will thereby enable antireductionists to be more specific about what they are denying. Reductionists and antireductionists agree in a certain minimal physicalism. To my knowledge, there are no major figures in contemporary biology who dispute the claim that each biological event, state, or process is a complex physical event, state, or process. The most intricate part of ontogeny or phylogeny involves countless changes of physical state. What antireductionists emphasize is the organization of nature and the "interactions among phenomena at different levels." The appeal to organization takes two different forms. When the subject of controversy is the proper form of evolutionary theory, then antireductionists contend that it is impossible to regard all selection as operating at the level of the gene.[14] What concerns me here is not this area of conflict between reductionists and their adversaries, but the attempt to block claims for the hegemony of molecular studies in understanding the physiology, genetics, and development of organisms.

A sophisticated reductionist ought to allow that, in the current practice of biology, nature is divided into levels which form the proper provinces of areas of biological study: molecular biology, cytology, histology, physiology, and so forth. Each of these sciences can be thought of as using certain language to formulate the questions it deems important and as supplying patterns of reasoning for resolving those questions. Reductionists can now set forth one of two main claims. The stronger thesis is that the explanations provided by any biological theories can be reformulated in the language of molecular biology and be recast so as to instantiate the patterns of reasoning supplied by molecular biology. the weaker thesis is that molecular biology provides explanatory extension of the other biological sciences.

Strong reductionism falls victim to the considerations that were advanced against R3. The distribution of genes to gametes is to be explained, not by rehearsing the gory details of the reshuffling of the molecules, but through the observation that chromosomes are aligned in pairs just prior to the meiotic division, and that one chromosome from each matched pair is transmitted to each gamete. We may formulate this point in the biologists' preferred idiom by saying that the assortment of alleles is to be understood at the cytological level. What is meant by this description is that there is a pattern of reasoning which is applied to derive the description of the assortment of alleles and which involves predicates that characterize cells and their large-scale internal structures. That pattern of reasoning is to be objectively preferred to the molecular pattern which would be instantiated by the derivation that charts the complicated rearrangement of individual molecules because it can be applied across a range of cases which would look heterogeneous from a molecular perspective. Intuitively, the cytological pattern makes connections which are lost at the molecular level, and it is thus to be preferred.

So far, antireductionism emerges as the thesis that there are *autonomous levels of biological explanation*. Antireductionism construes the current division of biology not simply as a temporary feature of our science stemming from our cognitive imperfections but as the reflection of levels of organization in nature. Explanatory patterns that deploy the concepts of cytology will endure in our science because we would foreswear significant unification (or fail to employ the relevant laws, or fail to identify the causally relevant properties) by attempting to derive the conclusions to which they are applied using the vocabulary and reasoning patterns of molecular biology. But the autonomy thesis is only the beginning of antireductionism. A stronger doctrine can be generated by opposing the weaker version of sophisticated reductionism.

In section 4, I raised the possibility that molecular genetics may be viewed as providing an explanatory extension of classical genetics through deriving the schematic sentence that assigns phenotypes to genotypes from a molecular pattern of reasoning. This apparent possibility fails in an instructive way. Antireductionists are not only able to contend that there are autonomous levels of biological explanation. They can also resist the weaker reductionist view that explanation always flows from the molecular level up. Even if reductionists retreat to the modest claim that, while there are autonomous levels of explanation, descriptions of cells and their constituents are always explained in terms of descriptions about genes, descriptions of tissue geometry are always explained in terms of descriptions of cells, and so forth, antireductionists can resist the picture of a unidirectional flow of explanation. Understanding the phenotypic manifestation of a gene, they will maintain, requires constant shifting back and forth across levels. Because developmental processes are complex and because changes in the timing of embryological events may produce a cascade of effects at several different levels, one sometimes uses descriptions at higher levels to explain what goes on at a more fundamental level. . . .

It would be premature to claim that I have shown how to reformulate the antireductionist appeals to the organization of nature in a completely precise way. My conclusion is that, to the extent that we can make sense of the present explanatory structure within biology—that division of the field into subfields corresponding to levels of organization in nature—we can also understand the antireductionist doctrine. In its minimal form, it is the claim that the commitment to several explanatory levels does not simply reflect our cognitive limitations; in its stronger form, it is the thesis that some explanations oppose the direction of preferred reductionistic explanation. Reductionists should not dismiss these doctrines as incomprehensible mush unless they are prepared to reject as unintelligible the biological strategy of dividing the field (a strategy which seems to me well understood, even if unanalyzed).

The examples I have given seem to support both antireductionist doctrines. To clinch the case, further analysis is needed. The notion of explanatory levels obviously cries out for explication, and it would be illuminating to replace the informal argument that the unification of our beliefs is best achieved by preserving multiple explanatory levels with an argument based on a more exact criterion for unification. Nevertheless, I hope that I have said enough to make plausible the view that, despite the immense value of the molecular biology that Watson and Crick launched in 1953, molecular studies cannot cannibalize the rest of biology. Even if geneticists must become ''physiological chemists'' they should not give up being embryologists, physiologists, and cytologists.

Notes

Earlier versions of this chapter were read at Johns Hopkins University and at the University of Minnesota, and I am very grateful to a number of people for comments and suggestions. In particular, I would like to thank Peter Achinstein, John Beatty, Barbara Horan, Patricia Kitcher, Richard Lewontin, Kenneth Schaffner, William Wimsatt, an anonymous reader, and the editors of the *Philosophical Review*, all of whom have had an important influence on the final version. Needless to say, these people should not be held responsible for residual errors. I am also grateful to the American Council of Learned Societies and the Museum of Comparative Zoology at Harvard University for support and hospitality while I was engaged in research on the topics of this paper.
[This article has been abridged. (ed.).]

1. ''Variation due to Change in the Individual Gene,'' reprinted in J. A. Peters ed., *Classic Papers in Genetics* (Englewood Cliffs, N.J.: Prentice-Hall, 1959), pp. 104–116. Citation from p. 115.

2. ''Molecular Structure of Nucleic Acids,'' *Nature* 171 (1953), pp. 737–738; reprinted in Peters, op. cit., pp. 241–243. Watson and Crick amplified their suggestion in ''Genetic Implications of the Structure of Deoxyribonucleic Acid,'' *Nature* 171 (1953), pp. 934–937.

3. The most sophisticated attempts to work out a defensible version of reductionism occur in articles by Kenneth Schaffner.

4. E. Nagel, *The Structure of Science* (New York: Harcourt Brace, 1961), Chapter 11. A simplified presentation can be found in Chapter 8 of C. G. Hempel, *Philosophy of Natural Science* (Englewood Cliffs, N.J.: Prentice-Hall, 1966).

5. Quite evidently, this is a weak version of what was once the "received view" of scientific theories, articulated in the works of Nagel and Hempel cited in the previous note.

6. Hull, *Philosophy of Biological Science* (Englewood Cliffs, N.J.: Prentice-Hall, 1974), p. 23, adapted from Theodosius Dobzhansky, *Genetics of the Evolutionary Process* (New York: Columbia University Press, 1970), p. 167.

7. A *locus* is the place on a chromosome occupied by a gene. Different genes which can occur at the same locus are said to be *alleles*. In diploid organisms, chromosomes line up in pairs just before the meiotic division that gives rise to gametes. The matched pairs are pairs of *homologous chromosomes*. If different alleles occur at corresponding loci on a pair of homologous chromosomes, the organism is said to be *heterozygous* at these loci.

8. My notion of a practice owes much to some neglected ideas of Sylvain Bromberger and Thomas Kuhn. See, in particular, Bromberger, "A Theory about the Theory of Theory and about the Theory of Theories" (W. L. Reese ed., *Philosophy of Science*, *The Delaware Seminar*, New York, 1963); and "Questions" (*Journal of Philosophy* 63 [1966], pp. 597–606); and Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962) Chapters II–V. The relation between the notion of a practice and Kuhn's conception of a paradigm is discussed in Chapter 7 of my book *The Nature of Mathematical Knowledge* (New York: Oxford University Press, 1983).

9. *Polygeny* occurs when many genes affect one characteristic; *pleiotropy* occurs when one gene affects more than one characteristic.

10. *Incomplete dominance* occurs when the phenotype of the heterozygote is intermediate between that of the homozygotes; *epistasis* occurs when the effect of a particular combination of alleles at one locus depends on what alleles are present at another locus.

11. See Watson, *Molecular Biology of the Gene* (Menlo Park, CA: W. A. Benjamin, 1976), Chapter 9; and Arthur Kornberg, *DNA Synthesis* (San Francisco: W. H. Freeman, 1974).

12. See. G. P. Georghiou, "The Evolution of Resistance to Pesticides," *Annual Review of Ecology and Systematics* 3 (1972), pp. 133–168.

13. There are numerous examples of such modifications from the history of chemistry. I try to do justice to this type of case in "Theories, Theorists, and Theoretical Change," *Philosophical Review* 87 (1978), pp. 519–547 and in "Genes," *British Journal for the Philosophy of Science* 33 (1982), pp. 337–359.

14. The extreme version of reductionism is defended by Richard Dawkins in *The Selfish Gene* (New York: Oxford University Press, 1976) and *The Extended Phenotype* (San Francisco: W. H. Freeman, 1982). For an excellent critique, see Elliott Sober and Richard C. Lewontin, "Artifact, Cause, and Genic Selection," *Philosophy of Science* 49 (1982), pp. 157–180.

14 Why the Antireductionist Consensus Won't Survive the Case of Classical Mendelian Genetics

C. Kenneth Waters

Philosophers now treat the relationship between Classical Mendelian Genetics and molecular biology as a paradigm of nonreduction and this example is playing an increasingly prominent role in debates about the reducibility of theories ranging from macrosocial science to folk psychology. Patricia Churchland (1986), for example, draws an analogy between the alleged elimination of the "causal mainstay" of classical genetics and her view that today's psychological theory will be eliminated by neuroscience. Patricia Kitcher takes an autonomous rather than eliminativist view of the reported nonreduction in genetics and reasons that psychology will retain a similar autonomy from lower level sciences (1980, 1982). Although Churchland and Kitcher offer different interpretations of the apparent failure of molecular biology to reduce classical genetics, they agree that this failure will help illuminate theoretical relations between psychology and lower level sciences. The appearance of the Mendelian example alongside the usual ones from physics and chemistry marks a turning point in philosophy of science. Philosophers now look to biology in general, and the case of genetics in particular, for insights into the nature of theoretical relations. If I am correct, however, the current autireductionist consensus about genetics is mistaken and threatens to misguide our attempt to understand relations between other scientific theories. My aim is to defuse the arguments offered in support of the antireductionist consensus. Although the question of whether molecular biology is reducing Classical Mendelian Genetics will not be settled in any single chapter, my critical analysis will reveal the signs of a significant theoretical reduction and uncover issues relevant to gaining a better understanding of what is now happening in genetics and of what we might expect to occur in other sciences.

The current consensus among philosophers is that, despite the appearances, Classical Mendelian Genetics (hereafter called CMG) is not being reduced to molecular

From *PSA 1990*, Philosophy of Science Association, Volume 1, 125–139. © 1990 by the Philosophy of Science Association. All rights reserved. Reprinted by permission of The University of Chicago Press.

biology, at least *not in the spirit of Nagel's (1961) postpositivist conception of theoretical reduction*[1] (Hull 1972 and 1974, Wimsatt 1976, Maull 1977, Darden and Maull 1977, Hooker 1981, Kitcher 1984, and Rosenberg 1985, but Schaffner 1969 and 1976, Ruse 1976, and Richardson 1979 and 1982 disagree). There are important differences within the consensus view, but according to the general antireductionist thrust, the relations between the levels of organization represented by the classical and molecular theories are too complex to be connected in the systematic way essential for a successful theoretical reduction. Antireductionists support this view by arguing that the gene concepts of the respective theories cannot be linked in an appropriate way. If the concepts cannot be linked, the reasoning goes, neither can the theoretical claims couched in terms of them. Hence, reduction will never be achieved. Before considering the antireductionists' arguments in greater detail, I will briefly describe the conception of reduction at issue and review CMG and the molecular theory of the gene.

Preliminaries

The Spirit of Postpositivist Reduction

The consensus against reductionism in genetics has focused on Nagel's (1961) formal analysis of theoretical reduction. One of the two formal requirements set out by Nagel was that the laws of the reduced theory must be derivable from the laws and associated coordinating definitions of the reducing theory. The second formal requirement was that all terms of the reduced theory must either be contained within or be appropriately connected to the reducing theory by way of "additional assumptions." It is this condition of connectability that proponents of the consensus think cannot be satisfied in the case of genetics because of the contrasting gene concepts in the classical and molecular theories. A difficulty of relying on this formal conception is that it is couched within an account of theories discarded by most philosophers of biology.[2] In order to render the antireductionist consensus nontrivial, the spirit behind Nagel's conception of theoretical reduction will need to be separated from his formal analysis.

Nagel's discussion of nonformal conditions for reduction provides an opening for freeing his conception of theoretical reduction from his outmoded account of theories. In a section on these conditions, Nagel admitted, "The two formal conditions for reduction discussed in the previous section [connectability and derivability] do not suffice to distinguish trivial from noteworthy scientific achievements" (p. 358). He identified two sets of nonformal considerations to explain why the reduction of thermodynamics was a significant achievement. The first set concerned the establishment of new experimental laws that were in better agreement with a broader range of facts than were the original ones. The second set involved the discovery of surprising connections between various experimental laws.

Nagel's reliance on nonformal conditions indicates that he had an unarticulated notion of theoretical reduction which he failed to capture in his formal account. I would like to suggest, therefore, that his underlying conception of theoretical reduction can be separated from his formal treatment and in fact reformulated with respect to an updated account of theories. When I say that intertheoretical relations satisfy *the spirit of postpositivist reduction*, I simply mean that they would satisfy conditions set out in such a reformulation. Since postpositivists tended to view both explanation and reduction as special kinds of derivation, it is natural to suppose that their conception of theoretical reduction centered on the idea that reducing theories explain the success of reduced theories. Hence, the fundamental question for us is whether CMG is being explained by molecular biology. According to the antireductionist consensus, CMG is not and will not be systematically explained by molecular biology.

Classical Mendelian Genetics (CMG)

The consensus view concerns the reducibility of the theory of Classical Mendelian Genetics (CMG), not the reducibility of Mendel's theory. CMG was developed during the first decades of this century, in large part by Thomas Hunt Morgan and his graduate students who worked on the genetics of *Drosophila*. According to the classical theory, patterns of inheritance can be explained by postulating the existence of genes. Differences in outward appearances (or phenotypes) of organisms are explained as the result of organisms' inheriting different genes (or genotypes). Genes in *Drosophila* come in twos on corresponding pairs of linear chains. Each gene of a given pair has 50 percent chance of having a copy distributed to a particular gamete (an egg or sperm). Genes located on different (nonpaired) chains assort independently from one another. Genes located on the same chain tend to be assorted together, but are sometimes distributed separately because paired chains occasionally exchange segments. The relative positions of genes can be determined by the frequency of such exchanges (on the assumption that genes located farther apart from one another are assorted separately more often than genes located closer together). CMG concerns a wide range of gene behavior including, but not limited to, mutation, expression, interaction, recombination, and distribution.

The classical account of gene expression is complicated. In the simplest kind of system, two alleles with complete dominance, there are two contrasting phenotypic traits and two kinds of genes, one of which is dominant. Each trait is associated with one kind of gene and every organisms has two genes. If an organism has two copies of the same gene, it exhibits the trait associated with the matching genes. If an organism has a pair of contrasting genes, it exhibits the characteristic associated with the dominant gene. This is but the simplest model of gene expression; classical geneticists have constructed models to represent systems of much greater complexity.

This abstract theory has a cytological interpretation. Gene chains are identified as chromosomes. Meiosis, the process in which chromosomes are distributed to gametes, offers an explanation of segregation and assortment. During the first division of this process, homologous chromosomes pair and then separate as two daughter cells are produced. The lack of complete linkage of genes located on the same chromosome is explained in terms of the crossing over (the exchange) of chromosomal segment.[3]

The Molecular Theory

The molecular theory of the gene is based on the Watson and Crick Model of DNA. According to molecular theory, a gene is a relatively short segment of a DNA molecule, which consists of two very long chains of nucleotides held together by hydrogen bonds. The genetic information is encoded in the linear sequence of nucleotides making up individual genes. On the basis of this model and empirical studies, molecular biologists soon succeeded in explaining a number of important genetic phenomena including: gene replication; the multistep process by which the information encoded in structural genes eventually gets translated during polypeptide synthesis; and mechanisms of gene regulation. Polypeptides are the constituents of proteins and the regulation of biosynthetic pathways is for the most part directed by enzymatic proteins. Hence, the molecular explanation of how genes direct polypeptide synthesis offered an abstract picture of the biochemistry of gene expression.[4]

These successes led Kenneth Schaffner (1969) to conclude that CMG was being reduced to molecular biology. But enthusiasm for reductionism soon waned (at least among philosophers) when Michael Ruse (1971) and David Hull (1972) criticized Schaffner's specific account of the apparent reduction. Since then, these rather narrowly focused criticisms have been generalized into self-contained arguments against the general idea that CMG is being reduced. I now turn to these antireductionist objections.

Defusing the Antireductionist Objections

Arguments against the idea that CMG is being reduced (in the spirit of postpositivist reduction) fall into two general categories. The most prominent arguments are those aimed at showing that there are unbridgeable conceptual gaps between CMG and molecular biology. According to these arguments, subtle differences in the meaning of parallel terms from the classical and molecular theories obstruct reduction. The second category consists of arguments which conclude that molecular theory cannot deliver the explanatory power that reductionism requires. These arguments allegedly show that the explanatory relations between the classical and molecular theories are incomplete and that if a fuller explanation of Mendelian genetics is possible, it will come from a variety of biological fields, not just from molecular genetics as reductionism seems to imply. My critical analysis of these objections will not only show that the relationship between CMG and molecular biology is misunderstood, it will also reveal signs of a successful theoretical reduction in progress.

The Unconnectability Objection

The unconnectability objection can be traced to David Hull's seminal works (1972 and 1974) where he proposed and defended the then-heretical notion that Mendelian genetics is not being reduced by molecular biology, at least not according to Nagel's conception of theoretical reduction. The most rigorous formulation of this objection can be found in Alexander Rosenberg's provocative text (1985).

Rosenberg's opposition to reductionism in genetics rests on an alleged conceptual gap between the classical and molecular theories of genetics. He argues that relations between the gene concepts of the two theories are hopelessly complicated "many-many relations" that will forever frustrate any attempt to systematically connect the two theories. Rosenberg begins his analysis by pointing out that in CMG, genes are always identified by way of their phenotypic effects. Classical geneticists identified the gene for red eye color in *Drosophila*, for example, by following the distribution of red and white phenotypes in successive generations of a laboratory population. The reason CMG will never be reduced to molecular biology, Rosenberg argues, is that there is no manageable connection between the concept of a Mendelian phenotype and that of a molecular gene. The relation between them is complicated by the fact that scores of Mendelian phenotypes are potentially affected by an individual molecular gene and that a vast array of molecular genes are responsible for the production of any given Mendelian phenotype. Rosenberg explains the problem as follows:

> Suppose we have set out to explain the inheritance of normal red eye color in *Drosophila* over several generations. The pathway to red eye pigment production begins at many distinct molecular genes and proceeds through several alternative branched pathways. Some of the genes from which it begins are redundant, in that even if they are prevented from functioning the pigment will be produced. Others are interdependent, so that if one is blocked the other will not produce any product. Still others are "ambiguous"—belonging to several distinct pathways to different phenotypes. The pathway from the genes also contains redundant, ambiguous, and interdependent paths. If we give a biochemical characterization of the gene for red eye color either by appeal to the parts of its pathway of synthesis, or by appeal to the segments of DNA that it begins with, our molecular description of this gene will be too intricate to be of any practical explanatory upshot. (Rosenberg 1985, p. 101)

Rosenberg reasons that since Mendelian genes are identified through their phenotypes, and since the relation between molecular genes and Mendelian phenotypes is exceedingly complex, the connection between the molecular and Mendelian gene concepts must also be exceedingly complex. Hence, he concludes, CMG will forever remain beyond the reductive grasp of molecular biology. Rosenberg does not deny that molecular biologists will occasionally furnish individual accounts of various Mendelian phenomena on a piecemeal basis (as they have done with the genetics of sickle-cell anemia). He insists, however, that the unmanageable complex relations between the gene concepts of the two theories will prevent any systematic, reductive explanation of CMG in terms of molecular theory.

What Rosenberg's persuasive argument does not take into consideration is the relationship between the Mendelian gene and the Mendelian phenotype. According to the classical theory, one gene can affect different phenotypic traits and each phenotypic trait can be affected by different (nonallelic) genes. I will argue that the relationship between the Mendelian gene and the Mendelian phenotype exhibits the same complexity that Rosenberg discusses from the molecular perspective. My argument will not depend upon historical hindsight. Alfred H. Sturtevant, one of the architects of CMG, discussed the complex relation between the Mendelian gene and phenotype in his Ph.D. thesis (1916), which he wrote under T. H. Morgan. Ironically, he illustrated the point with the very same example that Rosenberg considers:

> The difference between normal red eyes and colorless (white) ones in *Drosophila* is due to a difference in a single gene. Yet red is a very complex color, requiring the interaction of at least five (and probably of very many more) different genes for its production. And these genes are quite independent, each chromosome bearing some of them. Moreover, eye-color is indirectly dependent upon a large number of other genes, such as those on which the life of the fly depends. We can then, in no sense identify a given gene with the red color of the eye, even though there is a single gene differentiating it from the colorless eye. So it is for all characters—as Wilson (1912) has put it "... the entire germinal complex is directly or indirectly involved in the production of every character.[5]"

The parallel between Sturtevant's and Rosenberg's accounts of the complex relationship between Mendelian phenotypes and Mendelian genes (Sturtevant's) and between Mendelian phenotypes and molecular genes (Rosenberg's) is striking. Both identify a web of relations too complex for the kind of explanation that Rosenberg seeks. My claim is that the molecular perspective offers a reductive interpretation of the complex picture offered by the classical theory. Our understanding of the biosynthetic pathways explains why there should be many-many relations between classical genes and Mendelian phenotypes.

The problem with Rosenberg's antireductionist line of reasoning is that it assumes that the existence of a particular gene can explain the presence of particular traits in an individual when in fact *genes can only explain* phenotypic *differences* and only *in given populations*. The presence of a gene for red eyecolor on the X chromosome explains why the red-eyed *Drosophila* in a certain population have red eyes instead of white ones. The reason why classical geneticists found manageably simple relations between genes and phenotypic differences is because the genetic backgrounds against which particular genes produced differences were sufficiently uniform from one organism to another in the laboratory populations (of highly related individuals) under study. This can be explained from the molecular perspective in terms of a uniformity in relevant portions of the DNA, which in turn provided a uniform potential for bringing about certain results within the complex web of biosynthetic reactions.

Rosenberg's is but one of several lines of reasoning against the idea that the concepts of CMG and molecular biology can be systematically connected. Others focus on the problem of specifying a precise biochemical definition of a Mendelian gene. If the behavior of Mendelian genes can be explained in terms of molecular biology, some critics reason, then the central concepts of Mendelian theory must be defined in purely biochemical terms. The attempt to define a gene as a relatively short stretch of DNA won't do, the antireductionists point out, because not all relatively short stretches are genes. Furthermore, the attempt to define the gene in terms of a finer structure associated with a specific molecular mechanism will not work because of the diversity of molecular ways in which genes produce their effects. For example, Mendelian genes cannot be identified with reading frames (sections of DNA that are transcribed into RNA) because regulatory genes function without being transcribed. Such considerations reveal that a simple molecular definition of a Mendelian gene is not forthcoming.

The obvious response for the reductionist is simply to hold out for a disjunctive connection.[6] As we learn more about the molecular nature of Mendelian genes, we have discovered that they do not all function by way of the same mechanism. Some genes function by being transcribed into segments of RNA which code for polypeptides. Others function by regulating the transcription of neighboring genes. Furthermore, although all Mendelian genes are relatively short segments of DNA (or perhaps RNA), their finer structure varies with their role. Hence, any definition of Mendelian gene in terms of fine molecular structure will be disjunctive.

While I am not prepared to insist that molecular biology already provides the means for completing a disjunctive definition *in terms of molecular structure*, I do think the elements for such a definition are falling in place. For the time being, I believe it suffices to point out that the behavior of specific Mendelian genes has been explained by identifying them with relatively short segments of DNA which function as units to influence the course of chemical reactions within a biochemical system. The fact that such a characterization has been sufficient for the development of molecular models of a variety of Mendelian phenomena leads me to think that the philosophers' attempt to formulate precise syntactical connections (in the form of explicit and detailed definitions) has been counterproductive. The focus on formal aspects of the postpositivist conception of reduction has led to too much haggling over syntax and not enough analysis of whether genetics exhibits the sort of semantic and pragmatic features that motivated the formal account in the first place.

The Mendelian gene can be specified in molecular biology as a relatively short segment of DNA that functions as a biochemical unit. This specification provides an appropriate interpretation of the many-many relation between a Mendelian gene and phenotype. In addition, it provides a general statement of the precise connections that practicing molecular biologists have drawn between genes and phenotypes in individual cases. Most important, however, it has proved to be tremendously fruitful in research. For it has enabled molecular biologists to apply traditional strategies from classical genetics to uncover the biochemistry underlying many life processes. I conclude that the antireductionist thesis that there is some unbridgeable conceptual gap lurking between CMG and its molecular interpretation is wrong.

The Explanatory Incompleteness Objection

The idea that CMG is being reduced to molecular biology has also been opposed on the grounds that molecular biology will never explain, and hence will never reduce, the classical theory of genetics. Since the postpositivist account of theoretical reduction is centered on the idea that the reducing theory explains the reduced one, this complaint strikes at the very heart of the claim that CMG is being reduced in the spirit of postpositivism. Although this kind of objection can be found interspersed throughout the antireductionistic literature and seems to be an important element motivating the consensus against reductionism in genetics, it is seldom put forth as rigorously as the unconnectability objection. Nevertheless, I will reconstruct and defuse two separate arguments falling under this category.[7]

The Gory Details Argument Antireductionists have argued that knowledge of the molecular makeup of genes does not enhance our understanding of their classical Mendelian behavior. For example, Philip Kitcher (1984), in his brilliant essay which marks the culmination of the antireductionist literature, argues that the assortment of genes is best understood at the cytological level: "The distribution of genes to gametes is to be explained, not by rehearsing the gory details of the reshuffling of the molecules, but through the observation that chromosomes are aligned in pairs just prior to the meiotic division, and that one chromosome from each matched pair is transmitted to each gamete" (Kitcher 1984, p. 370). He goes on to argue that the cytological pattern of explanation is objectively preferable because it can uniformly account for a wide range of cases that would look heterogeneous from a molecular perspective.

Kitcher does not describe a diversity of molecular processes responsible for the segregation of genes during meiosis. Instead, he offers an abstract account of the cytological explanation of gene distribution. According to his account, the distribution of genes is explained by identifying meiosis a belonging to the natural kind of "pair-separation processes." This natural kind of process, he says, is heterogeneous from the molecular perspective because different kinds of forces are responsible for bringing together and pulling apart different paired "entities." The separation of paired entities, he claims, "may occur because of the action of electromagnetic forces or even of nuclear forces;

but it is easy to think of examples in which the separation is effected by the action of gravity'' (Kitcher 1984, p. 350). Kitcher, I think, is *not* making the claim that some paired chromosomes are pulled apart by nuclear forces and others by the force of gravity (such a claim would be completely at odds with today's evidence). Rather, when he is discussing the multiple realizations of pairseparation processes he seems to be conceiving of a natural kind that includes processes quite unlike anything that occurs during meiosis. Hence, his reasoning only suggests that at some high level of abstraction, it is possible to draw an analogy between the process of meiosis and (yet to be specified) processes that have quite different molecular mechanisms. This is a far cry from showing that cytological theory offers a uniform explanation of a range of cases that would appear heterogeneous at the molecular level.

Although meiosis appears to be an unpromising candidate, there are other phenomena that are explained uniformly by CMG, but which are caused by a variety of molecular mechanisms. Phenomena of gene expression provide obvious examples. CMG, for instance, lumps together different kinds of gene expression under the category of dominance. This Mendelian category includes genes that code for structural proteins, genes which code for enzymes, and even regulatory genes. The molecular mechanisms by which these different kinds of genes are eventually expressed are quite different. Yet, when examining concrete cases where CMG offers a more uniform perspective, it is difficult to accept the antireductionist judgment that the shallow explanations of CMG are objectively preferable to the deeper accounts provided by molecular theory.

The idea that the uniformity provided by CMG gives it some sort of explanatory edge over the less uniform molecular account seems plausible only when our attention is called away from the actual biology. But even if uniformity of explanation did provide a potentially decisive advantage, there would be no reason to suppose that the uniformity represented by CMG could not also be captured within the molecular perspective through the familiar scientific practices of abstraction and idealization. The reductionists' view is not that the pictures offered by the reduced and reducing theory are the same, but that they can be connected by auxiliary assumptions such that the reducing theory stands in an explanatory relation to the reduced one. The fact that the reducing theory, when not accompanied by such auxiliary assumptions, more accurately represents the true diversity of mechanisms responsible for various processes should not be held against it.

Antireductionists, of course, do not deny the fact that molecular biology has greatly improved our understanding of genetics. Kitcher (1984), for example, provides an interesting discussion of various ways that molecular genetics has advanced our understanding. But they seem pessimistic when it comes to the issue of whether molecular theory will help us understand what (they think) are the essentials of CMG: the processes by which genes are distributed to gametes. The phenomenon of independent assortment of nonlinked genes, it is claimed, depends only on the pairwise separation of chromosomes. The classical theory apparently tells us all we need to know: nonlinked genes are located on separate nonhomologous chromosomes and nonhomologous chromosomes segregate independently. The identification of genes as segments of a molecular double helix allegedly adds nothing to this account.

This antireductionist argument is problematic for two reasons: first, it becomes less plausible when we flesh it out within CMG (as opposed to Mendel's genetics) and second, it seems unduly pessimistic. To flesh the argument out within CMG, we need to consider not just the independent assortment of nonlinked genes, but also the distribution of linked ones. Recall that of central importance to the classical theory was the fact that linkage is incomplete because of the process of crossing over. At the cytological level, not much can be said about this process except that homologous chromosomes sometimes wrap around each other and swap segments during cellular division. Shortly after the double helical structure of the genetic chains was understood, however, molecular models of crossing over were proposed. The basic Holliday Model (Holliday 1964), illustrated in figure 14.1, has been especially fruitful. Since then, laboratory studies have led to a more detailed, though admittedly tentative, biochemical understanding of the individual steps outlined in this model (see Potter and Dressler 1988). Our understanding of the exchange of segments between paired chains of genes is being greatly enhanced by our knowledge of the molecular structure of those chains. The biochemistry of genetic recombination is a tremendously active area of research and will bring our understanding of the classical Mendelian process of crossing over to the molecular level.

Antireductionists might respond by insisting that although the molecular perspective will contribute to our understanding of this bit of CMG, reductionism is a global thesis and requires that it contribute to all bits of the theory. "What about the independent assortment of nonlinked genes?" they might ask. "How do the molecular details improve the cytological explanation according to which nonlinked genes segregate independently because they are located on different chromosomes, which have been observed (via the microscope) to segregate independently?" This is the point at which I think the gory details objection becomes unduly pessimistic. Surely, the conjugation and separation of homologous chromosomes depends upon molecular mechanisms. While our understanding of why homologous chromosomes pair, why nonhomologous chromosomes don't pair, why separately paired chromosomes segregate independently, and so forth is not well developed, antireductionists haven't offered sufficient reason for thinking these questions won't eventually be answered. The answers to these questions will be given from the molecular perspective and will enhance our understanding of why nonlinked genes assort independently.

Research in the general area of genetic recombination has already displayed signs identified by Nagel as the distinguishing features of an important theoretical reduc-

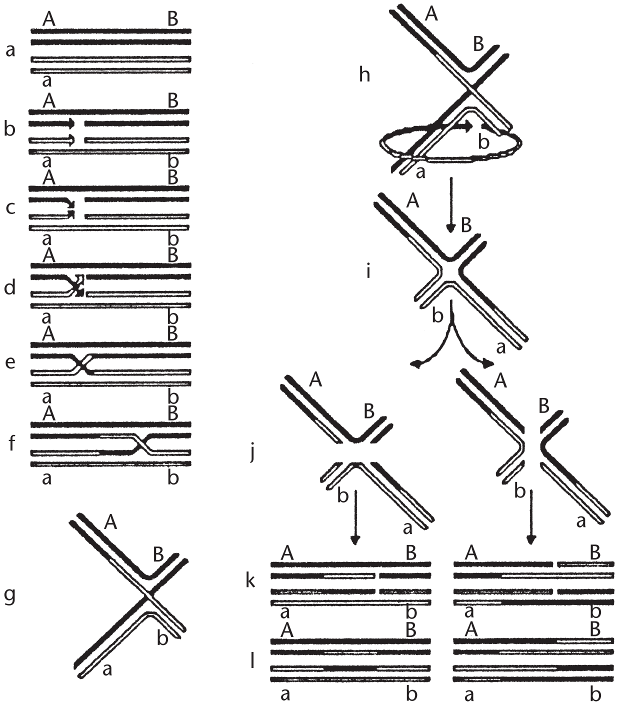


Figure 14.1

The Holliday Model for genetic recombination. (*a*) Two homologous double helices are aligned. (*b*) The two + or − strands are cut. (*c*) The free ends leave the complementary strands to which they have been hydrogen bonded. (*d*) The free ends become associated with the complementary strands in the homologous double helix. (*e*) Ligation creates partially heteroduplex double helices. (*f*) Migration of the branch point occurs by continuing strand transfer by two polynuclotide chains involved in a crossover. (*g*) The Holliday structure shown in extended form. (*h*) The rotation of the structure shown in (*g*) can yield the form depicted in (*i*). Resolution of the structure shown in (*i*) can proceed in two ways, depending on the points of enzymatic cleavage, yielding the structures shown in (*j*). Figure from Potter and Dresler (1979, p. 970). Explanation quoted from Suzuki et al. (1986, p. 360).

tion. One sign is the discovery of surprising connections between seemingly unrelated processes. Recent biochemical research has revealed unexpected connections among the processes of recombination, replication, and repair (see Low 1988). Another sign of a significant reduction is the establishment of new experimental laws that are in better agreement with the facts. Recent lines of biochemical research hold promise for explaining why recombination is not entirely random and for helping us discover the finer patterns of genetic recombination (e.g., patterns of interference in closely spaced exchanges). Hence, even with respect to the Mendelian phenomena for which molecular explanations have tended to lag (i.e., transmission phenomena), the relation between CMG and molecular theory is beginning to exhibit characteristics corresponding to the two nonformal conditions set out by Nagel (1961) in his classic account of theoretical reduction.

The claim that the gory details of molecular biology do not enhance our understanding of key processes underlying CMG is quickly becoming outdated. There is no question that molecular theory has greatly improved our understanding of gene replication, expression, mutation, and recombination. Furthermore, it is just a matter of time before it accounts for the pair-wise coupling and separation of chromosomes during meiosis. Antireductionists need to justify their pessimism and explain why we should not expect molecular biology to continue on its path toward explaining CMG in accordance with the spirit of postpositivist reduction.

The Splintering Argument The antireductionist literature contains hints of a way to dodge reductionism without denying the impending molecular explanation of CMG. Antireductionists might argue that even if a molecular explanation is imminent, the explanation will not come from molecular genetics; instead, it will come from a multitude of theories or fields of molecular biology.[8] Following Hull (1974), antireductionists have typically classified CMG as a theory of transmission genetics and molecular genetics as a theory of development. Presupposing this taxonomy, it might be argued that the classical and molecular theories of genetics explain different aspects of heredity. Hence, antireductionists might argue that even if transmission is explained at the molecular level, it will not be explained by molecular genetics.

It is tempting to dismiss such an antireductionist response as a case of sour grapes. "After volumes of denial," the reductionist might complain, "when the antireductionists are finally forced to admit that molecular biology systematically improves our understanding of classical genetics, they turn around and say that the explanation does not count because it comes from the wrong parts of molecular biology." While tempting, such a reply might miss the crux of the antireductionist complaint.

The issue at stake is whether molecular theory will offer a reasonably coherent explanation of CMG. The possible complaint is that molecular explanations will splinter into numerous fields. Instead of a case of one theory reducing another, one might envision a number of distinct theories explaining bits or pieces of the higher-level theory. If unification is taken to be the hallmark of scientific explanation, the splintering of explanatory paths might appear to clinch the case against reductionism.

While such reasoning sounds plausible in the abstract, it depends on a number of slippery points in need of careful examination. The conceptual division between transmission and developmental genetics, for instance, though widely adopted in the philosophical literature and introductory chapters of genetic texts, has never been carefully analyzed and provides a weak footing for antireductionism. The chief reason offered in favor of this division, that CMG was developed on the basis of transmission studies, applies to molecular genetics as well (transmission studies have played and will continue to play an important role in the development of molecular genetics). Furthermore, the history of classical genetics supports the idea that the scope of CMG encompasses more than transmission. Debates about the presence and absence hypothesis and the position effect, to take just two examples, clearly went beyond issues of transmission.

Perhaps the most serious obstacle to developing the splintering argument is it rests on the idea that there are significant divisions between theories of molecular biology when in fact molecular theory seems to have a diffuse structure. It is far from clear that molecular biology contains a separate theory of molecular genetics. Perhaps molecular biology consists of numerous molecular models of various phenomena, which are not organized into more discrete theories, but are loosely unified by their grounding in a set of common biochemical and biophysical principles. If this is indeed the case, the molecular explanation of CMG will not splinter into a number of different theories at the molecular level.

Developing the splintering objection would also entail substantiating premises about the structure of CMG. Antireductionists minimize the explanatory fit between CMG and molecular theory by deemphasizing the parts of CMG that can be elegantly explained at the molecular level. The explanatory relations between CMG and molecular theory appear fractured, for example, when Kitcher characterizes the principle of gene replication as a "presupposition," as opposed to a "central law" of CMG (1984, p. 361). Such structural accounts of CMG depend on controversial philosophical views about the structure of scientific theories, which I believe are poorly motivated.[9] In any case, they should not be taken for granted.

The prospects for developing the splintering objection appear dim. The objection entails controversial philosophical views about the structure of theories and the nature of explanation as well as highly questionable assumptions about the taxonomy of genetics and the makeup of CMG and molecular biology.

Conclusion

The major objections to the view that CMG is being reduced by molecular biology have not withstood rigorous scrutiny. Perhaps the most surprising result is that the unconnectability objection was found to be so seriously flawed. In retrospect, however, the claimed unconnectability seems unlikely. After all, researchers are successfully identifying the molecular constitutents and pinpointing the exact locations of genes contributing to many classically characterized traits (e.g., Duchenne muscular dystrophy). With sufficient experimental ingenuity, the molecular constituents and locations of the *Drosophila* genes mapped by Morgan et al. could also be identified and pinpointed.[10] As a matter of fact, researchers have just determined the molecular identity of the first Mendelian gene ever discovered, the gene for wrinkled-seed character in pea plants (Bhattacharyya et al. 1990). While molecular biologists have had to conquer many obstacles in their search for the molecular identity of Mendelian genes, the alleged conceptual gap between gene concepts was not one of them.

My examination of the arguments aimed at showing that molecular theory will never explain (and hence never reduce) classical genetics provides a partial explanation of why philosophers and molecular biologists disagree about the reduction of Mendelian genetics. In each case, the antireductionist arguments were based on admittedly brilliant philosophical analyses that appeared plausible in the abstract. But when scrutinized with respect to the details of the actual science, the arguments were found to rest on undue pessimism, on implausible judgments of comparative explanatory value, and on highly questionable assumptions about the structure of CMG and molecular biology. Practicing geneticists believe that the classical theory can be systematically explained at the molecular level, I suggest, because they have a firm grasp of the explanatory power and structure of molecular biology.

Practicing geneticists are also well aware of the achievements that I have identified as signs of a significant theoretical reduction. These achievements include the discovery of unexpected molecular connections among several different genetic processes and promises to improve the precision of our generalizations in genetics. In addition, genetics has provided tremendously fruitful strategies for biochemical research. While such signs indicate that a theoretical reduction is in the making, I have not offered an account of that reduction. I will conclude by briefly anticipating some of the philosophical work that lies ahead.

The main philosophical task will involve reformulating the postpositivist conception of theoretical reduction. Reformulating the postpositivist conception will require and explicit account of explanation as well as an updated account of theories.[11] The antireductionist arguments are tacitly or explicitly linked to accounts of explanation that place a very high premium on unification. This premium is associated with the idea that theoretical reduction requires unification. While the postpositivist view assumes that reduction is accompanied by unification, it is not clear whether the view takes unification to be an essential ingredient or just an expected dividend. Unification is essential for reduction just in case it is essential for explanation. If, as I have hinted, the unificationist criterion for explanation is implausible when invoked within the nittygritty details of genetics, there will be strong incentive to treat unification as a valued bonus, rather than a necessary requirement in the reformulated account of theoretical reduction. The unificationist accounts of explanation and reduction, I suggest, should be assessed from the perspective of molecular biology rather than the other way around.

The reformulation of theoretical reduction will have to be carried out in terms of an explicit account of theories. Most philosophers of biology accept something akin to the semantic view, a view which holds some promise for helping us capture the spirit of postpositivist reduction. One advantage of the semantic view is that it can reportedly help us avoid the logical empiricists' preoccupation with syntactical matters, a preoccupation which plays a role in some antireductionist analyses. Another advantage is that its picture of piecemeal theorizing should enable us to formulate not just a conception of completed reduction, but also the conception of reduction in progress. A shortcoming of the original formulation is that it does not offer a dynamic picture of theoretical reduction. This is especially problematic with respect to genetics where the reduction is still being worked out.

Different philosophical views on the structure of theories and the nature of explanation will undoubtedly lead to different conceptions of theoretical reduction and different pictures of the theoretical relations between classical genetics and molecular biology. These, in turn, can be assessed on the basis of how well they illuminate the actual science. The question of whether CMG is being reduced deserves to be reconsidered, not just because we have good reason to suspect that the antireductionistic consensus is wrong, but also because it provides the opportunity to advance philosophical debates about the structure of theories and the nature of scientific explanation and theoretical reduction.[12]

Notes

I thank Bob Knox for stimulating discussions which influenced my thinking on this subject. Earlier versions of this chapter were presented at the University of Pittsburgh and Indiana University where audiences provided helpful feedback. The National Science Foundation funded this research (Grant No. DIR 89-12221) and the Center for Philosophy of Science at the University of Pittsburgh provided additional support and hospitality while I worked on this chapter.

1. ''Not in the spirit of the postpositivist conception of theoretical reduction'' is emphasized because some critics acknowledge that there are important theoretical relations between CMG and molecular biology, but insist that these relations cannot be understood in terms of the postpositivist conception of reduction. Wimsatt (1976), for example, attacks Nagel's conception and offers his own functional account of the activities related to ''explanatory reduction.'' The more recent literature (e.g., Kitcher 1984 and Rosenberg 1985), which heavily borrows from the earlier works, is less ambiguous and clearly denies that molecular biology will ever reduce CMG in any significant sense of ''reduction.'' I suspect that some of the earlier papers will appear less antireductionist when antireductionism is no longer taken for granted.

2. Rosenberg, however, clings to the old account of theories. See Waters (1990).

3. A good primary source of CMG is Morgan (1928). Carlson (1966) offers a provocative historical account and Hull (1974) gives a succinct and clear presentation of the theory.

4. More detailed accounts of molecular theory can be found in practically any contemporary genetics text.

5. Quoted from Carlson (1988, p. 69).

6. Although this is the obvious response, another is available. For, as some antireductionists have admitted (Hull 1974, Kitcher 1984), the derivation (or explanation) of the principles of CMG does not require the formulation of a set of necessary and sufficient molecular conditions for the terms of CMG. Necessary conditions would suffice.

7. The basic reasoning behind the first argument and hints of the second can be found in Kitcher (1984). Elements of them can also be found in Hull (1974), Wimsatt (1976), and perhaps Maull (1977), and Darden and Maull (1977). A third argument can be constructed on the basis of Beatty's point that molecular biology will never completely explain CMG because it will never be able to reduce the evolutionary explanation of Mendelian principles (see 1983). Beatty has developed an important point about the limits of molecular reductionism in biology and it would be decisive if I was arguing that all of biology can be reduced to a science of proximate causes. But my interest concerns the question of whether the proximate theory of Mendelian genetics will be reduced by the proximate theory of molecular biology. Evolutionary questions about Mendelian phenomena will not go away upon achievement of this reduction; they will simply be reduced to evolutionary questions about molecular phenomena.

8. Kitcher (1984), for example, suggests something along this line when he writes that ''molecular genetics on its own, cannot deliver the goods'' (p. 366) and that ''it would be folly to suggest that the [explanatory] extension is provided by molecular genetics alone'' (p. 368).

9. Kitcher's (1984, p. 361) defense of this characterization is enmeshed within his distinctive account of the structure of scientific theories. I have challenged the central motivation for his radical departure from the traditional view that theories contain law-like claims (Waters 1989 and forthcoming). If I'm correct, the principle of gene replication should be viewed as a law of CMG.

10. *Drosophila* researchers have shifted their attention to genes that play significant roles in developmental processes. So, the search is mainly for genes with developmental significance.

11. The depth of Kitcher's (1984) account of this case stems from the fact that he has taken into account these underlying philosophical issues. But I believe the denial of the unconnectability objection, a more explicit account of molecular biology, and different philosophical views on structure of theories and the nature of explanation will lead to a different and more illuminating picture of the situation.

12. The theses defended in this article are further developed in Waters (forthcoming).

References

Beatty, J. 1983. "The Insights and Oversights of Molecular Genetics: The Place of the Evolutionary Perspective." In P. Asquith and T. Nickles (eds.), *PSA 1982*, Volume 2, East Lansing, MI: Philosophy of Science Association.

Bhattacharyya, M. K., Smith, A. M., Ellis, T. H. N., Hedley, C., and Martin, C. (1990). "The Wrinkled-Seed Character of Pea Described by Mendel Is Caused by a Transposon-like Insertion in a Gene Encoding Starch-Branching Enzyme." *Cell* 60: 115–22.

Carlson, E. A. 1989. *The Gene, A Critical History*. Ames: Iowa State University Press.

Churchland, P. S. 1986. *Neurophilosophy*. Cambridge, Mass.: Bradford/MIT Press.

Darden, L., and N. Maull. 1977. "Interfield Theories." *Philosophy of Science* 44: 43–64.

Holliday, R. 1964. "A Mechanism for Gene Conversion in Fungi." *Genetics Research* 5: 282–303.

Hooker, C. A. 1981. "Towards a General Theory of Reduction, Part I: Historical and Scientific Setting, Part II: Identity in Reduction, Part III: Cross-categorical Reduction." *Dialogue* 20: 38–59, 201–36, 496–529.

Hull, D. 1972. "Reduction in Genetics—Biology or Philosophy?" *Philosophy of Science* 39: 491–99.

———. 1974. *Philosophy of Biological Science*. Englewood Cliffs, N.J.: Prentice-Hall.

Kitcher, P. 1984. "1953 and All That: A Tale of Two Sciences." *Philosophical Review* 93: 335–73.

Low, K. B. 1988. "Genetic Recombination: A Brief Overview." In K. B. Low (ed.), *The Recombination of the Genetic Material*. San Diego: Academic Press, pp. 1–23.

Maull, N. 1977. "Unifying Science without Reduction." *Studies in the History and Philosophy of Science* 8: 143–71.

Morgan, T. H. 1926. *The Theory of the Gene*. New Haven: Yale University Press.

Nagel, E. 1961. *The Structure of Science*. London: Routledge and Kegan Paul.

Potter, H., and Dressler, D. 1979. "Branch Migration in Recombination." *Cold Spring Harbor Symposium of Quantitative Biology* 43: 957–90.

———. 1988. "Genetic Recombination: Molecular Biology, Biochemistry, and Evolution." In K. Brooks Low (ed.), *The Recombination of the Genetic Material*. San Diego: Academic Press, pp. 218–83.

Richardson, R. C. 1979. "Functionalism and Reductionism." *Philosophy of Science* 48: 533–558.

———. 1982. "Discussion: How Not to Reduce a Functional Psychology." *Philosophy of Science* 49: 125–37.

Rosenberg, A. 1985. *The Structure of Biological Science*. Cambridge: Cambridge University Press.

Ruse, M. 1969. ‘‘Reduction in Genetics.’’ In R. S. Cohen et al. (eds.), *PSA 1974*. Boston: D. Reidel, pp. 653–670.

Schaffner, K. 1969. ‘‘The Watson-Crick Model and Reductionism.’’ *British Journal for the Philosophy of Science* 20, 325–348.

———. 1976. ‘‘Reductionism in Biology: Prospects and Problems.’’ In R. S. Cohen et al. (eds.), *PSA 1974*. Boston: D. Reidel, pp. 613–632.

Suzuki, D. T. et al. 1986. *Introduction to Genetic Analysis*. New York: W. H. Freeman and Company.

Waters, C. K. 1989. ‘‘The Universal Laws of Biology.’’ Presented to the 1989 Meeting of the Eastern Division of the American Philosophical Association, abstract in *Proceedings and Addresses of the American Philosophical Association* 62: 74.

———. 1990. ‘‘Rosenberg’s Rebellion.’’ *Biology and Philosophy* 5: 225–239.

———. Forthcoming. ‘‘Laws, Kinds and Generalities in Biology.’’ Unpublished manuscript.

———. Forthcoming. ‘‘Genes Mude Molecular.’’ *Philosophy of Science*.

15 The Multiple Realizability Argument Against Reductionism

Elliott Sober

Reductionism is often understood to include two theses: (1) every singular occurrence that the special sciences can explain also can be explained by physics; (2) every law in a higher-level science can be explained by physics. These claims are widely supposed to have been refuted by the multiple realizability argument, formulated by Putnam (1967, 1975) and Fodor (1968, 1975). The present chapter criticizes the argument and identifies a reductionistic thesis that follows from one of the argument's premises.

1 Introduction

If there is now a received view among philosophers of mind and philosophers of biology about reductionism, it is that reductionism is mistaken. And if there is now a received view as to why reductionism is wrong, it is the multiple realizability argument.[1] This argument takes as its target the following two claims, which form at least part of what reductionism asserts:

(1) Every singular occurrence that a higher-level science can explain also can be explained by a lower-level science.
(2) Every law in a higher-level science can be explained by laws in a lower-level science.

The "can" in these claims is supposed to mean "can in principle," not "can in practice." Science is not now complete; there is a lot that the physics of the present fails to tell us about societies, minds, and living things. However, a completed physics would not thus be limited, or so reductionism asserts (Oppenheim and Putnam 1958).

The distinction between higher and lower of course requires clarification, but it is meant to evoke a familiar hierarchical picture; it runs (top to bottom) as follows—the social sciences, individual psychology, biology, chemistry, and physics. Every society is

From *Philosophy of Science* 66 (December 1999): 542–564.

composed of individuals who have minds; every individual with a mind is alive;[2] every individual who is alive is an individual in which chemical processes occur; and every system in which chemical processes occur is one in which physical processes occur. The domains of higher-level sciences are subsets of the domains of lower-level sciences. Since physics has the most inclusive domain, immaterial souls do not exist and neither do immaterial vital fluids. In addition, since the domains are (properly) nested, there will be phenomena that lower-level sciences can explain, but that higher-level sciences cannot. Propositions (1) and (2), coupled with the claim of nested domains, generate an asymmetry between higher-level and lower-level sciences.

Reductionism goes beyond what these two propositions express. Events have multiple causes. This means that two causal explanations of the same event may cite different causes. A car skids off the highway because it is raining, and also because the tires are bald (Hanson 1958). Proposition (1) says only that if there is a psychological explanation of a given event, then there is also a physical explanation of that event. It does not say how those two explanations are related, but reductionism does. Societies are said to have their social properties *solely in virtue of* the psychological properties possessed by individuals; individuals have psychological properties *solely in virtue of* their having various biological properties; organisms have biological properties *solely in virtue of* the chemical processes that occur within them; and systems undergo chemical processes *solely in virtue of* the physical processes that occur therein. Reductionism is not just a claim about the explanatory capabilities of higher- and lower-level sciences; it is, in addition, a claim to the effect that the higher-level properties of a system are determined by its lower-level properties.[3]

These two parts of reductionism are illustrated in figure 15.1. The circled *e* represents the relation of diachronic explanation; the circled *d* represents the relation of synchronic determination. Reductionism says that if (x) explains (y), then (z) explains (y); it also asserts that (z) determines (x). The multiple realizability argument against reductionism does not deny that higher-level properties are determined by lower-level properties. Rather, it aims to refute propositions (1) and (2)—(z) does not explain (y), or so this argument contends.

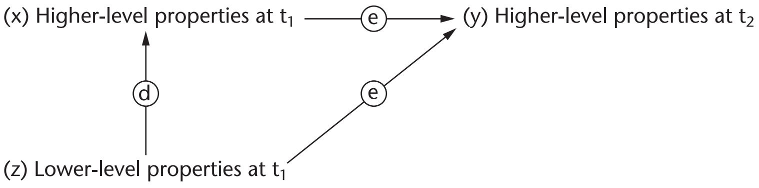


Figure 15.1

Relations of synchronic determination (*d*) and diachronic explanation (*e*) that may connect higher- and lower-level properties.

2 Multiple Realizability

Figure 15.2 is redrawn from the first chapter, entitled "Special Sciences," of Fodor's 1975 book, *The Language of Thought*. It describes a law in a higher-level science and how it might be related to a set of laws in some lower-level science. The higher-level law is couched in its own proprietary vocabulary; P and Q are higher-level properties and the higher-level law says that everything that has P also has Q. The lower-level science provides n laws, each of them connecting an A predicate to a B predicate; the lower-level laws say that everything that has A_i also has B_i (for each $i = 1, 2, \ldots, n$).

The higher-level property P is said to be multiply realizable; $A_1, A_2, \ldots, A_n$ are the different (mutually exclusive and collectively exhaustive) realizations that P might have. Similarly, Q has $B_1, B_2, \ldots, B_n$ as its alternative realizations. What does multiple realizability mean? First, it entails the relation of simultaneous determination; necessarily, if something has A_i at time t, then it has P at t, and if it has B_i at time t, then it has Q at t. But there is something more, and it is this second ingredient that is supposed to ensure that the multiple realizability relation is anti-symmetric. An individual that has P has that property *solely in virtue* of the fact that it has whichever A_i it possesses. Because the higher-level properties are multiply realizable, the mapping from lower to higher is many-to-one. You cannot tell which of the A_i properties is exhibited by a system just from knowing that it has property P, and you cannot tell which of the B_j properties the system has just from knowing that it has Q.[4]

Two examples will make the intended meaning of multiple realizability sufficiently clear. Suppose that different types of physical system can have minds; minds can be built out of neurons, but perhaps they also can be built out of silicon chips. An individual mind—you, for example—will have its psychological properties in virtue of the physical properties that the system possesses. But if you and someone else have some psychological properties in common, there is no guarantee that the two of you also will share physical properties; you and this other person may deploy different physical

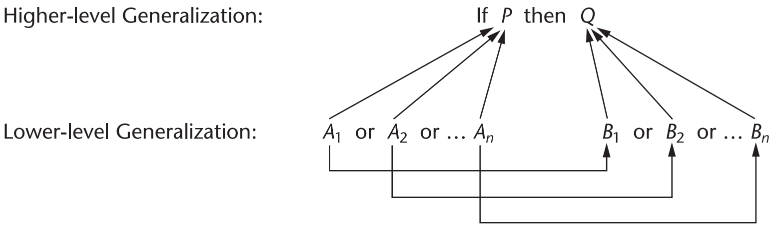


Figure 15.2
The lower-level properties A_i and B_j provide multiple realizations of the higher-level properties P and Q, respectively. One higher-level law and n lower-level laws are depicted, following Fodor 1975.

realizations of the same psychological properties. The same point can be made with respect to biological properties—you have various biological properties, and each of these is present in virtue of your possessing this or that set of physical properties. However, you and some other organism may share a given biological property even though you are physically quite different; this will be true if you and this other organism deploy different physical realizations of the same biological properties.

Since the multiple realizability relation obtains between simultaneously instantiated properties, the relation is not causal (assuming as I will that cause must precede effect). However, the diachronic laws I want to consider *are* causal—they say that a system's having one property at one time causes it to exhibit another property sometime later. The reason I will focus on causal diachronic laws is not that I think that all diachronic laws are causal, but that these provide the clearest cases of scientific explanations.[5] Thus, returning to propositions (1) and (2), we can ask the following two questions about the multiple realizability relations depicted in the second figure:

(1′) If an individual's having property P explains its having property Q, is it also true that its having property A_i explains its having property Q?
(2′) Do lower-level laws of the form "if A_i then B_i" explain the higher-level law "if P then Q"?

Let us assume that the properties described in higher-level sciences are multiply realized by properties discussed in a lower-level science. What consequences follow from this concerning reductionism?

3 The Explanation of Singular Occurrences—Putnam's Peg

Suppose a wooden board has two holes in it. One is circular and has a 1-inch diameter; the other is square and is 1 inch on a side. A cubical peg that is 15/16ths of an inch on each side will fit through the square hole, but not the circular one. What is the explanation? Putnam (1975) says that the explanation is provided by the *macro*-properties just cited of the peg and the holes. He denies that the *micro*-properties of molecules or atoms or particles in the peg and the piece of wood explain this fact. The micro-description is long and complicated and it brings in a welter of irrelevant detail. To explain why the peg goes through one hole but not the other, it does not matter what micro-properties the molecules have, as long as the peg and board have the macro-properties I mentioned. The macro-properties are explanatory; the micro-properties that realize those macro-properties are not. Hence, reductionism is false.

This is a delightfully simple example and argument, but it is possible to have one's intuitions run in the opposite direction. Perhaps the micro-details do not interest *Putnam*, but they may interest *others*, and for perfectly legitimate reasons. Explanations come with different levels of detail. When someone tells you more than you want to hear, this does not mean that what is said fails to be an explanation. There is a difference between explaining too much and not explaining at all.

Compare the micro-story that Putnam derides with a quite different story. Suppose someone suggested that the reason the peg goes through one hole but not the other is that the peg is *green*. Here it is obvious that a mistake has been made. If we demand that explanations be *causal* explanations, it will be quite clear why the color of the peg is not explanatory. It is causally irrelevant. This is an objective feature of the system under consideration and has nothing to do with our desire for brevity or detail.

It is possible to be misled by a superficial similarity that links the micro-story about the particles in the peg and board and the pseudo-explanation that cites the peg's color. Both of the following counterfactuals are true:

If the particles in the peg and board had been different, the peg still would have passed through one hole but not the other, as long as the macro-dimensions were as described.

If the peg had not been green, it still would have passed through one hole but not the other, as long as the macro-dimensions were as described.

If we say that causes are *necessary* for their effects (as does Lewis 1973a), we might be tempted to use these counterfactuals to conclude that the system's micro-features and the peg's color are both causally irrelevant, and hence should not be cited in a causal explanation. This proposal should be understood to mean that the effect would not have happened if the cause had not, *in the specific circumstances that actually obtained*; striking a match is not always necessary to get the match to light, but it may be necessary in various specific circumstances.

There are general questions that may be raised about the adequacy of this account of causation.[6] However, even if we waive these questions, it is important to examine more closely how the counterfactual test connects with Putnam's argument. Let us suppose that the micro-properties of the peg and board's molecules are not necessary for the peg to go through one hole but not the other, if we hold fixed the macro-dimensions. But are the macro-dimensions necessary, if we hold fixed the micro-properties? That is, are we prepared to affirm the following counterfactual?

If the macro-dimensions of the peg and board had been different, while the micro-properties were as described, the peg would not have passed through the one hole but not the other.

This counterfactual has a nomologically impossible antecedent. Many of us simply draw a blank when asked to assign a truth value to such assertions. The semantics of Stalnaker (1968) and Lewis (1973b) does not; it says that the counterfactual is vacuously true. However, before we interpret this as vindicating Putnam's argument, we also should note that the same semantic theory says that the following counterfactual is true as well:

> If the macro-dimensions of the peg and board had been different, while the micro-properties were as described, the peg still would have passed through the one hole but not the other.

It is hard to see how such counterfactuals can vindicate the judgment that the macro-properties are causally efficacious while their micro-realizations are not.[7]

I very much doubt that the concept of explanatory relevance means what Putnam requires it to mean in this argument. When scientists discover why smoking causes cancer, they are finding out which ingredients in cigarette smoke are carcinogenic. If smoking causes cancer, this is presumably because the micro-configuration of cigarette smoke is doing the work. If there turn out to be several carcinogenic ingredients and different cigarettes contain different ones, this does not make the molecular inquiry explanatorily irrelevant to the question of why people get cancer. The fact that *P* is multiply realizable does not mean that *P*'s realizations fail to explain the singular occurrences that *P* explains. A smoker may not want to hear the gory details, but that does not mean that the details are not explanatory.[8]

Putnam says he does not care whether we call the micro-story about the peg and the board a non-explanation, or simply describe it as a "terrible" explanation (Putnam 1975, 296). He thinks that the "goodness" of an explanation "is not a subjective matter." According to the objective concept of good explanation that Putnam has in mind, "an explanation is superior if it is more general" and he quotes with approval a remark by Alan Garfinkel—that "a good explanation is invariant under small perturbations of the assumptions" (301). What makes a more general (more invariant) explanation *objectively* better than one that is less? Putnam's answer is that "one of the things we do in science is to look for laws. Explanation is superior not just subjectively, but *methodologically*, in terms of facilitating the aims of scientific inquiry, if it brings out relevant laws" (301). My reply is that the goal of finding "relevant" laws cuts both ways. Macro-generalizations may be laws, but there also may be laws that relate micro-realizations to each other, and laws that relate micro- to macro- as well. Although "if *P* then *Q*" is more general than "if A_i then B_i,"[9] the virtue of the micro-generalization is that it provides more details. Science aims for depth as well as breadth. Some good explanations are fox-like; others are hedgehogian (Berlin 1953). There is no objective rule concerning which is better.

The claim that the preference for breadth over depth is a matter of taste is consistent with the idea that the difference between a genuine explanation and a nonexplanation is perfectly objective. In fact, it also is consistent with Hempel's (1965) view that the concept of scientific explanation should be explicated in terms of the notion of an ideally complete explanation, and that this is an objective notion. Perhaps an ideally complete scientific explanation of a singular occurrence in which an individual (or set of individuals) exhibits a multiply realizable property (or relation) would include the macro-story, the micro-story, and an account of how these are connected. If this is right, then reductionists and antireductionists alike are mistaken if they think that only part of this multilevel account deserves mention. But whatever the merits are of the idea of an ideally complete scientific explanation, we need to recognize that science in its currently incomplete state still is able to offer up ''explanations.'' Perhaps these should be termed ''explanation sketches,'' since they fall short of the Hempelian ideal. In any case, it remains true that science provides a plurality of such accounts of a given event. They vary in how detailed they are and in the level of organization described.[10]

Returning to Putnam's example, let us imagine that we fact *two* peg-plus-board systems of the type that he describes. If we opt for the macro-explanation of why, in each case, the peg goes through one hole but not the other, we will have provided a *unified explanation*. We will have explained similar effects by describing similar causes. However, if we choose a micro-explanation, it is almost inevitable that we will describe the two systems as being physically different, and thus our explanation will be *dis*unified. We will have explained the similar effects by tracing them back to different types of cause. Putnam uses the terms ''general'' and ''invariant'' to extol the advantages of macro-explanation, but he might just as well have used the term ''unified'' instead. In claiming that it is a matter of taste whether we prefer the macro- or the micro-explanation, I am claiming that there is no objective reason to prefer the unified over the disunified explanation. Science has room for both lumpers and splitters. Some people may not be interested in hearing that the two systems are in fact different; the fact that they have the same macro-properties may be all they wish to learn. But this does not show that discerning differences is less explanatory. Indeed, many scientists would find it more illuminating to be shown how the same effect is reached by different causal pathways.

In saying that the preference for unified explanation is merely a matter of taste, I seem to be contradicting a fundamental fact about scientific inference—that it counts in favor of the plausibility of a theory that the theory unifies disparate phenomena. Actually, no such consequence follows from what I am saying. Here, it is essential to distinguish the *context of justification* from the *context of explanation*.[11] When two theories are evaluated in the light of the evidence available, the fact that one is unified and the other is disunified is epistemologically relevant. In a wide range of circumstances, the unified theory can be expected to be more predictively accurate than the theory that is disunified, when they fit the data about equally well (Forster and Sober 1994). Whether a theory is unified is relevant to deciding whether we should accept it. However, the problem addressed by the multiple realizability argument is not about acceptance. We are supposed to assume that the macro-story and the micro-story are both *true*. Given this, we now are asked to decide which provides the better explanation of why the systems behave similarly. Unification is relevant to acceptance, but unification is not objectively relevant to deciding which accepted statements we should use in formulating explanations. The latter is simply a matter of taste—do we want more details or fewer? The context of justification and the context of explanation are different.

4 The Explanation of Laws—Fodor's Horror of Disjunctions

Whereas Putnam discusses the explanation of singular occurrences, Fodor uses the idea of multiple realizability to argue that laws in a higher-level science are not explained by laws in a lower-level science. This shift introduces some new considerations. Although many, if not all, explanations of singular occurrences are causal, the most familiar cases of explaining laws do not involve tracing effects back to their causes. Laws are usually explained by deriving them from "deeper" laws and initial condition statements; the explained laws and the explaining laws are true at the same time, so it is hard to think of the one as causing the other.

To understand Fodor's antireductionist position, let us consider the following derivation of a higher-level law:

If A_i then B_i (for each $i = 1, 2, \ldots, n$).
If A_1 or A_2 or … or A_n, then B_1 or B_2 or … or B_n.
P iff A_1 or A_2 or … or A_n.
Q iff B_1 or B_2 or … or B_n.

If P then Q.

The first premise describes a set of lower-level laws; the second premise follows from the first. The third and fourth premises state bridge principles that connect a property discussed in a higher-level science with its multiple, lower-level, realizations. By assumption, the premises are true and the conclusion follows from the premises. Why, then, is this derivation not an explanation of the higher-level law?

Fodor's answer is not that the premises involve concepts that come from the higher-level science. Given that the higher-level science and the lower-level science use different vocabularies, any derivation of the one from the other must include bridge principles that bring those different vocabularies into contact (Nagel 1961). Rather, Fodor's reason is that laws cannot be disjunctive. Although he grants that each statement of the form "if A_i then B_i" is a law, he denies that the second premise expresses a law. For the same reason, the third and fourth premises also fail to express laws. To reduce a law, one must explain why the proposition is not just true, but is a law; this is supposed to mean that one must derive it solely from lawful propositions. This is why Fodor thinks that multiple realizability defeats reductionism.

Even if laws cannot be disjunctive, why does the above derivation fail to explain why "if P then Q" is a law? After all, the conclusion will be nomologically necessary if the premises are, and Fodor does not deny that the premises are necessary. Are we really prepared to say that the truth and lawfulness of the higher-level generalization is *inexplicable*, just because the above derivation is peppered with the word "or"? I confess that I feel my sense of incomprehension and mystery palpably subside when I contemplate this derivation. Where am I going wrong?

It also is not clear that laws must be nondisjunctive, nor is it clear what this requirement really amounts to. Take a law that specifies a quantitative threshold for some effect—for example, the law that water at a certain pressure will boil if the ambient temperature exceeds 100°C. This law seems to be disjunctive—it says that water will boil at 101°C, at 102°C, and so on. Of course, we have a handy shorthand for summarizing these disjuncts; we just say that any temperature "above 100°C" will produce boiling water. But if this strategy suffices to render the law about water nondisjunctive, why can't we introduce the letter α to represent the disjunction "A_1 or A_2 or ... or A_n" and β to represent the disjunction "B_1 or B_2 or ... or B_n"? It may be replied that the different disjuncts in the law about water all bring about boiling by the same type of physical process, whereas the different physical realizations A_i that the higher-level property P might have are heterogeneous in the way they bring about the B_i's that are realizations of Q.[12] The point is correct, but it remains unclear why this shows that laws cannot be disjunctive.

Disjunctiveness makes sense when it is understood as a *syntactic* feature of sentences. However, what does it mean for a proposition to be disjunctive, given that the same proposition can be expressed by different sentences? The problem may be illustrated by way of a familiar example. Suppose that the sentence "every emerald is green" and the sentence "every emerald is grue and the time is before the year 2000, or every emerald is bleen and the time is after the year 2000" are equivalent by virtue of the definitions of the terms "grue" and "bleen" (Goodman 1965). If laws are language-independent propositions of a certain type, and if logically equivalent sentences pick out the same proposition, then both sentences express laws, or neither does. Nothing changes if green is a natural kind whereas grue and bleen are not.

Although Fodor (1975) does not mention grue and bleen, it is fairly clear that his thinking about natural kinds—and his horror of disjunctions—both trace back to that issue.[13] Goodman (1965) held that law-like generalizations are confirmed by their positive instances, whereas accidental generalizations are not. The statement "all emeralds are green" is supposed to be lawlike, and hence instance confirmable, in virtue of the fact that "emerald" and "green" name natural kinds (or are "projectible"); "all emeralds are grue," on the other hand, is supposed to be non-lawlike, and so not confirmable by its instances, because it uses the weird predicate "grue." However, subsequent work on the confirmation relation has thrown considerable doubt on the idea that all and only the lawlike statements are instance confirmable (see, e.g., Sober 1988).

If P and (A_1 or A_2 or ... or A_n) are known to be nomologically equivalent, then any probabilistic model of confirmation that takes that knowledge into account will treat them as *confirmationally* equivalent. For example, if a body of evidence confirms the hypothesis that a given individual has P, then that evidence also confirms the hypothesis that the individual has (A_1 or A_2 or ... or A_n). This is a feature, for example, of Bayesian theories of confirmation (on which, e.g., see Howson and Urbach 1989 and Earman 1992). Disjunctiveness has no special meaning within that framework.

Fodor (1975, 21) concedes that the claim that laws must be nondisjunctive is "not strictly mandatory," but then points out that "one denies it at a price." The price is that one loses the connection between a sentence's expressing a law and the sentence's containing kind predicates. "One thus inherits the need for an alternative construal of the notion of a kind"; I am with Fodor when he says that he does not "know what that alternative would be like" (22). Fodor is right here, but his argument is prudential, not evidential. Like Pascal, Fodor is pointing out the disutility of denying a certain proposition, but this is not to show that the proposition is true.

The multiple realizability argument against the reducibility of laws is sometimes formulated by saying that the disjunctions that enumerate the possible realizations of P and Q are "open-ended." This would defeat the derivation described above—the third and fourth premises would be false—but it is important to see that the rules of the game now have changed. The mere fact that P and Q are multiply realizable would no longer be doing the work. And if the point about "open-endedness" is merely epistemological (we now do not *know* all of the physical realizations that P and Q have), it is irrelevant to the claim that higher-level sciences are reducible *in principle*.[14]

5 Probabilistic Explanations

The multiple realizability argument is usually developed by considering deterministic laws. However, laws in many sciences are probabilistic. How would the argument be affected by assuming that P and Q are probabilistically related, and that the A_i and the B_i are too?

Suppose that A_1 and A_2 are the only two possible realizations that P can have, and that B_1 and B_2 are the only two realizations that Q can have (the points I'll make also hold for $n > 2$). Suppose further that the probabilistic law connecting P to Q has the form

$$\Pr(Q \mid P) = p.$$

Then it follows that

$$p = \Pr(Q \mid P) = \Pr(Q \mid A_1)\ \Pr(A_1 \mid P) + \Pr(Q \mid A_2)\ \Pr(A_2 \mid P).$$

If we substitute $p_1 = \Pr(Q \mid A_1)$ and $p_2 = \Pr(Q \mid A_2)$ into this expression, we obtain

$$p = (p_1)\ \Pr(A_1 \mid P) + (p_2)\ \Pr(A_2 \mid P).$$

The probability (p) described in the higher-level law is a *weighted average* of the two probabilities p_1 and p_2; the weighting is determined simply by how often systems with P happen to deploy one micro-realization rather than the other.

It is not inevitable that $p = p_1 = p_2$. For example, suppose that smoking (P) makes lung cancer (Q) highly probable and that cigarette smoke always contains one of two carcinogenic ingredients (A_1 or A_2), which are found only in cigarette smoke. It can easily turn out that one of these ingredients is more carcinogenic than the other.[15] This means that there can be an important difference between higher-level and lower-level explanations of the same event—they may differ in terms of the probabilities that *explanans* confers on *explanandum*. To see why, let us add one more detail to the example. Suppose that lung cancer can be realized by one of two types of tumor (B_1 or B_2) growing in the lungs. Given this, consider an individual who has lung cancer. How are we to explain why this person has that disease? One possible reply is to say that the person smoked cigarettes. A second possibility is to say that the cancer occurred because the person inhaled ingredient A_1. Putnam's multiple realizability argument entails that the second suggestion is either no explanation at all, or is a "terrible" explanation. I suggest, however, that it should be clear to the unjaundiced eye that the second explanation may have its virtues. Perhaps A_1 confers on lung cancer a different probability from the one entailed by A_2 $(p_1 \neq p_2)$, and so the first account entails a different probability of cancer than the second $(p \neq p_1)$. Furthermore, perhaps A_1 and A_2 confer different probabilities on the two tumors B_1 and B_2 and these tumors respond differently to different treatments. The additional details provided by the micro-explanation are not stupid and irrelevant. They make a difference—to the probability of the *explanandum*, and to much else.[16] Perhaps it is a good thing for cancer research that the multiple realizability argument has not won the hearts of oncologists.

6 Inference to the Best Explanation

I suspect that the multiple realizability argument has exerted so much influence because of a widespread misunderstanding concerning how *inference to the best explanation* works. The rough idea behind this mode of inference is that one should accept or reject hypotheses by deciding whether they are needed to explain observed phenomena. This inferential procedure seems to bear on the issue of reductionism as follows: We *now* need statements formulated in higher-level sciences because present day physics is not able to tell us how to understand societies, minds, and living things. However, if reductionism is correct, then these higher-level statements will not be needed once we have an ideally complete physics, and so they *then* should be rejected. But surely an ideally complete physics would not make it reasonable to reject all statements in higher-level sciences. This means that those statements must be needed to explain something that statements in an ideal physics could not explain. The multiple realizability argument presents itself as a diagnosis of why this is so.

This line of argument rests on a misunderstanding of inference to the best explanation. If you think that A_1 is one of the micro-realizations that P has, then you should not view "P causes Q" and "A_1 causes Q" as competing hypotheses (Sober 1999). The evidence you have may justify accepting both. Inference to the best explanation is a procedure that belongs to the context of justification. Once you have used that technique to accept a variety of different hypotheses, it is perfectly possible that your set of beliefs will furnish several explanations of a given phenomenon, each perfectly compatible with the others. Some of those explanations will provide more details while others will provide fewer. Some may cite proximal causes while others will cite causes that are more distal. The mistake comes when one applies the principle of inference to the best explanation a *second* time—to the set of hypotheses one *already* believes, and rejects hypotheses that one does not "need" for purposes of explanation. Inference to the best explanation is a rule for deciding what to believe; it is not a principle for retaining or eliminating beliefs that one already has perfectly good evidence for accepting. If hypotheses in higher-level sciences can be accepted on the basis of evidence, they will not be cast into the outer darkness simply because physics expands.

It is worth bearing in mind that the phrase "inference to 'the' best explanation" can be misleading. The hypothesis singled out in such inferences is not the best of all explanations (past, present, and future) that could be proposed; it is merely the best of the competing hypotheses under evaluation. Hypothesis testing is essentially a contrastive activity; a given hypothesis is tested by testing it *against* one or more alternatives (Sober 1994). When psychological hypotheses compete against each other, inference to the best explanation will select the best of the competitors; of necessity, the winner in this competition will be a psychological hypothesis, because all the competitors are. Likewise, when physicalistic explanations of a behavior compete against each other, the resulting selection will, of course, be a physicalistic explanation. It is perfectly consistent with these procedures that a given phenomenon should have a psychological *and* a physicalistic explanation. Both reductionists and antireductionists go wrong if they think that the methods of science force one to choose among hypotheses that, in fact, are not in competition at all.[17]

7 Two Other Criticisms of the Multiple Realizability Argument

The multiple realizability argument, when it focuses on the explanation of singular occurrences, has three premises:

Higher-level sciences describe properties that are multiply realizable and that provide good explanations.

If a property described in a higher-level science is multiply realizable at a lower level, then the lower-level science will not be able to explain, or will explain only feebly, the phenomena that the higher-level science explains well.

If higher-level sciences provide good explanations of phenomena that lower-level sciences cannot explain, or explain only feebly, then reductionism is false.

Reductionism is false.

I have criticized the second premise, but the first and third have not escaped critical scrutiny (see, e.g., Lewis 1969, Churchland 1982, Enç 1983, and Kim 1989; Bickle 1998 provides a useful discussion). I will consider these other objections separately.

Philosophers with eliminitivist leanings have criticized the first premise. They have suggested that if "pain," for example, is multiply realizable, then it probably does not have much explanatory power. Explanations that cite the presence of "pain" will be decidedly inferior to those that cite more narrow-gauged properties, such as "human pain," or "pain with thus-and-such a neural realization." Philosophers who advance this criticism evidently value explanations for being deep, but not for being general. I disagree with this one-dimensional view, just as I disagree with the multiple realizability argument's single-minded valuation of generality at the expense of depth. Higher-level explanations often provide fewer explanatory details, but this does not show that they are inferior *tout court*.

It might interest philosophers of mind who have these worries about multiply realized psychological properties to consider the multiply realized properties discussed in evolutionary biology. In cognitive science, it is difficult to point to many present-day models that are well-confirmed and that are articulated by describing multiply realizable properties; this is mostly a hoped-for result of scientific advance. However, in evolutionary biology, such models are extremely common. Models of the evolution of altruism (Sober and Wilson 1998), for example, use the concept of fitness and it is quite clear that fitness is multiply realizable. These models have a useful generality that descriptions of the different physical bases of altruism and selfishness would not possess.

The third premise in the multiple realizability argument also has come in for criticism. Perhaps *pain* is multiply realizable, but *human pain* may not be. And if *human pain* is multiply realizable, then some even more circumscribed type of pain will not be. What gets reduced is not pain in general, but specific physical types of pain (Nagel 1965). The multiple realizability argument is said to err when it assumes that reductionism requires *global* reduction; *local* reduction is all that reductionism demands. To this objection, a defender of the multiple realizability argument might reply that there are many questions about reduction, not just one. If human pain gets reduced to a neurophysiological state, but pain in general does not, then reductionism is a correct claim about the former, but not about the latter. If psychology provides explanations in which pain—and not just *human* pain—is an *explanans*, then reductionism fails as a claim about *all* of psychology.

Scientists mean a thousand different things by the term "reductionism." Philosophers have usually been unwilling to tolerate this semantic pluralism, and have tried to say what reductionism "really" is. This quest for univocity can be harmless as long as philosophers remember that what they call the "real" problem is to some degree stipulative. However, philosophers go too far when they insist that reductionism requires local reductions but not global reductions. There are many reductionisms—focusing on one should not lead us to deny that others need to be addressed.

8 A Different Argument Against a Different Reductionism

Although the multiple realizability argument against reductionism began with the arguments by Putnam and Fodor that I have reviewed, more recent appeals to multiple realizability sometimes take a rather different form. The claim is advanced that higher-level sciences "capture patterns" that would be invisible from the point of view of lower-level science. Here the virtue attributed to the higher-level predicate "P" is not that it *explains* something that the lower-level predicate "A_i" cannot explain, but that the former *describes* something that the latter does not. The predicate "P" describes what the various realizations of the property P have in common. The disjunctive lower-level predicate "A_1 or A_2 or ... or A_n" does not do this in any meaningful sense. If I ask you what pineapples and prime numbers have in common and you reply that they both fall under the disjunctive predicate "pineapple or prime number," your remark is simply a joke. As a result, "if P then Q" is said to describe a regularity that "if (A_1 or A_2 or ... or A_n) then (B_1 or B_2 or ... or B_n)" fails to capture.

Whether or not this claim about the descriptive powers of higher- and lower-level sciences is right, it involves a drastic change in subject. Putnam and Fodor were discussing what higher- and lower-level sciences are able to *explain*. The present argument concerns whether a lower-level science is able to *describe* what higher-level sciences *describe*. I suspect that this newer formulation of the multiple realizability argument has seemed to be an elaboration, rather than a replacement, of the old arguments in part because "capturing a pattern" (or a generalization) has seemed to be more or less equivalent with "explaining a pattern" (or a generalization). However, there is a world of difference between describing a fact and explaining the fact so described. This new argument does not touch the reductionist claim that physics can explain everything that higher-level sciences can explain.

9 Concluding Comments

Higher-level sciences often provide more general explanations than the ones provided by lower-level sciences of the same phenomena. This is the kernel of truth in the multiple realizability argument—higher-level sciences ''abstract away'' from the physical details that make for differences among the micro-realizations that a given higher-level property possesses. However, this does not make higher-level explanations ''better'' in any absolute sense. Generality is one virtue that an explanation can have, but a distinct—and competing—virtue is depth, and it is on this dimension that lower-level explanations often score better than higher-level explanations. The reductionist claim that lower-level explanations are *always* better and the antireductionist claim that they are *always* worse are both mistaken.

Instead of claiming that lower-level explanations are always better than higher-level explanations of the same phenomenon, reductionists might want to demure on this question of better and worse, and try to build on the bare proposition that physics in principle can explain any singular occurrence that a higher-level science is able to explain. The level of detail in such physical explanations may be more than many would want to hear, but a genuine explanation is provided nonetheless, and it has a property that the multiple realizability argument has over-looked. For reductionists, the interesting feature of physical explanations of social, psychological, and biological phenomena is that they use the same basic theoretical machinery that is used to explain phenomena that are nonsocial, nonpsychological, and nonbiological. This is why reductionism is a thesis about the *unity* of science. The special sciences unify by abstracting away from physical details; reductionism asserts that physics unifies because everything can be explained, and explained *completely*, by adverting to physical details. It is ironic that ''unification'' is now a buzz word for antireductionists, when not so long ago it was the *cri de coeur* of their opponents.

To say that physics is capable in principle of providing a complete explanation does not mean that physical explanations will mention everything that might strike one as illuminating. As noted above, the explanations formulated by higher-level sciences can be illuminating, and physics will not mention *them*. Illumination is to some degree in the eye of the beholder; however, the sense in which physics can provide complete explanations is supposed to be perfectly objective. If we focus on *causal* explanation, then an objective notion of explanatory completeness is provided by the concept of *causal completeness*:

$$\Pr(\text{higher-level properties at } t_2 \mid \text{physical properties at } t_1 \;\&\; \text{higher-level properties at } t_1) = \Pr(\text{higher-level properties at } t_2 \mid \text{physical properties at } t_1).$$

To say that physics is causally complete means that (a complete description of) the physical facts at t_1 *determines* the probabilities that obtain at t_1 of later events; adding information about the higher-level properties instantiated at t_1 makes no difference.[18] In contrast, multiple realizability all but guarantees that higher-level sciences are causally incomplete:

$$\Pr(\text{higher-level properties at } t_2 \mid \text{physical properties at } t_1 \ \&\ \text{higher-level properties at } t_1) \neq \Pr(\text{higher-level properties at } t_2 \mid \text{higher-level properties at } t_1).$$

If A_1 and A_2 are the two possible realizations of P, then one should not expect that $\Pr(Q \mid P \ \&\ A_1) = \Pr(Q \mid P \ \&\ A_2) = \Pr(Q \mid P)$ (Sober 1999).

Is physics causally complete in the sense defined? It happens that causal completeness follows from the thesis of simultaneous determination described earlier (Sober 1999). This fact does not settle whether physics *is* causally complete, but merely pushes the question back one step. Why think that the physical facts that obtain at a given time determine all the nonphysical facts that obtain at that time? This is a question I will not try to answer here. However, it is worth recalling that defenders of the multiple realizability argument usually assume that the lower-level physical properties present at a time determine the higher-level properties that are present at that same time. This commits them to the thesis of the causal completeness of physics. If singular occurrences can be explained by citing their causes, then the causal completeness of physics insures that physics has a variety of explanatory completeness that other sciences do not possess. This is reductionism of a sort, though not the sort that the multiple realizability argument aims to refute.

Acknowledgments

My thanks to Martin Barrett, John Beatty, Tom Bontly, Ellery Eells, Berent Enç, Branden Fitelson, Jerry Fodor, Martha Gibson, Daniel Hausman, Dale Jamieson, Andrew Levine, Brian McLaughlin, Terry Penner, Larry Shapiro, Chris Stephens, Richard Teng, Ken Waters, Ann Wolfe, and an anonymous referee for this journal for comments on earlier drafts.

Notes

1. Putnam (1967, 1975) and Fodor (1968, 1975) formulated this argument with an eye to demonstrating the irreducibility of psychology to physics. It has been criticized by Lewis (1969), Churchland (1982), Enç (1983), and Kim (1989), but on grounds distinct from the ones to be developed here. Their criticisms will be discussed briefly towards the end of this paper.

The multiple realizability argument was first explored in philosophy of biology by Rosenberg (1978, 1985), who gave it an unexpected twist; he argued that multiple realizability entails a kind of reductionism (both about the property of fitness and also about the relation of classical Mendelian genetics to molecular biology). In contrast, Sober (1984) and Kitcher (1984) basically followed the Putnam/Fodor line. The former work argues that the multiple realizability of fitness entails the irreducibility of theoretical generalizations about fitness; the latter argues for the irreducibility of classical Mendelian genetics to molecular biology. Waters (1990) challenges the specifics of Kitcher's argument; much of what he says is consonant with the more general criticisms of the multiple realizability argument to be developed here. Sober (1993) defends reductionism as a claim about singular occurrences, but denies that it is correct as a claim about higher-level laws.

2. If some computers (now or in the future) have minds, then the reducibility of psychology to biology may need to be revised (if the relevant computers are not "alive"); the obvious substitute is to have reductionism assert that psychology reduces to a physical science. Similarly, if some societies are made of mindless individuals (consider, for example, the case of the social insects), then perhaps the reduction will have to "skip a level" in this instance also.

3. Reductionism should not be formulated so that it is committed to individualism of the sort discussed in philosophy of mind. For example, if wide theories of content are correct, then the beliefs that an individual has at a time depend not just on what is going on inside the skin of that individual at that time, but on what is going on in the individual's environment, then and earlier.

4. Although multiple realizability induces an asymmetry between P and each A_i, it does not entail that there is an asymmetry between P and the disjunctive property (A_1 or A_2 or ... or A_n). Fodor would say that this disjunctive predicate fails to pick out a natural kind, a point that will be discussed later.

5. Here I waive the question of whether *all* explanations are causal explanations, on which see Sober 1983 and Lewis 1986.

6. I will mention two. The first concerns how this theory of causation analyzes putative cases of overdetermination by multiple actual causes. Suppose Holmes and Watson each simultaneously shoot Moriarty through the heart. The theory entails that Holmes did not cause Moriarty's death, and Watson did not either. Rather, the cause is said to be disjunctive—Holmes shot him or Watson did. The second question comes from thinking about the possibility of indeterministic causation. Just as the totality of the antecedent causal facts need not suffice for the effect to occur, so the effect could have happened even if the causes had been different.

7. I am grateful to Brian McLaughlin for drawing my attention to this line of argument.

8. It is worth considering a curious remark that Putnam makes in a footnote before he introduces the example of the peg and board. He says:

> Even if it were not physically possible to realize human psychology in a creature made of anything but the usual protoplasm, DNA, etc., it would still not be correct to say that psychological states are identical with their physical realizations. For, as will be argued below, such an identification has no *explanatory* value in *psychology*. (1975, 293)

He then adds the remark: "on this point, compare Fodor, 1968," presumably because Fodor thought that antireductionism depends on higher-level properties being *multiply* realizable.

If we take Putnam's remark seriously, we must conclude that he thinks that the virtue of higher-level explanations does not reside in their greater generality. If a higher-level predicate (P) has just one possible physical realization (A_1), then P and A_1 apply to exactly the same objects. Putnam presumably would say that citing A_1 in an explanation provides extraneous information, whereas citing P does not. It is unclear how this concept of explanatory relevance might be explicated. In any event, I have not taken this footnote into account in describing the "multiple realizability argument," since Putnam's point here seems to be that *multiple* realizability does not bear on the claims he is advancing about explanation. This is not how the Putnam/Fodor argument has been understood by most philosophers.

9. I grant this point for the sake of argument, but it bears looking at more closely. Intuitively, "if P then Q" is more general than "if A_i then B_i" because the extension of P properly contains the extension of A_i. However, each of these conditionals is logically equivalent with its contrapositive, and it is equally true that the extension of not-B_i properly contains the extension of not-Q. This point is not a mere logical trick, to be swept aside by saying that the "right" formulation of a law is one that uses predicates that name natural kinds. After all, some laws (specifically, zero force laws) are typically stated as conditionals but their applications usually involve the predicates that occur in the contrapositive formulation. For example, the Hardy–Weinberg Law in population genetics describes how gamete frequencies will be related to genotype frequencies when no evolutionary forces are at work; its typical applications involve noting a departure from Hardy–Weinberg genotype frequencies, with the conclusion being drawn that some evolutionary forces are at work (Sober 1984). To say that the Hardy–Weinberg law has zero generality because every population is subject to evolutionary forces is to ignore the standard way in which the law is applied, and applied frequently, to nature.

10. Putnam's argument also has implications about the explanatory point of citing distal and proximate causes of a given effect. Imagine a causal chain from C_d to C_p to E. Suppose that C_d suffices for the occurrence of C_p, but is not necessary, and that the only connection of C_d to E is through C_p. Then Putnam's argument apparently entails that C_p explains E, and that C_d is either not an explanation of E, or is a terrible explanation of that event. But surely there can be an explanatory point to tracing an effect more deeply into the past. And surely it does not automatically increase explanatory power to describe more and more proximate causes of an effect.

11. The distinction between justification and explanation was clearly drawn by Hempel (1965), who points out that why-questions can be requests for evidence or requests for explanation. This distinction supplements the familiar logical empiricist distinction between the *context of discovery* and the *context of justification.*

12. Fodor (1998, 16) says that a disjunction may occur in a bridge law if and only if the disjunction is "independently certified," meaning that "it also occurs in laws at its own level." The disjunction in the law about boiling presumably passes this test.

13. See, for example, Davidson's (1966) discussion of "all emeroses are gred" and also Davidson 1970.

14. Moreover, the multiple realizability argument is not needed to show that the thesis of *reducibility in practice* is false; one can simply inspect present-day science to see this.

15. If laws must be time-translationally invariant, then it is doubtful that "$\Pr(Q \mid P) = p$" expresses a law, if P is multiply realizable (Sober 1999).

16. This argument would not be affected by demanding that a probabilistic explanation must cite the positive and negative causal factors that raise and lower the probability of the *explanandum* (see, e.g., Salmon 1984). Cigarette smoke may raise the probability of lung cancer to a different extent than inhaling A_1 does, and so the two explanations will differ in important ways.

17. This point bears on an argument that Fodor (1998) presents to supplement his (1975) argument against reductionism. I am grateful to Fodor for helping me to understand this new argument. Fodor compares two hypotheses (which I state in the notation I have been using): (i) "if (A_1 or A_2 or . . . or A_n), then Q" and (ii) "if (A_1 or A_2 or . . . or A_n) then P (because the A_i's are possible realizations of P), and if P then Q." Fodor points out that the latter generalization is logically stronger (19); he then claims that it is sound inductive practice to "prefer the strongest claim compatible with the evidence, all else being equal" (20). Since we should accept the stronger claim instead of the weaker one, Fodor concludes that reductionism is false.

I have three objections to this argument. First, I do not think that the two generalizations are in competition with each other. If one thinks that the first conditional is true, and wants to know whether, in addition, it is true that the A_i's are realizations of P, then the proper competitor for this conjecture is that at least one of the A_i's is *not* a realization of P. Second, even if the two hypotheses were competitors, Fodor's Popperian maxim is subject to the well-known "tacking problem"—that irrelevant claims can be conjoined to a well-confirmed hypothesis to make it logically stronger. Fodor, of course, recognizes that *H&I* is not always preferable to *H*, ceteris paribus; however, he thinks that a suitably clarified version of the maxim he describes is plausible and that it will have the consequence he says it has for the example at hand. I have my doubts. It is illuminating, I think, to compare this inference problem to a structurally similar problem concerning intervening variables. If the A_i's are known to cause Q, should one postulate a variable (P) that the A_i's cause, and which causes Q? I do not think that valid inductive principles tell one to prefer the intervening variable model over one that is silent on the question of whether the intervening variable exists, when both models fit the data equally well (see Sober 1998 for further discussion). Third, even if the stronger hypothesis should be accepted in preference to the weaker one, I do not see that this refutes reductionism (though it does refute "eliminativist reductionism"). After all, the reductionist can still maintain that "if P then Q" is explained by theories at the lower level.

Notice that Fodor's argument does not depend on whether the A_i's listed are some or all of the possible realizations that P can have; it also does not matter whether the modality involved is metaphysical or nomological. Notice, finally, that this argument concerns inductive inference (the "context of justification," mentioned earlier), not explanation, which is why it differs from the argument of Fodor 1975.

18. Let $M =$ all the higher-level properties a system has at time t_1. Let $P =$ all the physical properties that the system has at t_1. And let $B =$ some property that the system might have at the later time t_2. We want to show that

$$\Pr(M \mid P) = 1.0$$

entails

$$\Pr(B \mid P) = \Pr(B \mid P \;\&\; M).$$

First note that $\Pr(B \mid P)$ can be expanded as follows:

$$\begin{aligned}\Pr(B \mid P) &= \Pr(B \;\&\; P)/\Pr(P)\\ &= [\Pr(B \;\&\; P \;\&\; M) + \Pr(B \;\&\; P \;\&\; \text{not-}M)]/\Pr(P)\\ &= [\Pr(B \mid P \;\&\; M)\,\Pr(P \;\&\; M) + \Pr(B \;\&\; \text{not-}M \mid P)\,\Pr(P)]/\Pr(P)\\ &= \Pr(B \mid P \;\&\; M)\,\Pr(M \mid P) + \Pr(B \;\&\; \text{not-}M \mid P)\end{aligned}$$

From this last equation, it is clear that if $\Pr(M \mid P) = 1.0$, then $\Pr(B \mid P) = \Pr(B \mid P \;\&\; M)$.

References

Berlin, Isaiah (1953), *The Hedgehog and the Fox*. New York: Simon and Shuster.

Bickle, John (1998), *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.

Churchland, Paul (1982), "Is 'Thinker' a Natural Kind?," *Dialogue* 21: 223–238.

Davidson, Donald (1966), "Emeroses by Other Names," *Journal of Philosophy* 63: 778–780. Reprinted in *Essays on Actions and Events*. Oxford: Oxford University Press, 1980, 225–227.

———. (1970), "Mental Events," in L. Foster and J. Swanson (eds.), *Experience and Theory*. London: Duckworth. Reprinted in *Essays on Actions and Events*. Oxford: Oxford University Press, 1980, 207–225.

Earman, John (1992), *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Enç, Berent (1983), "In Defense of the Identity Theory," *Journal of Philosophy* 80: 279–298.

Fodor, Jerry (1968), *Psychological Explanation*. Cambridge, MA: MIT Press.

———. (1975), *The Language of Thought*. New York: Thomas Crowell.

———. (1998), "Special Sciences—Still Autonomous After All These Years," in *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. Cambridge, MA: MIT Press.

Forster, Malcolm and Elliott Sober (1994), "How to Tell When Simpler, More Unified, or Less *Ad Hoc* Theories Will Provide More Accurate Predictions," *British Journal for the Philosophy of Science* 45: 1–35.

Goodman, Nelson (1965), *Fact, Fiction, and Forecast*. Indianapolis: Bobbs-Merrill.

Hanson, N. Russell (1958), *Patterns of Discovery*. Cambridge: Cambridge University Press.

Hempel, Carl (1965), "Aspects of Scientific Explanation," in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Howson, Colin and Peter Urbach (1989), *Scientific Reasoning: The Bayesian Approach.* La Salle: Open Court.

Kim, Jaegwon (1989), "The Myth of Nonreductive Materialism," *Proceedings and Addresses of the American Philosophical Association* 63: 31–47. Reprinted in *Supervenience and Mind.* Cambridge: Cambridge University Press, 1993.

Kim, Sungsu (unpublished), "Physicalism, Supervenience, and Causation—a Probabilistic Approach."

Kitcher, Philip (1984), "1953 and All That: A Tale of Two Sciences," *Philosophical Review* 93: 335–373. Reprinted in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 1994, 379–399.

Lewis, David (1969), "Review of *Art, Mind, and Religion*," *Journal of Philosophy* 66: 22–27. Reprinted in N. Block (ed.), *Readings in Philosophy of Psychology*, vol. 1. Cambridge, MA: Harvard University Press, 1983, 232–233.

———. (1973a), "Causation," *Journal of Philosophy* 70: 556–567. Reprinted with a "Postscript" in D. Lewis, *Philosophical Papers*, vol. 2. Oxford: Oxford University Press, 1986, 159–213.

———. (1973b), *Counterfactuals.* Cambridge, MA: Harvard University Press. Revised edition 1986.

———. (1986), "Causal Explanation," in D. Lewis, *Philosophical Papers*, vol. 2. Oxford: Oxford University Press, 214–240.

Nagel, Ernest (1961), *The Structure of Science*. New York: Harcourt Brace.

Nagel, Thomas (1965), "Physicalism," *Philosophical Review* 74: 339–356.

Oppenheim, Paul, and Hilary Putnam (1958), "Unity of Science as a Working Hypothesis," in H. Feigl, G. Maxwell, and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, Minneapolis: University of Minnesota Press, 3–36.

Putnam, Hilary (1967), "Psychological Predicates," in W. Capitan and D. Merrill (eds.), *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, 37–48. Reprinted as "The Nature of Mental States" in *Mind, Language, and Reality*. Cambridge: Cambridge University Press, 1975, 429–440.

———. (1975), "Philosophy and our Mental Life," in *Mind, Language, and Reality*. Cambridge: Cambridge University Press, 291–303.

Rosenberg, Alexander (1978), "The Supervenience of Biological Concepts," *Philosophy of Science* 45: 368–386.

———. (1985), *The Structure of Biological Science*. Cambridge: Cambridge University Press.

Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Sober, Elliott (1983), "Equilibrium Explanation," *Philosophical Studies* 43: 201–210.

———. (1984), *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, MA: MIT Press. 2nd edition, University of Chicago Press, 1994.

———. (1988), ''Confirmation and Lawlikeness,'' *Philosophical Review* 97: 93–98.

———. (1994), ''Contrastive Empiricism,'' in *From a Biological Point of View*. Cambridge: Cambridge University Press, 114–135.

———. (1998), ''Black Box Inference: When Should an Intervening Variable be Postulated?,'' *British Journal for the Philosophy of Science* 49: 469–498.

———. (1999), ''Physicalism from a Probabilistic Point of View,'' *Philosophical Studies* 95: 135–174.

Sober, Elliott and David S. Wilson (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Stalnaker, Robert (1968), ''A Theory of Conditionals,'' in N. Rescher (ed.), *Studies in Logical Theory*. Oxford: Blackwell, 98–112.

Waters, Kenneth (1990), ''Why the Antireductionist Consensus Won't Survive the Case of Classical Mendelian Genetics,'' *PSA 1990*. E. Lansing, MI: Philosophy of Science Association, 125–139. Reprinted in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 1994, 402–417.

VIII Essentialism and Population Thinking

16 Typological versus Population Thinking

Ernst Mayr

Rather imperceptibly, a new way of thinking began to spread through biology soon after the beginning of the nineteenth century. It is now most often referred to as population thinking. What its roots were is not at all clear, but the emphasis of animal and plant breeders on the distinct properties of individuals was clearly influential. The other major influence seems to have come from systematics. Naturalists and collectors realized increasingly often that there are individual differences in collected series of animals, corresponding to the kind of differences one would find in a group of human beings. Population thinking, despite its immense importance, spread rather slowly, except in those branches of biology that deal with natural populations.

In systematics it became a way of life in the second half of the nineteenth century, particularly in the systematics of the better-known groups of animals, such as birds, mammals, fishes, butterflies, carabid beetles, and land snails. Collectors were urged to gather large samples at many localities, and the variation within populations was studied as assiduously as differences between localities. From systematics, population thinking spread, through the Russian school, to population genetics and to evolutionary biology. By and large it was an empirical approach with little explicit recognition of the rather revolutionary change in conceptualization on which it rested. So far as I know, the following essay, excerpted from a paper originally published in 1959, was the first presentation of the contrast between essentialist and population thinking, the first full articulation of this revolutionary change in the philosophy of biology.

The year of publication of Darwin's *Origin of Species*, 1859, is rightly considered the year in which the modern science of evolution was born. It must not be forgotten, however, that preceding this zero year of history there was a long prehistory. Yet, despite the existence in 1859 of a widespread belief in evolution, much published evidence on its course, and numerous speculations on its causation, the impact of Darwin's publication was so immense that it ushered in a completely new era.

It seems to me that the significance of the scientific contribution made by Darwin is threefold:

From *Evolution and the Diversity of Life*, Harvard University Press, 1975, 26–29.

1. He presented an overwhelming mass of evidence demonstrating the occurence of evolution.
2. He proposed a logical and biologically well-substantiated mechanism that might account for evolutionary change, namely, natural selection. Muller (1949 : 459) has characterized this contribution as follows:

> Darwin's theory of evolution through natural selection was undoubtedly the most revolutionary theory of all time. It surpassed even the astronomical revolution ushered in by Copernicus in the significance of its implications for our understanding of the nature of the universe and of our place and role in it. . . . Darwin's masterly marshaling of the evidence for this [the ordering effect of natural selection], and his keen-sighted development of many of its myriad facets, remains to this day an intellectual monument that is unsurpassed in the history of human thought.

3. He replaced typological thinking by population thinking.

The first two contributions of Darwin are generally known and sufficiently stressed in the scientific literature. Equally important but almost consistently overlooked is the fact that Darwin introduced into the scientific literature a new way of thinking, "population thinking." What is this population thinking, and how does it differ from typological thinking, the then-prevailing mode of thinking? Typological thinking no doubt had its roots in the earliest efforts of primitive man to classify the bewildering diversity of nature into categories. The *eidos* of Plato is the formal philosophical codification of this form of thinking. According to it, there are a limited number of fixed, unchangeable "ideas" underlying the observed variability, with the *eidos* (idea) being the only thing that is fixed and real, while the observed variability has no more reality than the shadows of an object on a cave wall, as it is stated in Plato's allegory. The discontinuities between these natural "ideas" (types), it was believed, account for the frequency of gaps in nature. Most of the great philosophers of the seventeenth, eighteenth, and nineteenth centuries were influenced by the idealistic philosophy of Plato, and the thinking of this school dominated the thinking of the period. Since there is no gradation between types, gradual evolution is basically a logical impossibility for the typologist. Evolution, if it occurs at all, has to proceed in steps or jumps.

The assumptions of population thinking are diametrically opposed to those of the typologist. The populationist stresses the uniqueness of everything in the organic world. What is true for the human species—that no two individuals are alike—is equally true for all other species of animals and plants. Indeed, even the same individual changes continuously throughout its lifetime and when placed into different environments. All organisms and organic phenomena are composed of unique features and can be described collectively only in statistical terms. Individuals, or any kind of organic entities, form populations of which we can determine only the arithemetic mean and the statistics of variation. Averages are merely statistical abstractions; only the individuals of which the populations are composed have reality. The ultimate conclusions of the population thinker and of the typologist are precisely the opposite. For the typologist, the type (*eidos*) is real and the variation an illusion, while for the populationist the type (average) is an abstraction and only the variation is real. No two ways of looking at nature could be more different.

The importance of clearly differentiating these two basic philosophies and concepts of nature cannot be overemphasized. Virtually every controversy in the field of evolutionary theory, and there are few fields of science with as many controversies, was a controversy between a typologist and a populationist. Let me take two topics, race and natural selection, to illustrate the great difference in interpretation that results when the two philosophies are applied to the same data.

Race

The typologist stresses that every representative of a race has the typical characteristics of that race and differs from all representatives of all other races by the characteristics ''typical'' for the given race. All racist theories are built on this foundation. Essentially, it asserts that every representative of a race conforms to the type and is separated from the representatives of any other race by a distinct gap. The populationist also recognizes races but in totally different terms. Race for him is based on the simple fact that no two individuals are the same in sexually reproducing organisms and that consequently no two aggregates of individuals can be the same. If the average difference between two groups of individuals is sufficiently great to be recognizable on sight, we refer to such groups of individuals as different races. Race, thus described, is a universal phenomenon of nature occurring not only in man but in two thirds of all species of animals and plants.

Two points are especially important as far as the views of the populationist on race are concerned. First, he regards races as potentially overlapping population curves. For instance, the smallest individual of a large-sized race is usually smaller than the largest individual of a small-sized race. In a comparison of races the same overlap will be found for nearly all examined characters. Second, nearly every character varies to a greater or lesser extent independently of the others. Every individual will score in some traits above, in others below the average for the population. An individual that will show in all of its characters the precise mean value for the population as a whole does not exist. In other words, the ideal type does not exist.

Natural Selection

A full comprehension of the difference between population and typological thinking is even more necessary as a basis for a meaningful discussion of the most important and most controversial evolutionary theory—namely, Darwin's theory of evolution through natural selection. For the typologist everything in nature is either "good" or "bad," "useful" or "detrimental." Natural selection is an all-or-none phenomenon. It either selects or rejects, with rejection being by far more obvious and conspicuous. Evolution to him consists of the testing of newly arisen "types." Every new type is put through a screening test and is either kept or, more probably, rejected. Evolution is defined as the preservation of superior types and the rejection of inferior ones, "survival of the fittest" as Spencer put it. Since it can be shown rather easily in any thorough analysis that natural selection does not operate in this described fashion, the typologist comes by necessity to the conclusions: (1) that natural selection does not work, and (2) that some other forces must be in operation to account for evolutionary progress.

The populationist, on the other hand, does not interpret natural selection as an all-or-none phenomenon. Every individual has thousands or tens of thousands of traits in which it may be under a given set of conditions selectively superior or inferior in comparison with the mean of the population. The greater the number of superior traits an individual has, the greater the probability that it will not only survive but also reproduce. But this is merely a probability, because under certain environmental conditions and temporary circumstances, even a "superior" individual may fail to survive or reproduce. This statistical view of natural selection permits an operational definition of "selective superiority" in terms of the contribution to the gene pool of the next generation.

Reference

Muller, H. J. 1949. The Darwinian and modern conceptions of natural selection. *Proc. Amer. Phil. Soc.* 93: 459–470.

17 Evolution, Population Thinking, and Essentialism

Elliott Sober

Philosophers have tended to discuss essentialism as if it were a *global* doctrine—a philosophy which, for some uniform reason, is to be adopted by all the sciences, or by none of them. Popper (1972) has taken a negative global view because he sees essentialism as a major obstacle to scientific rationality. And Quine (1953b, 1960), for a combination of semantical and epistemological reasons, likewise wishes to banish essentialism from the whole of scientific discourse. More recently, however, Putnam (1975) and Kripke (1972) have advocated essentialist doctrines and have claimed that it is the task of each science to investigate the essential properties of its constitutive natural kinds.

In contrast to these global viewpoints is a tradition which sees the theory of evolution as having some special relevance to essentialist doctrines within biology. Hull (1965) and Mayr (1959) are perhaps the two best known exponents of this attitude; they are *local* anti-essentialists. For Mayr, Darwin's hypothesis of evolution by natural selection was not simply a new theory, but a new *kind of theory*—one which discredited essentialist modes of thought within biology and replaced them with what Mayr has called "population thinking." Mayr describes essentialism as holding that

> [t]here are a limited number of fixed, unchangeable "ideas" underlying the observed variability [in nature], with the *eidos* (idea) being the only thing that is fixed and real, while the observed variability has no more reality than the shadows of an object on a cave well.... [In contrast], the populationist stresses the uniqueness of everything in the organic world.... All organisms and organic phenomena are composed of unique features and can be described collectively only in statistical terms. Individuals, or any kind of organic entities, from populations of which we can determine the arithmetic mean and the statistics of variation. Averages are merely statistical abstractions only the individuals of which the population are composed have reality. The ultimate conclusions of the population thinker and of the typologist are precisely the opposite. For the typologist the type (*eidso*) is real and the variation an illusion, while for the populationist, the type (average) is

From *Philosophy of Science* 47 (1980): 350–383.

> an abstraction and only the variation is real. No two ways of looking at nature could be more different. (Mayr 1959, 28–9)

A contemporary biologist reading this might well conclude that essentialists had no scientifically respectable way of understanding the existence of variation in nature. In the absence of this, typologists managed to ignore the fact of variability by inventing some altogether mysterious and unverifiable subject matter for themselves. The notion of *types* and the kind of anti-empiricism that seems to accompany it, appear to bear only the most distant connection with modern conceptions of evidence and argument. But this reaction raises a question about the precise relation of evolution to essentialism. How could the *specifics* of a particular scientific theory have mattered much here, since the main obstacle presented by essentialist thinking was just to get people to be scientific about nature by paying attention to the evidence? The problem was to bring people down to earth by rubbing their noses in the diversity of nature. Viewed in this way, Mayr's position does not look much like a form of *local* anti-essentialism.

Other perplexities arise when a contemporary biologist tries to understand Mayr's idea of population thinking as applying to his or her own activity. If "only the individuals of which the population are composed have reality," it would appear that much of population biology has its head in the clouds. The Lotke-Volterra equations, for example, describe the interactions of predator and prey populations. Presumably, population thinking, properly so called, must allow that there is something real over and above individual organisms. Population thinking countenances organisms and populations; typological thinking grants that both organisms and types exist. Neither embodies a resolute and ontologically austere focus on individual organisms alone. That way lies nominalism, which Mayr (1969) himself rejects.

Another issue that arises from Mayr's conception of typological and population thinking is that of how we are to understand his distinction between "reality" and "abstraction." One natural way of taking this distinction is simply to understand reality as meaning existence. But presumably no population thinker will deny that there are such things as averages. If there are groups of individuals, then there are numerous properties that those groups possess. The *average* fecundity within a population is no more a property which we invent by "mere abstraction" than is the fecundity of individual organisms. Individual and group properties are equally "out there" to be discovered. And similarly, it is unclear how one could suggest that typologists held that variability is unreal; surely the historical record shows that typologists realized that differences between individuals *exist*. How, then, are we to understand the difference between essentialism and population thinking in terms of what each holds to be "real" about biological reality?

Answering these questions about the difference between essentialist and population modes of thought will be the main purpose of this chapter. How did essentialists propose to account for variability in nature? How did evolutionary theory undermine the explanatory strategy that they pursued? In what way does post-Darwinian biology embody a novel conception of variability? How has population thinking transformed our conception of what is *real*? The form of local anti-essentialism which I will propound in what follows will be congenial to many of Mayr's views. In one sense, then, our task will be to explicate and explain Mayr's insight that the shift from essentialist to populationist modes of thinking constituted a shift in the concept of biological reality. However, I will try to show why essentialism was a manifestly *scientific* working hypothesis. Typologists did not close their eyes to variation but rather tried to explain it in a particular way. And the failure of their explanatory strategy depends on details of evolutionary theory in ways which have not been much recognized.[1]

The approach to these questions will be somewhat historical. Essentialism about species is today a dead issue, not because there is no conceivable way to defend it, but because the way in which it was defended by biologists was thoroughly discredited. At first glance, rejecting a metaphysics or a scientific research program because one of its formulations is mistaken may appear to be fallacious. But more careful attention vindicates this pattern of evaluation. It is pie-in-the-sky metaphysics and science to hold on to some guiding principle simply because *it is possible* that there might be some substantive formulation and development of it. Thus, Newtonianism, guided by the maxim that physical phenomena can be accounted for in terms of matter in motion, would have been rejected were it not for the success of particular Newtonian explanations. One evaluates regulative principles by the way in which they regulate the actual theories of scientists. At the same time, I will try in what follows to identify precisely what it is in essentialism and in evolutionary theory that makes the former a victim of the latter. It is an open question to what degree the source of this incompatibility struck working biologists as central. As I will argue at the end of this section, one diagnosis of the situation which seems to have been historically important is much less decisive than has been supposed.

The essentialist's method of explaining variability, I will argue, was coherently formulated in Aristotle, and was applied by Aristotle in both his biology and in his physics. Seventeenth- and eighteenth-century biologists, whether they argued for evolution or against it, made use of Aristotle's natural state model. And to this day, the model has not been refuted in mechanics. Within contemporary biology, however, the model met with less success. Twentieth-century population genetics shows that the model cannot be applied in the way that the essentialist requires. But the natural state model is not wholly without a home in contemporary biology; in fact, the way in which it finds an application there highlights some salient facts about what population thinking amounts to.

An essentialist view of a given species is committed to there being some property which all and only the members of that species possess. Since there are almost certainly only finitely many individuals in any given species, we are quite safe in assuming there is some finitely statable condition which all and only the members of the species satisfy.[2] This could trivially be a list of the spatiotemporal locations of the organisms involved. But the fact that such a condition exists is hardly enough to vindicate essentialism. The essentialist thinks that there is a diagnostic property which any *possible* organism must have if it is to be a member of the species. It cannot be the case that the property in question is possessed by all organisms belonging to *Homo sapiens*, even though there might exist a member of *Homo sapiens* who lacked the trait. It must be necessarily true, and not just accidental, that all and only the organisms in *Homo sapiens* have the characteristic.

However, even this requirement of essentialism is trivially satisfiable. Is it not necessarily true that to be a member of *Homo sapiens* an organism must be a member of *Homo sapiens*? This is guaranteed if logical truths are necessary. But essentialism about biology is hardly vindicated by the existence of logical truths. In a similar vein, if it is impossible for perpetual motion machines to exist, then it is necessarily true that something belongs to *Homo sapiens* if and only if it belongs to *Homo sapiens* or is a perpetual motion machine. This necessary truth is not a truth of logic; it is a result of the theory of thermodynamics. But it too fails to vindicate biological essentialism. What more, then, is required?

The key idea, I think, is that the membership condition must be *explanatory*. The essentialist hypothesizes that there exists some characteristic unique to and shared by all members of *Homo sapiens* which explains why they are the way they are. A species essence will be a causal mechanism which works on each member of the species, making it the kind of thing that it is.

The characterization of essentialism just presented is fairly vague. For one thing, a great deal will depend on how one understands the crucial idea of *explanation*. But since explanation is clearly to be a scientific notion, I hope that, on my sketch, essentialism has the appearance of a scientific thesis, although perhaps one that is not terribly precise Although historically prey to obscurantism, essentialism has nothing essentially to do with mystery mongering, or with the irrational injunction that one should ignore empirical data. It is a perfectly respectable claim about the existence of hidden structures which unite diverse individuals into natural kinds.

Besides its stress on the giving of explanations, there is another feature of our characterization of essentialism which will be important in what follows. The essentialist requires that *a species* be defined in terms of the characteristics of the *organisms* which belong to it. We might call this kind of definition a *constituent definition*; wholes are to be defined in terms of their parts, sets are to be defined in terms of their members, and so on. Pre-Darwinian critics of the species concept, like Buffon and Bonnet, argued that species are unreal, because no such characteristics of organisms can be singled out (see Lovejoy 1936), and pre-Darwinian defenders of the species concept likewise agreed that the concept is legitimate only if constituent definitions could be provided. Constituent definitions are *reductionistic*, in that concepts at higher levels of organization (e.g., species) are legitimate only if they are definable in terms of concepts applying at lower levels of organization (e.g., organisms). It is quite clear that if there are finitely many levels of organization, one cannot demand constituent definitions for concepts at every level of organization (Kripke 1978). As we will see in what follows, evolutionary theory emancipated the species concept from the requirement that it be provided with a constituent definition. The scientific coherence of discourse at the population level of organization was to be assured in another way, one to which the label "population thinking" is especially appropriate.

Chemistry is prima facie a clear case in which essentialist thinking has been vindicated. The periodic table of elements is a taxonomy of chemical kinds. The essence of each kind is its atomic number. Not only is it the case that all actual samples of nitrogen happen to have atomic number 14; it is necessarily the case that a thing is made of nitrogen if and only if it is made of stuff having atomic number 14. Moreover, this characteristic atomic number plays a central role in explaining other chemical properties of nitrogen. Although things made of this substance differ from each other in numerous respects, underlying this diversity there is a common feature. It was hardly irrational for chemists to search for this feature, and the working assumption that such essences were out there to be found, far from stifling inquiry, was a principal contributor to that inquiry's bearing fruit.

Can an equally strong case be made for an essentialist view of biological species? One often hears it said that evolution undermined essentialism because the essentialist held that species are static, but from 1859 on we had conclusive evidence that species evolve. This comment makes a straw man of essentialism and is in any case historically untrue to the thinking of many essentialists. For one thing, notice that the discovery of the transmutation of elements has not in the slightest degree undermined the periodic table. The fact that nitrogen can be changed into oxygen does not in any way show that nitrogen and oxygen lack essences. To be nitrogen is to have one atomic number; to be oxygen is to have another. To change from nitrogen into oxygen, a thing must therefore shift from one atomic number to another. The mere fact of evolution does not show that species lack essences.

As a historical matter, some essentialists, like Agassiz (1859), did assert a connection between essentialism and stasis. But others considered the possibility that new species should have arisen on earth since the beginning (if they thought that there was a beginning). Thus, Linnaeus originally hypothesized that all species were created once and for all at the beginning, but later in his career he changed his mind because he thought that he had discovered a species, *Peloria*, which arose through cross-species hybridization (Rabel 1939, Ramsbottom 1938). And in *Generation of Animals* (II 746a30), Aristotle himself speculates about the possibility of new species arising as fertile hybrids.

Countenancing such species need have no effect on binomial nomenclature or on deciding which characteristics of organisms to view as diagnostic. The question of when there started to be various kinds of things in the universe seems to be quite independent of what makes for differences between kinds.

Another, more plausible, suggestion, concerning how evolution undermined essentialism, is this: The fact that species evolve *gradually* entails that the boundaries of species are vague. The essentialist holds that there are characteristics which all and only the members of a given species possess. But this is no longer a tenable view; it is just as implausible as demanding that there should be a precise number of dollars which marks the boundary between rich and poor. This is the Sorites problem. Since ancient Greece, we have known that being a heap of stones, being bald, and being rich are concepts beset by line-drawing problems. But, the suggestion goes, it was only since 1859 that we have come to see that *Homo sapiens* is in the same boat. Thus, Hull (1965) has argued that essentialism was refuted because of its Aristotelian theory of *definition*; the requirement that species have nontrivial necessary and sufficient conditions runs afoul of the kind of continuity found in nature.

Unfortunately, this limpid solution to our problem becomes clouded a bit when we consider the historical fact that many essentialists conceded the existence of line-drawing problems. Thus, Aristotle in his *History of Animals*, (5888b4 ff.), remarks:

> nature proceeds little by little from inanimate things to living creatures, in such a way that we are unable, in the continuous sequence to determine the boundary line between them or to say which side an intermediate kind falls. Next, after inanimate things come the plants: and among the plants there are differences between one kind and another in the extent to which they seem to share in life, and the whole genus of plants appears to be alive when compared with other objects, but seems lifeless when compared with animals. The transition from them to the animals is a continuous one, as remarked before. For with some kinds of things found in the sea one would be at a loss to tell whether they are animals or plants.

It is unclear exactly how one should interpret this remark. Does it indicate that there are in fact no boundaries in nature, or does it mean that the boundaries are difficult to discern? From the time of Aristotle up to the time of Darwin, the principle of continuity seems to have coexisted peacefully with strong essentialist convictions in the minds of many thinkers (Lovejoy 1936). Bonnet, Akenside, and Robinet are eighteenth-century biologists who exemplify this curious combination of doctrines. Does this coexistence imply that the two doctrines are in fact compatible, or rather, does it show that their conceptual dissonance was a long time in being appreciated? To answer this question, let us return to our analogy with the transmutation of elements.

In what sense are the boundaries between chemical kinds any more definite than those which we encounter in biology? At first glance, there appears to be all the difference in the world: in the periodic table, we have discrete jumps—between atomic number 36 and atomic number 37 there are no intermediate atomic numbers to blur distinctions. But let us reflect for a moment on the mechanism of transmutation. Consider, as an example, the experiment which settled the question of how nitrogen can be transmuted into oxygen (Ihde 1964, 509):

$${}^{4}_{2}\mathrm{He} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{17}_{8}\mathrm{O} + {}^{1}_{1}\mathrm{H}.$$

In this reaction, the α-particle is absorbed and a proton is expelled. Let us ask of this process a typical Sorites question: At what point does the bombarded nucleus cease to be a nitrogen nucleus and when does it start being a nucleus of oxygen?

There *may* be a precise and principled answer to this question which is given by the relevant physical theory. But then again there may not.[3] I would suggest that which of these outcomes prevails really does not matter to the question of whether essentialism is a correct doctrine concerning the chemical kinds. It may well be that having a particular atomic number is a vague concept. But this is quite consistent with that (vague) property's being the essence of a chemical kind. This really does not matter, as long as the vagueness of "nitrogen" and that of "atomic number 14" coincide. Essentialism is in principle consistent with *vague essences*.[4] In spite of this, one wonders what the history of chemistry, and its attendant metaphysics, would have looked like, if the transmutation of elements had been a frequent and familiar phenomenon during the second half of the nineteenth century. Just as the fact of evolution at times tempted Darwin to adopt a nominalist attitude toward species, so in chemistry the impressive taxonomy which we now have in the form of the periodic table might never have been arrived at, line-drawing problems having convinced chemists that chemical kinds are unreal.[5]

As a historical matter, Hull (1965) was right in arguing that essentialism was standardly associated with a theory of definition in which vagueness is proscribed. Given this association, nonsaltative evolution was a profound embarassment to the essentialist. But, if I am right, this theory of definition is inessential to essentialism. Our argument that the gradualness of evolution is not the decisive issue in undermining essentialism is further supported, I think, by the fact that contemporary evolutionary theory contains proposals in which evolutionary gradualism is rejected. Eldredge and Gould (1972) have argued that the standard view of speciation (as given, for example, in Ayala 1978 and Mayr 1963) is one in which phylogeny is to be seen as a series of "punctuated equilibria." Discontinuities in the fossil record are not to be chalked up to incompleteness, but rather to the fact that, in geological time, jumps are the norm. I would suggest that this theory of discontinuous speciation is cold comfort to the essentialist. Whether lines are easy or hard to draw is not the main issue, or so I shall argue.[6]

Another local anti-essentialist argument has been developed by Ghiselin (1966, 1969, 1974) and Hull (1976, 1978). They have argued that evolutionary theory makes it more plausible to view species as spatiotemporally extended individuals than as natural kinds. A genuine natural kind like gold may ''go extinct'' and then reappear; it is quite possible for there to be gold things at one time, for there to be no gold at some later time, and then, finally, for gold to exist at some still later time. But the conception of species given by evolutionary theory does not allow this sort of flip-flopping in and out of existence: once a biological taxon goes extinct, it must remain so. Hull (1978) argues that the difference between chemical natural kinds and biological species is that the latter, but not the former, are historical entities. Like organisms, biological species are individuated in part by historical criteria of spatiotemporal continuity. I am inclined to agree with this interpretation; its impact on pre-Darwinian conceptions of species could hardly be more profound. But what of its impact on essentialism? If essentialism is simply the view that species have essential properties (where a property need not be purely qualitative), then the doctrine remains untouched (as Hull himself realizes). Kripke (1972) has suggested that each individual human being has the essential property of being born of precisely the sperm and the egg of which he or she was born. If such individuals as organisms have essential properties, then it will presumably also be possible for individuals like *Drosophila melanogaster* to have essential properties as well. Of course, these essences will be a far cry from the ''purely qualitative'' characteristics which traditional essentialism thought it was in the business of discovering.

My analysis of the impact of evolutionary theory on essentialism is parallel, though additional. Whether species are natural kinds or spatiotemporally extended individuals, essentialist theories about them are untenable. Two kinds of arguments will be developed for this conclusion. First, I will describe the way in which essentialism seeks to explain the existence of variability, and will argue that this conception is rendered implausible by evolutionary theory. Second, I will show how evolutionary theory has removed *the need* for providing species with constituent definitions; population thinking provides another way of making species scientifically intelligible. This consideration, coupled with the principle of parsimony, provides an additional reason for thinking that species do not have essences.

Aristotle's Natural State Model

One of the fundamental ideas in Aristotle's scientific thinking is what I will call his natural state model. This model provides a technique for explaining the great diversity found in natural objects. Within the domain of physics, there are heavy and light objects, ones that move violently and ones that do not move at all. How is one to find some order that unites and underlies all this variety? Aristotle's hypothesis was that there is a distinction between the *natural state* of a kind of object and those states which are not natural. These latter are produced by subjecting the object to an *interfering force*. In the sublunar sphere, for a heavy object to be in its natural state is for it to be located where the center of the Earth is now (*On the Heavens*, ii, clr 296b and 310b, 2–5). But, of course, many heavy objects fail to be there. The cause for this divergence from what is natural is that these objects are acted on by interfering forces which prevent them from achieving their natural state by frustrating their natural tendency. Variability within nature is thus to be accounted for as a deviation from what is natural; were there no interfering forces, all heavy objects would be located in the same place (Lloyd 1968).

Newton made use of Aristotle's distinction, but disagreed with him about what the natural state of physical objects is. The first law of motion says that if a body is not acted upon by a force, then it will remain at rest or in uniform motion. And even in general relativity, the geometry of space-time specifies a set of geodesics along which an object will move as long as it is not subjected to a force. Although the terms "natural" and "unnatural" no longer survive in Newtonian and post-Newtonian physics, Aristotle's distinction can clearly be made within those theories. If there are no forces at all acting on an object, then, *a fortiori*, there are no interfering forces acting on it either. A natural state, within these theories, is a zero-force state.

The explanatory value of Aristotle's distinction is fairly familiar. If an object is not in its natural state, we know that the object must have been acted on by a force, and we set about finding it. We do this by consulting our catalog of known forces. If none of these is present, we might augment our catalog, or perhaps revise our conception of what the natural state of the system is. This pattern of analysis is used in population genetics under the rubric of the Hardy-Weinberg law. This law specifies an equilibrium state for the frequencies of genotypes in a panmictic population; this natural state is achieved when the evolutionary forces of mutation, migration, selection, and drift are not at work.

In the biological world, Aristotle sets forth the same sort of explanatory model. Diversity was to be accounted for as the joint product of natural regularities and interfering forces. Aristotle invokes this model when he specifies the regularities governing how organisms reproduce themselves: "[for] any living thing that has reached its normal development and which is unmutilated, and whose mode of generation is not spontaneous, the most natural act is the production of another like itself, an animal producing an animal, a plant a plant" (*De Anima*, 415a26). Like producing like, excepting the case of spontaneous generation, is the natural state, subject to a multitude of interferences, as well shall see.

In the case of spontaneous generation, the natural state of an organism is different. Although in the *Metaphysics* and the *Physics* "spontaneous" is used to mean unusual or random, in the later biological writings, *History of Animals* and *Generation of Animals*, Aristotle uses the term in a different way (Balme 1962, Hull 1967). Spontaneous generation obeys its own laws. For a whole range of organisms classified between the intermediate animals and the plants, like *never* naturally produces like. Rather, a bit of earth will spontaneously generate an earthworm, and the earthworm will then produce an eel. Similarly, the progression from slime to ascarid to gnat and that from cabbage leaf to grub to caterpillar to chrysallis to butterfly likewise counts as the natural reproductive pattern for this part of the living world (*History of Animals*, 570a5, 551b26, 551a13).

So much for the natural states. What counts as an interference for Aristotle? According to Aristotle's theory of sexual reproduction, the male semen provides a set of instructions which dictates how the female matter is to be shaped into an organism.[7] Interference may arise when the form fails to completely master the matter. This may happen, for example, when one or both parents are abnormal, or when the parents are from different species, or when there is trauma during fetal development Such interferences are anything but rare, according to Aristotle. Mules—sterile hybrids—count as deviations from the natural state (*Generation of Animals*, ii, 8). In fact, the females of a species do too, even though they are necessary for the species to reproduce itself (*Generation of Animals*, ii, 732a; ii, 3, 737a27; iv, 3, 767b8; iv, 6, 775a15). In fact, reproduction that is completely free of interference would result in an offspring which exactly resembles the father.[8] So failure to exactly resemble the male parent counts as a departure from the natural state. Deviations from type, whether mild or extreme, Aristotle labels "*terata*"—monsters. They are the result of interfering forces (*biaion*) deflecting reproduction from its natural pattern.

Besides trying to account for variation within species by using the natural state model, Aristotle at time seems to suggest that there are entire species which count as monsters (Preuss 1975, 215–16; Hull 1968). Seals are deformed as a group because they resemble lower classes of animals, owing to their lack of ears. Snails, since they move like animals with their feet cut off, and lobsters, because they use their claws for locomotion, are likewise to be counted as monsters (*Generation of Animals*, 19, 714b, 18–19; *Parts of Animals*, iv, 8, 684a35). These so-called dualizing species arise because they are the best possible organisms that can result from the matter out of which they are made. The scale of nature, it is suggested, arises in all its graduated diversity because the quality of the matter out of which organisms are made also varies—and nature persists in doing the best possible, given the ingredients at hand.

One cannot fault Aristotle for viewing so much of the biological domain as monstrous. Natural state models habitually have this characteristic; Newton's first law of motion is not impugned by the fact that no physical object is wholly unaffected by an outside force. Even so, Aristotle's partition of natural state and non-natural state in biology sounds to the modern ear like a reasonable distinction run wild. "Real terrata are one thing," one might say, "but to call entire species, and all females, and all males who don't exactly resemble their fathers monsters, seems absurd." Notice that our "modern" conceptions of health and disease and our notion of normality as something other than a statistical average enshrine Aristotle's model. We therefore are tempted to make only a conservative criticism of Aristotle's biology: we preserve the form of model he propounded, but criticize the applications he made of it. Whether this minimal critique of Aristotle is possible in the light of evolutionary theory remains to be seen.

The natural state model constitutes a powerful tool for accounting for variation. Even when two species seem to blend into each other continuously, it may still be the case that all the members of one species have one natural tendency while the members of the other species have a quite different natural tendency. Interfering forces may, in varying degrees, deflect the individuals in both species from their natural states, thus yielding the surface impression that there are no boundaries between the species. This essentialist response to the fact of diversity has the virtue that it avoids the ad hoc maneuver of contracting the boundaries of species so as to preserve their internal homogeneity.[9] This latter strategy was not unknown to the essentialist, but its methodological defects are too well known to be worth recounting here. Instead of insisting that species be defined in terms of some surface morphological feature, and thereby having each species shrink to a point, the essentialist can countenance unlimited variety in, and continuity between, species, as long as underlying this plenum one can expect to find discrete natural tendencies. The failure to discover such underlying mechanisms is no strong reason to think that none exists; but the development of a theory which implies that natural tendencies are not part of the natural order is another matter entirely.

Aristotle's model was a fixed point in the diverse conjectures to be found in pre-Darwinian biology. Preformationists and epigeneticists, advocates of evolution and proponents of stasis, all assumed that there is a real difference between natural states and states caused by interfering forces. The study of monstrosity—teratology—which in this period made the transition from unbridled speculation to encyclopedic catalogues of experimental oddities (Meyer 1939), is an especially revealing example of the power exerted by the natural state model. Consider, for example, the eighteenth-century disagreement between Maupertuis and Bonnet over the proper explanation of polydactyly. Both had at their fingertips a genealogy; it was clear to both that somehow or other the trait regularly reappeared through the generations. Maupertuis conjectured that defective hereditary material was passed along, having originally made its appearance in the family because of *an error in nature* (Glass 1959b, 62–67). Maupertuis, a convinced Newtonian, thought that traits, both normal and anomalous, resulted from the lawful combination of hereditary particles (Roger 1963). When such particles have normal quantities of attraction for each other, normal characteristics result. However, when particles depart from this natural state, either too many or too few of them combine, thus resulting in *monstres par exces* or *monstres par defaut*. Bonnet, a convinced ovist, offered a different hypothesis. For him, polydactyly is never encoded in the germ, but rather results from abnormal interuterine conditions or from male sperm interfering with normal development (Glass 1959a, 169). Thus whether polydactyly is ''naturalized'' by Maupertuis's appeal to heredity or by Bonnet's appeal to environment, the trait is never regarded as being completely natural. Variability in nature—in this case variability as to the number of digits—is a deviation from type.

In pre-Darwinian disputes over evolution, natural states loom equally large. Evolutionary claims during this period mainly assumed that living things were programmed to develop in a certain sequence, and that the emergence of biological novelty was therefore in conformity with some natural plan. Lovejoy (1936) discusses how the Great Chain of Being was ''temporalized'' during the eighteenth century; by this, he has in mind the tendency to think that the natural ordering of living things from those of higher type down to those of lower type also represented a historical progression. Such programmed, directed evolution—in which some types naturally give rise to others—is very much in the spirit of the natural state model. Whether species are subject to historical unfolding, or rather exist unchanged for all time, the concept of species was inevitably associated with that of type; on either view, variation is deviation caused by interfering forces.

It was generally presupposed that somewhere within the possible variations that a species is capable of, there is a privileged state—a state which has a special causal and explanatory role. The laws governing a species will specify this state, just as the laws which make sense of the diversity of kinematic states found in physics tell us what is the natural state of a physical object. The diversity of individual organisms is a veil which must be penetrated in the search for invariance. The transformation in thinking which we will trace in the next two sections consisted in the realization that this diversity itself constituted an invariance, obeying its own laws.

The Law of Errors and the Emergence of Population Thinking

So far, I have sketched several of the applications that have been made of Aristotle's model within biology. This strategy for explaining variation, I will argue in the next section, has been discredited by modern evolutionary theory. Our current theories of biological variation provide no more role for the idea of natural state than our current physical theories do for the notion of absolute simultaneity. Theories in population genetics enshrine a different model of variation, one which only became possible during the second half of the nineteenth century. Some brief account of the evolution within the field of statistics of our understanding of the law of errors will lay the groundwork for discussing the modern understanding of biological variation.

From its theoretical formulation and articulation in the eighteenth century, up until the middle of the nineteenth century, the law of errors was understood as a law about *errors*. Daniel Bernouilli, Lagrange, and Laplace each tried to develop mathematical techniques for determining how a set of discordant observations was to be interpreted

(Todhunter 1865). The model for this problem was, of course, that there is a single true value for some observational variable, and a multiplicity of inconsistent readings that have been obtained. Here we have a straightforward instance of Aristotle's model: interfering forces cause variation in opinion; in nature there is but one true value. The problem for the theory of errors was to penetrate the veil of variability and to discover behind it the single value which was the constant cause of the multiplicity of different readings. Each observation was thus viewed as the causal upshot of two kinds of factors: part of what determines an observational outcome is the real value of the variable, but interfering forces which distort the communication of this information from nature to mind, also play a role. If these interfering forces are random—if they are as likely to take one value as any other—then the mean value of the readings is likely to represent the truth, when the number of observations is large. In this case, one reaches the truth by ascending to the summit of the bell curve. It is important to notice that this application of the natural state model is epistemological, not ontological. One seeks to account for variation in our observations of nature, not variation in nature itself. The decisive transition, from this epistemological to an ontological application, was made in the 1830s by the influential Belgian statistician Adolphe Quetelet.

Quetelet's insight was that the law of errors could be given an ontological interpretation by invoking a distinction which Laplace had earlier exploited in his work in Newtonian mechanics.[10] Laplace decomposed the forces at work in the solar system into two kinds. First, there are the *constant causes* by which the planets are affected by the sun's gravitation; second, there are the particular *disturbing causes* which arise from the mutual influences of the planets, their satellites, and the comets. Laplace's strategy was a familiar analytic one. He tried to decompose the factors at work in a phenomenon into components, and to analyze their separate contributions to the outcome. The character of this decomposition, however, is of special interest: one central, causal agent is at work on the components of a system, but the effects of this force are complicated by the presence of numerous interferences which act in different directions.

In his book of 1835, *Sur l'homme et le développement de ses facultés, ou essai de physique social*, Quetelet put forward his conception of the *average man* which for him constituted the true subject of the discipline of social physics. By studying the average man, Quetelet hoped to filter out the mutifarious and idiosyncratic characteristics which make for diversity in a population, and to focus on the central facts which constitute the social body itself. Like Weber's later idea of an ideal type, Quetelet's conception of the average man was introduced as a "fiction" whose utility was to facilitate a clear view of social facts by allowing one to abstract from the vagaries of individual differences. But unlike Weber, Quetelet quickly came to view his construct as real—a subject matter in its own right. Quetelet was struck by the analogy between a society's average man and a physical system's center of gravity. Since the latter could play a causal role, so too could the former; neither was a mere abstraction. For Quetelet, variability within a populations is *caused* by deviation from type. When the astronomer John Herschel reviewed Quetelet's *Lettres sur les probabilités* in 1850, he nicely captured Quetelet's idea that the average man is no mere artifact of reflection:

> An average may exist of the most different objects, as the heights of houses in a town, or the sizes of books in a library. It may be convenient to convey a general notion of the things averaged; but it involves no conception of a natural and recognizable central magnitude, all differences from which ought to be regarded as deviations from a standard. The notion of a mean, on the other hand, does imply such a conception, standing distinguished from an average by this very feature, *viz*. The regular marching of the groups, increasing to a maximum and thence again diminishing. An average gives us no assurance that the future will be like the past. A mean may be reckoned on with the most implicit confidence. (Hilts 1973, 217)

Quetelet found little theoretical significance in the fact of individual differences. Concepts of correlation and amount of variation were unknown to him. For Quetelet, the law of errors is still a law about errors, only for him the mistakes are made by nature, not by observers. Our belief that there is variation in a population is no mistake on our part. Rather, it is the result of interferences confounding the expression of a prototype. Were interfering forces not to occur, there would be no variation.

It may strike the modern reader as incredible that anyone could view a trait like girth on this mode. However, Quetelet, who was perhaps the most influential statistician of his time, did understand biological differences in this way. He was impressed, not to say awe struck, by the fact that the results of accurately measuring the waists of a thousand Scottish soldiers would assume the same bell-shaped distribution as the results of inaccurately measuring the girth of a single, average, soldier a thousand times. For Quetelet, the point of attending to variation was to *see through* it—to render it transparent. Averages were the very antitheses of artifacts; they alone were the true objects of inquiry.[11]

Frances Galton, Darwin's cousin, was responsible for fundamental innovations in the analysis of individual differences.[12] He discovered the standard deviation and the correlation coefficient. His work on heredity was later claimed by both Mendelians and biometricians as seminal, and thus can be viewed as a crucial step toward the synthetic theory of evolution (Provine 1971). But his interest to our story is more restricted. Galton, despite his frequently sympathetic comments about the concept of type, helped to displace the average man and the idea of deviation from type.[13] He did this, not by attacking these typological constructs directly, but by developing an alternative model for accounting for variability. This model is a nascent form of the kind of population thinking which evolutionary biologists today engage in.

One of Galton's main intellectual goals was to show that heredity is a central cause of individual differences. Although the arguments which Galton put forward for his hereditarian thesis were weak, the conception of variability he exploited in his book *Hereditary Genius* (1869) is of great significance. For Galton, variability is not to be explained away as the result of interference with a single prototype. Rather, variability within one generation is explained by appeal to variability in the previous generation and to facts about the transmission of variability. Galton used the law of errors, but no longer viewed it as a law *about* errors. As Hilts (1973, 223–24) remarks: "Because Galton was able to associate the error distribution with individual differences caused by heredity, the distinction between constant and accidental causes lost much of its meaning." At the end of his life, Galton judged that one of his most important ideas was that the science of heredity should be concerned with deviations measured in statistical units. Quetelet had earlier denied that such units exist. Galton's discovery of the standard deviation gave him the mathematical machinery to begin treating variability as obeying its own laws, as something other than an idiosyncratic artifact.

Eight years after the publication of *Hereditary Genius*, Galton was able to sketch a solution for the problem he had noted in that work: What fraction of the parental deviations from the norm is passed on to offspring? Galton described a model in which hereditary causes and nonhereditary causes are partitioned. Were only the former of these at work, he conjectured, each child would have traits that are intermediate between those of its parents. In this case, the amount of variation would decrease in each generation. But Galton suspected that the amount of variation is constant across generations. To account for this, he posited a second, counteracting force which causes variability within each family. Were this second force the only one at work, the amount of variation would increase. But in reality, the centrifugal and centripetal forces combine to yield a constant quantity of variability across the generations. An error distribution is thus accounted for by way of a hypothesis which characterizes it as the sum of two other error distributions.

In his *Natural Inheritance* of 1889, Galton went on to complete his investigations of the correlation coefficient, and introduced the name "normal law" as a more appropriate label for what had traditionally been called the law of errors.[14] Bell curves are normal; they are found everywhere, Galton thought. This change in nomenclature crystallized a significant transformation in thinking. Bell curves need not represent mistakes made by fallible observers or by sportive nature. Regardless of the underlying etiology, they *are real*; they enter into explanations because the variability they represent is lawful and causally efficacious.

The transition made possible by statistical thinking from typological to population thinking was not completed by Galton.[15] Although his innovations loosened the grip of essentialism, he himself was deeply committed to the idea of racial types and believed that evolutionary theory presupposes the reality of types. Both Galton and Darwin (1859, ch. 5; 1868, ch. 13) spoke sympathetically about the ideas of unity of type and of reversion to type, and sought to provide historical justifications of these ideas in terms of common descent. Unity of type was just similarity owing to common ancestry; reversion to type was the reappearance of latent ancestral traits.

But the presence of these ideas in their writings should not obscure the way in which their theorizing began to undermine typological thinking.

Darwin and Galton focused on the population as a unit of organization. The population is an entity, subject to its own forces, and obeying its own laws. The details concerning the individuals who are parts of this whole are pretty much irrelevant. Describing a single individual is as theoretically peripheral to a populationist as describing the motion of a single molecule is to the kinetic theory of gases. In this important sense, population thinking involves *ignoring individuals*: it is holistic, not atomistic. This conclusion contradicts Mayr's (1959, 28) assertion that for the populationist, "the individual alone is real."

Typologists and populationists agree that averages exist; and both grant the existence of variation. They disagree about the explanator character of these. For Quetelet, and for typologists generally, variability does not explain anything. Rather it is something to be explained or explained away. Quetelet posited a process in which uniformity gives rise to diversity; a single prototype—the average man—is mapped onto a variable resulting population. Galton, on the other hand, explained diversity in terms of an earlier diversity and constructed the mathematical tools to make this kind of analysis possible.

Both typologists and populationists seek to transcend the blooming, buzzing confusion of individual variation. Like all scientists, they do this by trying to identify properties of systems which remain constant in spite of the system's changes. For the typologist, the search for invariances takes the form of a search for natural tendencies. The typologist formulates a causal hypothesis about the forces at work on each individual within a population. The invariance underlying this diversity is the possession of a particular natural tendency *by each individual organism*. The populationist, on the other hand, tries to identify invariances by ascending to a different level of organization. For Galton, the invariant property across generations within a lineage is the amount of variability, and this is a property *of populations*. Again we see a way in which the essentialist is more concerned with individual organisms than the populationist is. Far from ignoring individuals, the typologist, *via* his use of the natural state model, resolutely focuses on individual organisms as the entities which possess invariant properties. The populationist, on the other hand, sees that it is not just individual organisms which can be the bearers of unchanging characteristics. Rather than looking for a reality that *underlies* diversity, the populationist can postulate a reality *sustained* by diversity.

I have just argued that there is an important sense in which typologists are more concerned with individual organisms than populationists are. However, looked at in another way, Mayr's point that populationists assign a more central role to organisms than typologists do can be established. In models of natural selection in which organisms enjoy different rates of reproductive success because of differences in fitness, natural selection is a force that acts on individual (organismic) differences. This standard way of viewing evolution assigns a causal role to individual idiosyncrasies. Individual differences are not *the effects* of interfering forces confounding the expression of a prototype; rather they are *the causes* of events that are absolutely central to the history of evolution. It is in this sense that Mayr is right in saying that evolutionary theory treats individuals as real in a way that typological thought does not (see also Lewontin 1974, 5–6). Putting my point and Mayr's point, thus interpreted, together, we might say that population thinking endows individual organisms with more reality *and* with less reality than typological thinking attributes to them.

To be real is to have causal efficacy; to be unreal is to be a mere artifact of some causal process. This characterization of what it is to be real, also used by Hacking (1975), is markedly different from the one used in traditional metaphysical disputes concerning realism, verificationism, and idealism (Sober 1980b). There, the problem is not how things are causally related, but rather it concerns what in fact *exists*, and whether what exists exists "independently" of us. The causal view of what it is to be real offers an explanation of a peculiar fact that is part of the more traditional metaphysical problem. Although two predicates may name real physical properties, natural kinds, theoretical magnitudes, or physical objects, simple operations on that pair of predicates may yield predicates which fail to name anything real. Thus, for example, "mass" and "charge" may name real physical magnitudes, even though "$\text{mass}^2/\text{charge}^3$" fails to name anything real. This is hard to explain, if reality is simply equated with existence (or with existence-that-is-independent-of-us). After all, if an object has a mass and if it has a charge, then there must be such a thing as what the square of its mass over the cube of its charge is. While this is quite true, it is *not* similarly correct to infer that because an object's mass causes some things and its charge causes other things, then there must be something which is caused by appeal to the square of its mass divided by the cube of its charge. Realism, in this case at least, is a thesis about what is cause and what is effect.

If we look forward in time, from the time of Galton and Darwin to the Modern Synthesis and beyond, we can see how population models have come to play a profoundly important role in evolutionary theorizing. In such models, properties of populations are identified and laws are formulated about their interrelations. Hypotheses in theoretical ecology and in island biogeography, for example, *generalize over populations* (see, for example, Wilson and Bossert 1971, chs. 3 and 4). The use of population concepts is not legitimized in those disciplines by defining them in terms of concepts applying at some lower level of organization. Rather, the use of one population concept is vindicated by showing how it stands in lawlike relations with other concepts *at the same level of organization*. It is in this way that we can see that there is an alternative to constituent definition. Here, then, is one way in which evolutionary theorizing undermined essentialism: Essentialism requires that species concepts be legitimized by constituent definition, but evolutionary theory, in its articulation of population models, makes such demands unnecessary. Explanations can proceed without this reductionistic requirement's being met.

If this argument is correct, there is a standard assumption made in traditional metaphysical problems having to do with identity which needs to be reevaluated. There could hardly be a more central category in our metaphysics, both scientific and everyday, than that of an enduring physical object. The way philosophers have tried to understand this category is as follows: Imagine a collection of instantaneous objects—i.e., objects at a moment in time. How are these various instantaneous objects united into the temporally enduring objects of our ontology? What criteria do we use when we lump together some time slices, but not others? This approach to the problem is basically that of looking for a constituent definition: enduring objects are to be defined out of their constituent time-slices. But, if populations can be scientifically legitimized in ways other than by using constituent definitions, perhaps the same thing is true of the category of physical object itself. I take it that Quine's (1953a) slogan "no entity without identity" is basically a demand for constituent definitions; this demand, which has been so fruitful in mathematics, should not be generalized into a universal maxim (nor can it be, if there are finitely many levels of organization. See Kripke 1978).

If constituent definitions for population concepts are theoretically unnecessary, then we have one argument, via the principle of parsimony (Sober 1980a), for the view that species do not have essences. However, there are equally pressing problems which essentialism faces when the natural state model is evaluated in the light of our current understanding of the origins of variability. It is to these problems that we now turn.

The Disappearance of a Distinction

The fate of Aristotle's model at the hands of population biology bears a striking resemblance to what happened to the notion of absolute simultaneity with the advent of relativity theory. Within classical physics, there was a single, well-defined answer to the question, "What is the temporal separation of two events *x* and *y*?" However, relativity theory revealed that answering this question at all depends on one's choice of a rest frame; given different rest frames, one gets different answers. We might represent the way the temporal separation of a pair of events may depend on a choice of frame as in the graph in figure 17.1. As is well known, the classical notions of temporal separation and spatial separation gave way in relativity theory to a magnitude that is not relative at all: this is the spatiotemporal separation of the two events. How large this quantity is does not depend on any choice of rest frame; it is frame invariant. Minkowski (1908) took this fact about relativity theory to indicate that space and time are not real physical properties at all, since they depend for their values on choices that are wholly arbi-

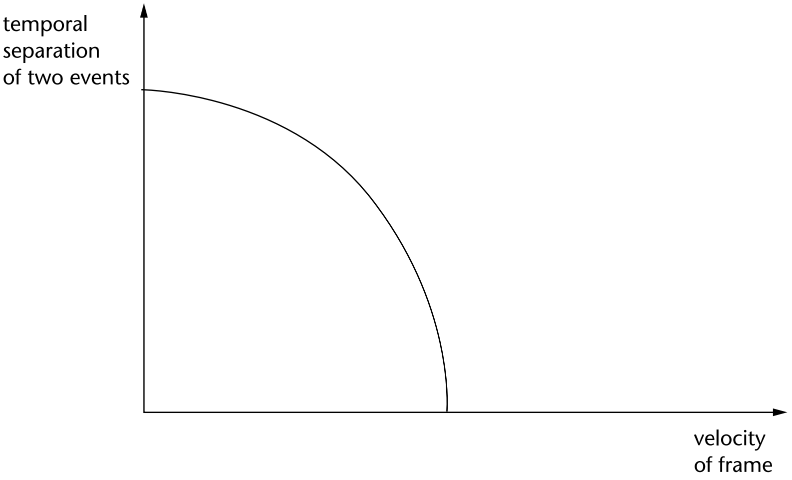


Figure 17.1
The temporal separation of a pair of events, relative to choices of rest frame.

trary. For Minkowski, to be real is to be invariant, and space and time become mere shadows.

Special relativity fails to discriminate between the various temporal intervals represented in figure 17.1; they are all on a par. No one specification of the temporal separation is any more correct than any other. It would be utterly implausible to interpret this fact as indicating that there is a physically real distinction which special relativity fails to make. The fact that our best theory fails to draw this distinction gives us a very good reason for suspecting that the distinction is unreal, and this is the standard view of the matter which was crystallized in the work of Minkowski.

According to the natural state model, there is one path of fetal development which counts as the realization of the organism's natural state, while other developmental results are consequences of unnatural interferences. Put slightly differently, for a given genotype, there is a single phenotype which it can have that is the natural one. Or, more modestly, the requirement might be that there is some restricted range of phenotypes which count as natural. But when one looks to genetic theory for a conception of the relation between genotype and phenotype, one finds no such distinction between natural state and states which are the results of interference. One finds, instead, the *norm of reaction*, which graphs the different phenotypic results that a genotype can have in different environments.[16] Thus the height of a single corn plant genotype might vary according to environmental differences in temperature, as is shown in figure 17.2. How would one answer the question: "Which of these phenotypes is the natural one for the corn plant to have?" One way to take this obscure question is indicated by the following answer: Each of the heights indicated in the norm of reaction is

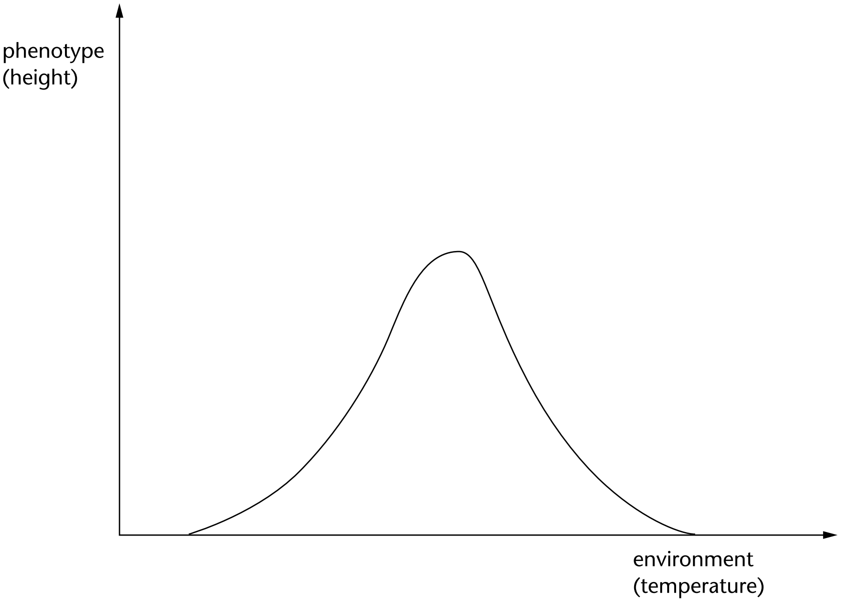


Figure 17.2
The norm of reaction of a given corn plant genotype, showing height as a function of temperature.

as "natural" as any other, since each happens in nature. Choose an environment, and relative to that choice we know what the phenotypic upshot in that environment is. But, of course, if the question we are considering is understood in terms of the natural state model, this sort of answer will not do. The natural state model presupposes that there is some phenotype which is the natural one *which is independent of a choice of environment.* The natural state model presupposes that there is some environment which is the natural environment for the genotype to be in, which determines, in conjunction with the norm of reaction, what the natural phenotype for the genotype is. But these presuppositions find no expression in the norm of reaction: all environments are on a par, and all phenotypes are on a par. The required distinctions simply are not made.

When one turns from the various phenotypes that a single genotype might produce, to the various genotypes that a population might contain the same result obtains. Again, according to the natural state model, there is a single genotype or restricted class of genotypes, which count as the natural states of the population or species, all other genotypes being the result of interfering forces. But again, statistical profiles of genotypic variance within a population enshrine no such difference. Genotypes differ from each other in frequency; but unusual genotypes are not in any literal sense to be understood as deviations from type.

When a corn plant of a particular genotype withers and dies, owing to the absence of trace elements in the soil, the natural state model will view this as an outcome that is not natural. When it thrives and is reproductively successful, one wants to say that *this* environment might be the natural one. Given these ideas, one might try to vindicate the natural state model from a selectionist point of view by identifying the natural environment of a genotype with the environment in which it is fittest.[17]

This suggestion fails to coincide with important intuitions expressed in the natural state model. First of all, let us ask the question: What is the range of environments relative to which the fittest environment is to be understood? Shall we think of the natural state as that which obtains when the environment is the fittest *of all possible environments*? If so, the stud bull, injected with medications, its reproductive capacities boosted to phenomenal rates by an efficient artificial insemination program, has achieved its natural state. And in similar fashion, the kind of environment that biologists use to characterize the intrinsic rate of increase (*r*) of a population—one in which there is no disease, no predation, no limitations of space or food supplies—will likewise count as the natural environment. But these optimal environments are *not natural*, the natural state model tells us. They involve ''artificially boosting'' the fitness of resulting phenotypes by placing the genotypes in environments that are more advantageous than the natural environment.

Let us consider another, perhaps more plausible, way to understand the range of environments with respect to which the fittest environment is to be calculated. Instead of taking the best of all possible environments, why not, more modestly, consider the best of all environments that have been historically represented? This suggestion evades the second, but not the first, counterexample mentioned above. However, other problems present themselves. The natural state of a genotype is often understood to be one which has yet to occur. Perhaps every environment that a species has historically experienced is such that a given genotype in that environment results in a *diseased* phenotypes, or one which is developmentally impaired in some way. The natural state of a genotype is often taken to be some sort of ideal state which may or may not be closely approximated in the history of the species.

I have just argued that the idea of a fittest environment does not allow one to impose on the norm of reaction the kind of distinction that the natural state model requires. Precisely the same reasons count against construing the idea of a genotype's being the natural state of a species in terms of maximal fitness. It is part of the natural state model that the natural genotypes for a species can be less fit (in some range of environments) than the best of all possible genotypes. And the natural genotype can likewise fail to be historically represented.

Aristotle is typical of exponents of the natural state model in holding that variation is introduced into a population by virtue of interferences with normal sexual reproduction. Our current understanding of the mechanisms of reproduction shows that precisely the opposite is the case. Even if one dismisses mutations as ''unnatural interferences,'' the fact of genetic recombination in meiosis looms large. Generally, the number of total genotypes that a gene pool can produce by recombination is the product of the number of diploid genotypes that can be constructed at each locus. For species like *Homo sapiens* and *Drosophila melanogaster*, the number of loci has been estimated to be about 10,000 or more. What this means is that the number of genotypes that can be generated by recombination is greater than the number of atoms in the visible universe (Wilson and Bossert 1971, 39). For species with this number of loci, even a single male and a single female can themselves reproduce a significant fraction of the variation found in a population from which they are drawn. All sorts of deleterious phenotypes may emerge from the recombination process initiated by a founder population.

A doctrinaire advocate of the natural state model may take these facts to show that recombination has the status of an interference with what is natural. But this desperate strategy conflicts with the received evolutionary view of the function of sexuality. The deploying of prodigious quantities of variability is not a dysfunction which sexual organisms are vulnerable to. Rather it is the principal advantage of sexuality; it is standardly construed to be *what sexuality is for* (but see Williams 1975 for a dissenting opinion). If the notion of a natural state is to make any sense at all, then variability must be viewed as the upshot of natural forces.

The natural state model is a *causal*, and thereby a *historical, hypothesis*. The essentialist attempts to understand variation within a species as arising through a process of deviation from type. By tracing back the origins of this variability we discover the natural state of a species. To do this is to uncover that natural tendency possessed by each member of the species. But the science which describes the laws governing the historical origins of variation within species—population genetics—makes no appeal to such ''natural tendencies.'' Rather, this frame invariant ''natural tendency''—this property that an organism is supposed to have regardless of what environment it might be in—has been replaced by a frame relative property—namely, the phenotype that a genotype will produce *in a given environment*. The historical concept of a natural state is discredited in much the same way that the kinematic concept of absolute simultaneity was.

Our current concepts of function and dysfunction, of disease and health, seem to be based on the kinds of distinctions recommended by the natural state model. And both of these distinctions resist characterization in terms of maximum fitness. For virtually any trait you please, there can be environments in which that trait is selected for, or selected against. Diseases can be rendered advantageous, and health can be made to represent a reproductive cost. And even if we restrict our attention to historically actual environments, we still encounter difficulties. A perfectly healthy phenotype may be historically nonexistent; the optimum actually attained might still be some diseased state.

The functional notions just mentioned make distinctions which are sanctioned by the natural state model. Given the inadequacy of this model, does this show that the difference between disease and health and the difference between function and dysfunction are mere illusions? I do not think that this follows. What we should conclude is that these functional notions of normality are not to be characterized in terms of a historical notion of fitness. Perhaps they can be understood in some other way; that remains to be seen.

In addition to the influence that the natural state model continues to exert in scientific thinking, perhaps even more pervasive is the way that notions of naturalness have had, and continue to have, an influence in politics and in popular culture. Political theorists of both the left and the right have appealed to something called ''human nature'' (Lewontin 1977, Hull 1978).[18] Political optimists see human nature as essentially good; the evil that human beings have done is to be chalked up to interferences on the part of civilization, or of the state, or of particular economic institutions. Pessimists, on the other hand, see in human beings a natural tendency toward evil, which the restraints made possible by civilization can perhaps correct. The common presupposition here is that each human being has a particular dispositional property—a natural tendency—whose manifestation is contingent on whether environmental forces facilitate the expression of what is natural, or, on the other hand, go against nature by imposing unnatural interferences.

A more recent manifestation of the same habit of mind is to be found in debates about ''environmental policy.'' Current environmental controversy, both on the part of those who want further industrialization to take its course and on the part of those who want to check or alter the way in which industry impinges on wildlife, tends to picture nature as something apart from us. The question before us, both sides imply, is how we should behave toward this separate sphere. We are not part of what is natural, and what we do has the character of an intervention from the outside into this natural domain. Our pollution of lakes, disruption of ecosystems, and extinction of species is just not natural. Natural, it would seem, is a good thing to be nowadays. Civilization is more often than not an interfering force, deflecting us from what is natural.

The Victorians, too, had their unnatural acts, thus hoping to find their ethics at least consistent with, and possibly vindicated by, the natural order. But they, at least, maintained some distance from the automatic equation of natural and good. Although some unnatural acts were wrong, others were decidedly right: here natural tendencies had to be checked if morally desirable qualities were to emerge. Perhaps it is a sign of our crumbling moral confidence that we no longer find it possible to separate questions of what is natural from what is good. By equating the two, we hope to read off our ethics directly from what happens in nature, and this gives us the illusion of needing to make no moral decisions for ourselves. This moral buck-passing is incoherent. What happens in nature is simply everything that happens. There is no other sense of "natural." Human society is not external to nature but a special part of it. It is no more a part of human nature to be healthy than to be diseased. Both kinds of phenotypes are to be found, and the norm of reaction makes no distinction between them. If we prefer one and wish to create environments in which it is encouraged, let us say so. But our reasons cannot be given in terms of allowing what is natural to occur unimpeded—by letting nature take its course, as if it has only one. Our activity, and inactivity, requires a more substantive justification than this.

Conclusion

Essentialism is as much entitled to appeal to the principle of tenacity as any other scientific hypothesis or guiding principle. It was hardly irrational for nineteenth-century research on the chemical elements to persist in its assumption that chemical kinds exist and have essential properties. The same holds true for those who hold that species are natural kinds and have essential properties; repeated failure to turn up the postulated items may be interpreted as simply showing that inquiry has not proceeded far enough. Matters change, however, when theoretical reasons start to emerge which cast doubt on the existence claim. For example, if the existence claim is shown to be theoretically superfluous, that counts as one reason for thinking that no such thing exists, or so the principle of parsimony would suggest (Sober 1980a). In another vein, if the causal mechanism associated with the postulated entity is cast in doubt, that too poses problems for the rationality of the existence claim. Our discussion of how population thinking emancipated biology from the need for constituent definitions of species is an argument of the first kind. Our examination of the theory of variation presupposed by essentialism is an argument of the second kind.

No phenotypic characteristic can be postulated as a species essence; the norm of reaction for each genotype shows that it is arbitrary to single out as privileged one phenotype as opposed to any other. Similar considerations show that no genotypic characteristic can be postulated as a species essence; the genetic variability found in sexual populations is prodigious and, again, there is no biologically plausible way to single out some genetic characteristics as natural while viewing others as the upshot of interfering forces. Even if a species were found in which some characteristic is shared by all and only the organisms that are in the species, this could not be counted as a species essence. Imagine, for example, that some novel form of life is created in the laboratory and subjected to some extreme form of stabilizing selection. If the number of organisms is kept small, it may turn out that the internal homogeneity of the species, as well as its distinctness from all other species, has been assured. However, the explanation of this phenomenon would be given in terms of the selection pressures acting on the population. If the universal property were a species essence, however, explaining why it is universal would be like explaining why all acids are proton donors, or why all bachelors are unmarried, or why all nitrogen has atomic number 14. These latter necessary truths, if they are explainable at all, are not explained by saying that some contingent causal force acted on acids, bachelors, or samples of nitrogen, thereby endowing them with the property in question. Characteristics possessed by all and only the extant members of a species, if such were to exist, would not be species essences. It is for this reason that hypotheses of discontinuous evolution like that proposed by Eldredge and Gould (1972) in no way confirm the claims of essentialism.

The essentialist hoped to penetrate the veil of variability found within species by discovering some natural tendency which each individual in the species possesses. This natural tendency was to be a dispositional property which would be manifest, were interfering forces not at work. Heterogeneity is thus the result of a departure from the natural state. But, with the development of evolutionary theory, it turned out that no such property was available to the essentialist, and in fact our current model of variability radically differs from the essentialist's causal hypothesis about the origins of variability.

At the same time that evolutionary theory undermined the essentialist's model of variability, it also removed the need for discovering species essences. Characteristics of populations do not have to be defined in terms of characteristics of organisms for population concepts to be coherent and fruitful. Population biology attempts to formulate generalizations about kinds of populations. In spite of the fact that species cannot be precisely individuated in terms of their constituent organisms, species undergo evolutionary processes, and the character of such processes is what population biology attempts to describe. Laws generalizing over population will, of course, include the standard *ceteris paribus* rider: they will describe how various properties and magnitudes are related, as long as no other forces affect the system. At least one such law describes what happens when *no* evolutionary force is at work in a panmictic Mendelian population. This is the Hardy–Weinberg equilibrium law. This law describes an essential property—a property which is necessary for a population to be Mendelian. But, of course, such laws do not pick out *species'* essences. Perhaps essentialism can reemerge as a thesis, not about species, but about *kinds* of species. The natural state model arguably finds an application at that level of organization in that the Hardy–Weinberg zero-force state is distinguished from other possible population configurations.

The transposition of Aristotle's distinction is significant. The essentialist searched for a property *of individual organisms* which is invariant across the organisms in a species. The Hardy–Weinberg law and other more interesting population laws, on the other hand, identify properties of *populations* which are invariant across all populations of a certain kind. In this sense, essentialism pursued an individualistic (organismic) methodology, which population thinking supplants by specifying laws governing objects at a higher level of organization.[19] From the individualistic (organismic) perspective assumed by essentialism, species are real only if they can be delimited in terms of membership conditions applying to individual organisms. But the populationist point of view made possible by evolutionary theory made such reductionistic demands unnecessary. Since populations and their properties are subject to their own invariances and have their own causal efficacy, it is no more reasonable to demand a species definition in terms of the properties of constituent organisms than it is to require organismic biology to postpone its inquiries until a criterion for sameness of organism is formulated in terms of relations between constituent cells. Essentialism lost its grip when populations came to be thought of as real.[20] And the mark of this latter transformation in thought was the transposition of the search for invariances to a higher level of organization.[21]

Notes

Suggestions made by William Coleman, James Crow, Joan Kung, David Hull, Geoffrey Joseph, Steven Kimbrough, Richard Lewontin, Ernst Mayr, Terrence Penner, William Provine, Robert Stauffer, Dennis Stampe, and Victor Hilts helped me considerably in writing this paper.

1. Mayr (1963) has argued additionally that essentialist errors continue to be made in population biology in the form of the distortions of ''bean-bag genetics.'' The assumption that the fitness of single genes is independent of their genetic context is and has been known to be mistaken; but how this simplifying assumption is essentialist in character is obscure to me.

2. If species are *individuals*—spatiotemporally extended lineages—as Ghiselin (1966, 1969, 1974), and Hull (1976, 1978) have argued, then we have our assurance of finitude. If, on the other hand, species are kinds of things, which may in principle be found anywhere in the universe at any time, then a slightly different argument is needed for the claim that the same species is overwhelmingly unlikely to have evolved twice. Such an argument is provided by considering the way in which speciation depends on the coincidence of a huge number of initial conditions. See Ayala (1978) for a summary of the received view of this matter.

3. I would suggest that quantum mechanical considerations show that the concept of being a nucleus with a particular atomic number is a vague one. Presumably, a collection of protons constitutes a nucleus when the strong force which causes them to attract each other overcomes their mutual electromagnetic repulsion. Whether this happens or not is a function of the distances between the protons. But *this* concept—that of ''the'' distance between particles—is indeterminate. Hence, the question of whether something is or is not a nucleus with a particular atomic number can only be answered probabilistically.

4. It is probably a mistake to talk about concepts being vague *simpliciter*. Rather, one should formulate matters in terms of concepts being vague relative to particular application. The issue of

whether a concept is vague seems to reduce to the issue of whether there are cases in which it is indeterminate whether the concept applies. I would guess that practically every concept applying to physical objects is vague in this sense. Thus, even such concepts as "being two in number" are such that circumstances can be described in which it is indeterminate whether they apply to the objects in question. Degrees of vagueness can be partially defined as follows: If the set of circumstances in which concept *P* is indeterminate in its application is properly included in the set of circumstances in which concept *Q* is indeterminate, then *Q* is more vague than *P*.

5. Thus in his (1859), p. 52, Darwin says: "From these remarks it will be seen that I look at the term species, as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other, and that it does not essentially differ from the term variety, which is given to less distinct and more fluctuating forms. The term variety, again, in comparson with mere individual differences, is also applied arbitrarily, and for mere convenience sake." Elsewhere in (1859, e.g., pp. 432–33), Darwin espouses his perhaps more dominant populationist view that, in spite of line-drawing problems, species are real.

6. I am not arguing that Hull (1965) and others have misidentified the essence of essentialism and that their criticisms thereby fail to get to the heart of the matter. Essentialism, like most isms which evolve historically, probably does not even have an essence. Rather, I am trying to construe essentialism as a fairly flexible doctrine which, in at least some circumstances, can be seen to be quite consistent with the existence of insoluble line-drawing problems.

7. This characterization of Aristotle's view in terms of some information bearing entity is not completely anachronistic, as Delbrück (1971) points out when he (in jest) suggests that Aristotle should receive a Nobel Prize for having discovered DNA.

8. In this discussion of Aristotle's view of *terrata*, I have been much helped by Furth's (1975, section 11).

9. If one views Aristotle as excluding monstrous forms from membership on any species category, then one will have an extreme instance of this ad hoc strategy; *no* organism will belong to any species. Hull (1973, 39–40) sees Aristotle and scholastic science as hopelessly committed to this futile strategy. However, on the view I would attribute to Aristotle, most, if not all, monstrous forms are members of the species from which they arose. They, like Newtonian particles which fail to be at rest or in uniform motion, fail to achieve their natural states because of identifiable causal forces.

10. Hilts (1973, 209–210). My discussion of Quetelet and Galton in what follows leans heavily on Hilts (1973). It has a number of points in common with Hacking's (1975).

11. Boring (1929, 477) brings out the Aristotelian teleology contained in Quetelet's ideas quite well when he characterizes Quetelet as holding that "we might regard such human variation as if it occurred when nature aimed at an ideal and missed by varying amounts."

12. Although Galton found *The Origin of Species* an encouragement to pursue his own ideas, he indicates that his interest in variation and inheritance were of long standing. See Hilts (1973, 220).

13. In his *Hereditary Genius*, Galton compared the development of species with a many-faceted spheroid tumbling over from one facet or stable equilibrium to another. See Provine (1971, 14–15). This saltative process ensured unity of type. In spite of Galton's adherence to the idea of discontinuous evolution and certain other essentialist predilections (Lewontin 1974, 4), his innovations in population thinking were anti-essentialist in their consequences, or so I will argue.

14. Hilts (1973, 228). Walker (1929, 185) claims that the origin of the name "normal curve" is obscure. It occurs in Lexis and, she says, "It is not improbable that the term goes back to Quetelet." As natural and inevitable as Quetelet found his interpretation of the bell curve in terms of the Natural State Model, by the time Galton's *Natural Inheritance* appeared in 1889, there was growing sentiment that this interpretation was acceptable, if at all, only as a special case. Thus we find Galton, in that work (p. 58), saying that "the term Probable Error is absurd when applied to the subjects now in hand, such as Stature, Eye-colour, Artistic Faculty, or Disease." A year earlier, Venn, in his *The Logic of Chance* (p. 42), made a similar comment: "When we perform an operation ourselves with a clear consciousness of what we are aiming at, we may quite correctly speak of every deviation from this as being an error; but when Nature presents us with a group of objects of every kind, it is using a rather bold metaphor to speak in this case also of a law of error, as if she had been aiming at something all the time, and had like the rest of us missed her mark or less in every instance." Quotations are drawn from Walker (1929, 53).

15. It would be important to trace the development of statistical ideas from Galton through Pearson and his circle to R. A. Fisher, and to see whether Pearson's positivistic convictions had the effect of further proscribing the idea of types on the grounds that it is "unscientific." Cohen (1972) sees Galton as already adopting some positivistic attitudes in his idea that heredity was to be understood in terms of correlations, and not in terms of causal forces. Also, see Hacking (1975) for a bold attempt to link Galton's innovations to other developments in nineteenth-century thought. I should point out that a fuller treatment of the emergence of population thinking would have to ascribe a central role to Mendel. He, much more than Galton, provided the central elements of our present conception of the relation of heredity and variation. I have stressed Galton, however, because of his interpretation of statistics and because of his view of the population as a unit of explanation.

16. The discussion of the norm of reaction in what follows depends heavily on some points made in Lewontin (1977).

17. This selectionist suggestion needs to be made more precise by specifying the notion of fitness used. I will not lay out these different conceptions here. Rather, I invite the reader to choose the one that he or she finds most plausible. The upshot of my argument does not seem to depend on which biologically plausible characterization is chosen.

18. Lewontin (1977, 11) has argued that the idea of a "natural phenotype" has been used in some hereditarian thinking in the IQ controversy. He quotes Herrnstein (1971, 54) as talking about "artificially boosting" an individual's IQ score. The presupposition seems to be that each human genotype has associated with it an IQ score (or range of such scores) which counts as its natural phenotype. As in Aristotle, the individual can be deflected from what is natural by environmental interference.

19. It is significant that biologists to this day tend to use "individual" and "organism" interchangeably. For arguments that populations, and even species, are to be construed as individuals, see Ghiselin (1966, 1969, 1974), and Hull (1976, 1978).

20. I borrow this way of putting matters from Hacking (1975) in which he describes the series of transformations in thought which resulted in "chance becoming real."

21. The group selection controversy provides an interesting example of the question of whether, and in what respects, it is appropriate to view populations as objects. In some ways, this debate recapitulates elements of the dispute between methodological holism and methodological individualism in the social sciences. See Sober (1980c) for details.

References

Agassiz, L. (1859). *Essay on Classification*. Cambridge, Mass.: Harvard University Press.

Ayala, F. (1978). "The Mechanisms of Evolution." *Scientific American* 239, 3: 56–69.

Balme, D. (1962). "Development of Biology in Aristotle and Theophrastus: Theory of Spontaneous Generation." *Phronesis* 2, 1: 91–104.

Boring, E. (1929). *A History of Experimental Psychology*. New York: Appleton-Century-Crofts.

Buffon, L. (1749). *Histoire Naturelle*. Paris.

Cohen, R. (1972). "Francis Galton's Contribution to Genetics." *Journal of the History of Biology* 5, 2: 389–412.

Darwin, C. (1859). *On the Origin of Species*. Cambridge, Mass.: Harvard University Press.

———. (1868). *The Variation of Animals and Plants under Domestication*. London: Murray.

Delbrück, M. (1971). "Aristotle-totle-totle." In Monod, J., and Borek, J. (eds), *Microbes and Life* 50–55. New York: Columbia University Press.

Eldredge, N., and Gould, S. (1972). "Punctuated Equilibria: An alternative to Phyletic/Gradualism." In T. Schopf (ed.), *Models in Paleobiology*, 82–115. San Francisco: Freeman Cooper.

Furth, M. (1975). *Essence and Individual: Reconstruction of an Aristotelian Metaphysics*, chapter 11, duplicated for the meeting of the Society for Ancient Greek Philosophy, unpublished.

Ghiselin, M. (1966). "On Psychologism in the Logic of Taxonomic Controversies." *Systematic Zoology* 15: 207–15.

———. (1969). *The Triumph of the Darwinian Method*. Berkeley: University of California Press.

———. (1974). "A Radical Solution to the Species Problem." *Systematic Zoology* 23: 536–44.

Glass, B. (1959a). "Heredity and Variation in the eighteenth Century Concept of the Species." In Glass, B., et al. (eds.), *Forerunners of Darwin*, 144–72. Baltimore: The Johns Hopkins Press.

———. (1959b). "Maupertuis, Pioneer of Genetics and Evolution." in Glass, B., et al. (eds.), *Forerunners of Darwin*, 51–83. Baltimore: The Johns Hopkins Press.

Hacking, I. (1975). "The Autonomy of Statistical Law." Talk delivered to American Philosophical Association, Pacific Division, unpublished.

Herrnstein, R. (1971). "IQ." *Atlantic Monthly* 228, 3: 43–64.

Hilts, V. (1973). "Statistics and Social Science." In Giere, R. and Westfall, R. (eds.), *Foundations of Scientific Method in the Nineteenth Century*, 206–33. Bloomington: Indiana University Press.

Hull, D. (1965). "The Effect of Essentialism on Taxonomy: 2000 Years of Stasis." *British Journal for the Philosophy of Science* 15: 314–16; 16: 1–18.

———. (1967). "The Metaphysics of Evolution." *British Journal for the History of Science* 3, 12: 309–37.

———. (1968). "The Conflict between Spontaneous Generation and Aristotle's Metaphysics." *Proceedings of the Seventh Inter-American Congress of Philosophy* 2 (1968): 245–50. Quebec City: Les Presses de l'Université Laval.

———. (1973). *Darwin and His Critics*. Cambridge, Mass.: Harvard University Press.

———. (1976). "Are Species Really Individuals?" *Systematic Zoology* 25: 174–91.

———. (1978). "A Matter of Individuality." *Philosophy of Science* 45: 335–60.

Ihde, A. (1964). *The Development of Modern Chemistry*. New York: Harper & Row.

Kripke, S. (1972). "Naming and Necessity." In Davidson, D., and Harman, G. (eds.), *Semantics of Natural Languages*, 253–355; 763–9. Dordrecht: Reidel.

———. (1978). "Time and Identity." Lectures given at Cornell University, unpublished.

Lewontin, R. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

———. (1977). "Biological Determinism as a Social Weapon." In Ann Arbor Science for the People Editorial Collective, *Biology as a Social Weapon*, 6–20. Minneapolis: Burgess.

Lloyd, G. (1968). *Aristotle: The Growth and Structure of His Thought*. Cambridge: Cambridge University Press.

Lovejoy, A. (1936). *The Great Chain of Being*. Cambridge, Mass.: Harvard University Press.

Mayr, E. (1959). "Typological versus Population Thinking." In *Evolution and Anthropology: A Centennial Appraisal*, 409–12. Washington: Anthropological Society of Washington.

———. (1963). *Animal Species and Evolution*. Cambridge, Mass.: Belknap Press of Harvard University Press.

———. (1969). "The Biological Meaning of Species." *Biology Journal of the Linnean Society* 1: 311–20.

———. (1976). *Evolution and the Diversity of Life*. Cambridge, Mass.: Harvard University Press.

Meyer, A. (1939). *The Rise of Embryology*. Stanford, Calif.: Stanford University Press.

Minkowski, H. (1908). "Space and Time." In Lorentz, H., Einstein, A., et al., *The Principle of Relativity* 73–91. New York: Dover.

Popper, K. (1972). *Objective Knowledge*. Oxford: Oxford University Press.

Preuss, A. (1975). *Science and Philosophy in Aristotle's Biological Works*. New York: Georg Olms.

Provine, W. (1971). *The Origins of Theoretical Population Genetics*. Chicago: University of Chicago Press.

Putnam, H. (1975). "The Meaning of 'Meaning,'" *In Mind, Language and Reality*, 215–71. Cambridge: Cambridge University Press.

Quetelet, A. (1842). *A Treatise on Man and the Development of His Faculties*. Edinburgh.

Quine, W. V. O. (1953a). "Identity, Ostension, Hypostasis." In *From a Logical Point of View*, 65–79. New York: Harper Torchbooks.

———. (1953b). "Reference and Modality." In *From a Logical Point of View*, 139–59. New York: Harper Torchbooks.

———. (1960). *Word and Object*. Cambridge, Mass.: MIT Press.

Rabel, G. (1939). "Long before Darwin: Linne's Views on the Origin of Species," *Discovery* n.s., 2: 121–75.

Ramsbottom, J. (1938). "Linnaeus and the Species Concept." *Proceedings of the Linnean Society of London*: 192–219.

Roger, J. (1963). *Les Sciences de la vie dans la pensée française du XVIII siècle*. Paris: Armand Colin.

Sober, E. (1980a). "The Principle of Parsimony." *British Journal for Philosophy of Science*.

———. (1980b). "Realism and Independence." *Noûs*. 41: 369–386.

———. (1980c). "Holism, Individualism, and the Units of Selection." *Proceedings of the Biennial Meeting of the Philosophy of Science Association*: 93–101. E. Lansing, Michigan. Philosophy of Science Association.

Todhunter, I. (1865). *History of the Theory of Probability to the Time of Laplace*. New York: Chelsea Publishing.

Walker, H. (1929). *Studies in the History of Statistical Method*. Baltimore: Williams & Wilkins.

Williams, G. C. (1975). *Sex and Evolution*. Princeton, N.J.: Princeton University Press.

Wilson, E., and Bossert, W. (1971). *A Primer of Population Biology*. Sunderland, Mass.: Sinauer.

IX Species

18 A Matter of Individuality

David L. Hull

Biological species have been treated traditionally as spatiotemporally unrestricted classes. If they are to perform the function which they do in the evolutionary process, they must be spatiotemporally localized individuals, historical entities. Reinterpreting biological species as historical entities solves several important anomalies in biology, in philosophy of biology, and within philosophy itself. It also has important implications for any attempt to present an "evolutionary" analysis of science and for sciences such as anthropology which are devoted to the study of single species.

Introduction

The terms "gene," "organism," and "species" have been used in a wide variety of ways in a wide variety of contexts. Anyone who attempts merely to map this diversity is presented with a massive and probably pointless task. In this chapter I consciously ignore "the ordinary uses" of these terms, whatever they might be, and concentrate on their biological uses. Even within biology the variation and conflicts in meaning are sufficiently extensive to immobilize all but the most ambitious ordinary language philosopher. Thus I have narrowed my focus even further to concentrate on the role which these terms play in evolutionary biology. In doing so, I do not mean to imply that this usage is primary or that all other biological uses which conflict with it are mistaken. Possibly evolutionary theory is *the* fundamental theory in biology, and all other biological theories must be brought into accord with it. Possibly all biological theories, including evolutionary theory, eventually will be reduced to physics and chemistry. But regardless of the answers to these global questions, at the very least various versions of evolutionary theory are sufficiently important in biology to warrant an investigation of the implications which they have for the biological entities which they concern.

From *Philosophy of Science* 45, no. 3 (September 1978): 335–360. © 1978 by the Philosophy of Science Association. All rights reserved. Reprinted by permission of the University of Chicago Press.

Genes are the entities which are passed on in reproduction and which control the ontogenetic development of the organism. Organisms are the complex systems which anatomists, physiologists, embryologists, histologists, and others analyze into their component parts. Species have been treated traditionally as the basic units of classification, the natural kinds of the living world, comparable to the physical elements. But these entities also function in the evolutionary process. Evolution consists in two processes (mutation and selection) which eventuate in a third (evolution). Genes provide the heritable variation required by the evolutionary process. Traditionally organisms have been viewed as the primary focus of selection, although considerable disagreement currently exists over the levels at which selection takes place. Some biologists maintain that selection occurs exclusively at the level of genes; others that supragenic, even supraorganismic units can also be selected. As one might gather from the title of Darwin's book, species are the things which are supposed to evolve. Whether the relatively large units recognized by taxonomists as species evolve or whether much less extensive units such as populations are the effective units of evolution is an open question. In this chapter when I use the term "species," I intend to refer to those supraorganismic entities which evolve regardless of how extensive they might turn out to be.

The purpose of this chapter is to explore the implications which evolutionary theory has for the ontological status of genes, organisms, and species. The only category distinction I discuss is between individuals and classes. By "individuals" I mean spatiotemporally localized cohesive and continuous entities (historical entities.) By "classes" I intend spatiotemporal unrestricted classes, the sorts of things which can function in traditionally defined laws of nature. The contrast is between Mars and planets, the Weald and geological strata, Gargantua and organisms. The terms used to mark this distinction are not important; the distinction is. For example, one might distinguish two sorts of sets: those that are defined in terms of a spatiotemporal relation to a spatiotemporally localized focus, and those that are not. On this view, historical entities such as Gargantua become sets. But they are sets of a very special kind—sets defined in terms of a spatiotemporal relation to a spatiotemporally localized focus. Gargantua, for instance, would be the set of all cells descended from the zygote which gave rise to Gargantua.

The reason for distinguishing between historical entities and genuine classes is the differing roles which each plays in science according to traditional analyses of scientific laws. Scientific laws are supposed to be spatiotemporally unrestricted generalizations. No uneliminable reference can be made in a genuine law of nature to a spatiotemporally individuated entity. To be sure, the distinction between accidentally true generalizations (such as all terrestrial organisms using the same genetic code) and genuine laws of nature (such as those enshrined in contemporary versions of celestial mechanics) is not easy to make. Nor are matters helped much by the tremendous emphasis placed on laws in traditional philosophies of science, as if they were the be-all and end-all of science. Nevertheless, I find the distinction between those generalizations that are spatiotemporally unrestricted and those that are not fundamental to our current understanding of science. Whether one calls the former "laws" and the latter something else, or whether one terms both sorts of statements "laws" is of little consequence. The point I wish to argue is that genes, organisms, *and* species, as they function in the evolutionary process, are necessarily spatiotemporally localized individuals. They could not perform the functions which they perform if they were not.

The argument presented in this chapter is metaphysical, not epistemological. Epistemologically red light may be fundamentally different from infrared light and mammals from amoebae. Most human beings can see with red light and not infrared light. Most people can see mammals; few if any can see amoebae with the naked eye. Metaphysically they are no different. Scientists know as much about one as the other. Given our relative size, period of duration, and perceptual acuity, organisms appear to be historical entities, species appear to be classes of some sort, and genes cannot be seen at all. However, after acquainting oneself with the various entities which biologists count as organisms and the roles which organisms and species play in the evolutionary process, one realizes exactly how problematic our commonsense notions actually are. The distinction between an organism and a colony is not sharp. If an organism is the "total product of the development of the impregnated embryo," then as far back as 1899 T. H. Huxley was forced to conclude that the medusae set free from a hydrozoan "are as much organs of the latter as the multitudinous pinnules of a *Comatula*, with their genital glands, are organs of the Echinoderm. Morphologically, therefore, the equivalent of the individual *Comatula* is the Hydrozoic stock and all the Medusae which proceed from it" (24). More recently, Daniel Janzen (25) has remarked that the "study of dandelion ecology and evolution suffers from confusion of the layman's 'individual' with the 'individual' of evolutionary biology. The latter individual has 'reproductive fitness' and is the unit of selection in most evolutionary conceptualizations" (see also 2). According to evolutionists, units of selection, whether they be single genes, chromosomes, organisms, colonies, or kinship groups, are individuals. In this chapter I intend to extend this analysis to units of evolution.

If the ontological status of space-time in relativity theory is philosophically interesting in and of itself (and God knows enough philosophers have written on that topic), then the ontological status of species in evolutionary theory should also be sufficiently interesting philosophically to discuss without any additional justification. However, additional justification does exist. From Socrates and Plato to Kripke and Putnam, organisms have been paradigm examples of primary substances, particulars, and/or individuals, while species have served as paradigm examples of secondary substances, universals, and/or classes. I do not think that this chapter has any necessary implications for various solutions to the problem of universals, identity, and the like. However, if my main contention is correct, if species are as much spatiotemporally localized individuals as organisms, then some of the confusion among philosophers over these issues is understandable. One of the commonest examples used in the philosophical literature is inappropriate. Regardless of whether one thinks that ''Moses'' is a proper name, a cluster concept, or a rigid designator, ''*Homo sapiens*'' must be treated in the same way.

The Evolutionary Justification

Beginning with the highly original work of Michael Ghiselin (12, 13, 14), biologists in increasing numbers are beginning to argue that species as units of evolution are historical entities (15, 20, 21, 22, 23, 34, 38). The justification for such claims would be easier if there were one set of propositions (presented preferably in axiomatic form) which could be termed *the* theory of evolution. Unfortunately, there is not. Instead there are several, incomplete, partially incompatible versions of evolutionary theory currently extant. I do not take this state of affairs to be unusual, especially in periods of rapid theoretical change. In general the myth that some one set of propositions exists which can be designated unequivocally as Newtonian theory, relativity theory, etc., is an artifact introduced by lack of attention to historical development and unconcern with the primary literature of science. The only place one can find *the* version of a theory is in a textbook written long after the theory has ceased being of any theoretical interest to scientists.

In this section I set out what it is about the evolutionary process which results in species being historical entities, not spatiotemporally unrestricted classes. In doing so I have not attempted to paper over the disagreements which currently divide biologists working on evolutionary theory. For example, some disagreement exists over how abruptly evolution can occur. Some biologists have argued that evolution takes place saltatively, in relatively large steps. Extreme saltationists once claimed that in the space of a single generation new species can arise which are so different from all other species that they have to be placed in new genera, families, classes, etc. No contemporary biologist to my knowledge currently holds this view. Extreme gradualists, on the other side, argue that speciation *always* occurs very slowly, over periods of hundreds of generations, either by means of a single species changing into a new species (phyletic evolution) or else by splitting into two large subgroups which gradually diverge (speciation). No contemporary biologist holds this view either. Even the most enthusiastic gradualists admit that new species can arise in a single generation, e.g., by means of polyploidy. In addition, Eldredge and Gould (11), building on Mayr's founder principle (36, 37), have recently argued that speciation typically involves small, peripheral isolates which develop quite rapidly into new species. Speciation is a process of ''punctuated equilibria.''

However, the major dispute among contemporary evolutionary theorists is the level (or levels) at which selection operates. Does election occur *only* and *literally* at the level of genes? Does selection take place *exclusively* at the level of organisms, the selection of genes being only a consequence of the selection of organisms? Can selection also take place at levels of organization more inclusive than the individual organism, e.g., at the level of kinship groups, populations, and possibly even entire species? Biologists can be found opting for every single permutation of the answers to the preceding questions. I do not propose to go through all the arguments which are presented to support these various conclusions. For my purposes it is sufficient to show that the points of dispute are precisely those which one might expect if species are being interpreted as historical entities, rather than as spatiotemporally unrestricted classes. Richard Dawkins puts the crucial issue as follows:

> Natural selection in its most general form means the differential survival of entities. Some entities live and others die but, in order for this selective death to have any impact on the world, an additional condition must be met. Each entity must exist in the form of lots of copies, and at least some of the entities must be *potentially* capable of surviving—in the form of copies—for a significant period of evolutionary time.

The results of evolution by natural selection are *copies* of the entities being selected, not *sets*. Elements in a set must be characterized by one or more common characteristics. Even fuzzy sets must be characterized by at least a "cluster" of traits. Copies need not be.[1] A particular gene is a spatiotemporally localized individual which either may or may not replicate itself. In replication the DNA molecule splits down the middle producing two new molecules composed *physically* of half of the parent molecule while *largely* retaining its structure. In this way genes form lineages, ancestor-descendant copies of some original molecule. The relevant genetic units in evolution are not *sets* of genes defined in terms of structural similarity but lineages formed by the imperfect copying process of replications.[2] Genes can belong to the same lineage even though they are structurally different from other genes in that lineage. What is more, continued changes in structure can take place indefinitely. If evolution is to occur, not only *can* such indefinite structural variation take place within gene lineages, but it *must*. Single genes are historical entities, existing for short periods of time. The more important notion is that of a *gene lineage*. Gene lineages are also historical entities persisting while changing indefinitely through time. As Dawkins puts this point:

> Genes, like diamonds, are forever, but not quite in the same way as diamonds. It is an individual diamond crystal which lasts, as an unaltered pattern of atoms. DNA molecules don't have that kind of permanence. The life of any one physical DNA molecule is quite short—perhaps a matter of months, certainly not more than one lifetime. But a DNA molecule could theoretically live on in the form of *copies* of itself for a hundred million years. (8, p. 36)

Exactly the same observations can be made with respect to organisms. A particular organism is a spatiotemporally localized individual which either may or may not reproduce itself. In asexual reproduction, part of the parent organism buds off to produce new individuals. The division can be reasonably equitable, as in binary fission, or extremely inequitable, as in various forms of parthenogenesis. In sexual reproduction gametes are produced which unite to form new individuals. Like genes, organisms form lineages. The relevant organismal units in evolution are not sets of organisms defined in terms of structural similarity but lineages formed by the imperfect copying processes of reproduction. Organisms can belong to the same lineage even though they are structurally different from other organisms in that lineage. What is more, continued changes in structure can take place indefinitely. If evolution is to occur, not only *can* such indefinite structural variation take place within organism lineages, but it *must*. Single organisms are historical entities, existing for short periods of time. Organism lineages are also historical entities persisting while changing indefinitely through time.

Both replication and reproduction are spatiotemporally localized processes. There is no replication or reproduction at a distance. Spatiotemporal continuity through time is required. Which entities at which levels of organization are sufficiently cohesive to function as units of selection is more problematic. Dawkins presents one view:

> In sexually reproducing species, the individual [the organism] is too large and too temporary a genetic unit to qualify as a significant unit of natural selection. The group of individuals is an even larger unit. Genetically speaking, individuals and groups are like clouds in the sky or dust-storms in the desert. They are temporary aggregates of federations. They are not stable through evolutionary time. Populations may last a long while, but they are constantly blending with other populations and so losing their identity. They are subject to evolutionary change from within. A population is not a discrete enough entity to be a unit of natural selection, not stable and unitary enough to be "selected" in preference to another population. (8, p. 37)

From a commonsense perspective, organisms are paradigms of tightly organized, hierarchically stratified systems. Kinship groups such as hives also seem to be internally cohesive entities. Populations and species are not. Dawkins argues that neither organisms (in sexually reproducing species) nor populations in any species are sufficiently permanent and cohesive to function as units in selection. In asexual species, organisms do not differ all that much from genes. They subdivide in much the same way that genes do, resulting in progeny which are identical (or nearly identical) with them. In sexual species, however, organisms must pool their genes to reproduce. The resulting progeny contain a combined sample of parental genes, Populations lack even this much cohesion.

Other biologists are willing to countenance selection at levels more inclusive than the individual gene, possibly parts of chromosomes, whole chromosomes, entire organisms, or even kinship groups (32). The issues, both empirical and conceptual, are not simple. For example, G. C. Williams in his classic work (61) argues that selection occurs only at the level of individuals. By "individual," biologists usually mean "organism." However, when Williams is forced to admit that kinship groups can also function as units of selection, he promptly dubs them "individuals." One of the commonest objections to E. O. Wilson's (62) equally classic discussion of evolution is that he treats kin selection as a special case of group selection. According to the group selectionists, entities more inclusive than kinship groups can also function as units of selection (63).[3] Matters are not improved much by vagueness over what is meant by "units of selection." Gene frequencies are certainly altered from generation to generation, but so are genotype frequencies. Genes cannot be selected in isolation. They depend on the success of the organism which contains them for survival. Most biologists admit that similar observations hold for certain kinship groups. Few are willing to extend this line of reasoning to include populations and entire species.

Although the dispute over the level(s) at which selection takes place is inconclusive, the points at issue are instructive. In arguing that neither organisms nor populations function as units of selection in the same sense that genes do, Dawkins does not complain that the cells in an organism or the organisms in a population are phenotypically quite diverse, though they frequently are. Rather he denigrates their cohesiveness and continuity through time, criteria which are relevant to individuating historical entities, not spatiotemporally unrestricted classes. Difficulties about the level(s) at which selection can operate to one side, the issue with which we are concerned is the ontological status of species. Even if entire species are not sufficiently well integrated to function as units of selection, they are the entities which evolve as a result of selection at lower levels. The requirements of selection at these lower levels place constraints on the manner in which species can be conceptualized. Species as the results of selection are necessarily lineages, not sets of similar lineages, not sets of similar organisms. In order for differences in gene frequencies to build up in populations, continuity through time must be maintained. To some extent genes in sexual species are reassorted each generation, but the organisms which make up populations cannot be. To put the point in the opposite way, if such shuffling of organisms were to take place, selection would be impossible.

The preceding characteristic of species as evolutionary lineages by itself is sufficient to preclude species being conceptualized as spatiotemporally unrestricted sets or classes. However, if Eldredge and Gould are right, the case for interpreting species as historical entities is even stronger. They ask why species are so coherent, why groups of relatively independent local populations continue to display fairly consistent, recognizable phenotypes, and why reproductive isolation does not arise in every local population if gene flow is the only means of preventing differentiation:

> The answer probably lies in a view of species and individuals [organisms] as homeostatic systems—as amazingly well-buffered to resist change and maintain stability in the face of disturbing influences.... In this view, the importance of peripheral isolates lies in their small size and the alien environment beyond the species border that they inhabit—for only here are selective pressures strong enough and the inertia of large numbers sufficiently reduced to produce the "genetic revolution" (Mayr, 1963, p. 533) that overcomes homeostasis. The coherence of a species, therefore, is not maintained by interaction among its members (gene flow). It emerges, rather, as an historical consequence of the species' origin as a peripherally isolated population that acquired its own powerful homeostatic system. (11, p. 114)

Eldredge and Gould argue that, from a theoretical point of view, species appear so amorphous because of a combination of the gradualistic interpretation of speciation and the belief that gene exchange is the chief (or only) mechanism by which cohesion is maintained in natural populations. However, in the field, species of both sexual and asexual organisms seem amazingly coherent and unitary. If gene flow were the only mechanism for the maintenance of evolutionary unity, asexual species should be as diffuse as duststorms in the desert. According to Eldredge and Gould, new species arise through the budding off of peripheral isolates which succeed in establishing new equilibria in novel environments. Thereafter they remain largely unchanged during the course of their existence and survive only as long as they maintain this equilibrium.

Another possibility is that evolutionary unity is maintained by both internal and external means. Gene flow and homeostasis within a species are internal mechanisms of evolutionary unity. Perhaps the external environment in the form of unitary selection pressures also contributes to the integrity of the entities which are evolving (10). For example, Jews have remained relatively distinct from the rest of humankind for centuries, in part by internal means (selective mating, social customs, etc.) but also in part by external means (discrimination, prejudice, laws, etc.). An ecological niche is a relation between a particular species and key environmental variables. A different species in conjunction with the same environmental variables could define quite a different niche. In the past biologists have tended to play down the integrating effect of the environment, attributing whatever unity and coherence which exists in nature to the integrating effect of gene complexes. At the very least, if the coherence of asexual species is not illusory, mechanisms other than gene flow must be capable of bringing about evolutionary unity.

Individuating Organisms and Species

By and large, the criteria which biologists use to individuate organisms are the same as those suggested by philosophers—spatiotemporal continuity, unity, and location. Differences between these two analyses have three sources: first, philosophers have been most interested in individuating persons, the hardest case of all, while biologists have been content to individuate organisms; second, when philosophers have discussed the individuation of organisms, they have usually limited themselves to adult mammals, while biologists have attempted to develop a notion of organism adequate to handle the wide variety of organisms which exist in nature; and finally, philosophers have felt free to resort to hypothetical, science fiction examples to test their conceptions, while biologists rely on actual cases. In each instance, I prefer the biologists' strategy. A clear notion of an individual organism seems an absolute prerequisite for any adequate notion of a person, and this notion should be applicable to all organisms, not just a minuscule fraction. But most important, real examples tend to be much more detailed and bizarre than those made up by philosophers. Too often the example is constructed for the sole purpose of supporting the preconceived intuitions of the philosophers and has no life of its own. It cannot force the philosopher to improve his analysis the way that real examples can. Biologists are in the fortunate position of being able to test their analyses against a large stock of extremely difficult, extensively documented actual cases.

Phenotypic similarity is irrelevant in the individuation of organisms. Identical twins do not become one organism simply because they are phenotypically indistinguishable. Conversely, an organism can undergo massive phenotypic change while remaining the same organism. The stages in the life cycles of various species of organisms frequently are so different that biologists have placed them in different species, genera, families, and even classes—until the continuity of the organism was discovered. If a caterpillar develops into a butterfly, these apparently different organisms are stages in the life cycle of a single organism regardless of how dissimilar they might happen to be (figure 18.1a). In ontogenetic development, a single lineage is never divided successively in time into separate organisms; some sort of splitting is required. In certain cases, such as transverse fission in paramecia, a single organism splits equally into two

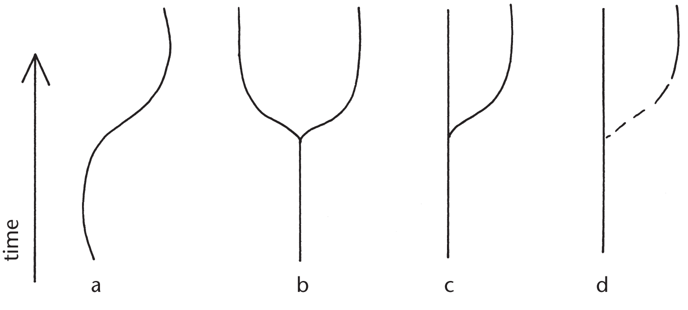


Figure 18.1

Diagrams which can be interpreted alternately as organisms undergoing ontogenetic change and the production of new organisms and as species undergoing phylogenetic change and speciation.

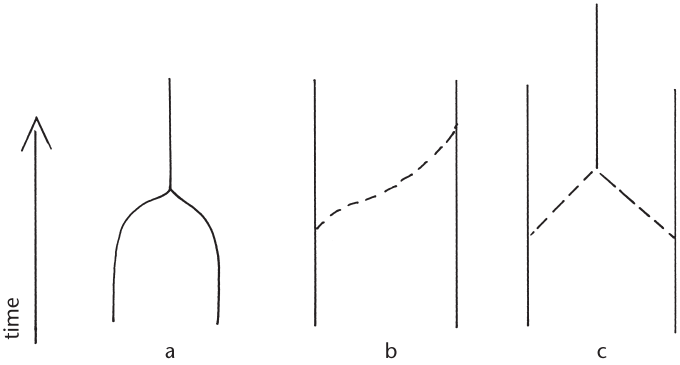


Figure 18.2
Diagrams which can be interpreted alternately as organisms merging totally or partially to give rise to new organisms and as species merging totally or partially to give rise to new species.

new organisms (figure 18.1b). In such cases, the parent organism no longer exists, and the daughter organisms are two new individuals. Sometimes a single individual will bud off other individuals which are roughly its own size but somewhat different in appearance, for example, strobilization in certain forms of Scyphozoa (figure 18.1c). At the other extreme, sometimes a small portion of the parent organism buds off to form a new individual, as in budding in Hydrozoa (figure 18.1d). In the latter two cases, the parent organism continues to exist while budding off new individuals. The relevant consideration is how much of the parent organism is lost and its internal organization disrupted.

Fusion also takes place at the level of individual organisms. For example, when presented with a prey too large for a single individual to digest, two amoebae will fuse cytoplasmically in order to engulf and digest it. However, the nuclei remain distinct and the two organisms later separate, genetically unchanged. The commonest example of true fusion occurs when germ cells unite to form a zygote. In such cases, the germ cells as individuals cease to exist and are replaced by a new individual (figure 18.2a). Sometimes one organism will invade another and become part of it. Initially, these organisms, even when they become obligate parasites, are conceived of as separate organisms, but sometimes they can become genuine parts of the host organism. For example, one theory of the origin of certain cell organelles is that they began as parasites. Blood transfusions are an unproblematic case of part of one organism's becoming part of another; conjugation is another (figure 18.2b). Sometimes parts of two different organisms can merge to form a third. Again, sexual reproduction is the commonest example of such an occurrence (figure 18.2c). In each of these cases, organisms are individuated on the basis of the amount of material involved and the effect of the change on the internal organization of the organisms. For example, after conjugation two paramecia are still two organisms and the same two organisms even though they have exchanged some of their genetic material.

If species are historical entities, then the same sorts of considerations which apply in the individuation of organisms should also apply to them, and they do (35). The only apparent discrepancy results from the fact that not all biologists have been totally successful in throwing off the old preevolutionary view of species as classes of similar organisms and replacing it with a truly evolutionary view. However, even these discrepancies are extremely instructive. For example, G. G. Simpson (50) maintains that a single lineage which changes extensively through time without speciating (splitting) should be divided into separate species (see figure 18.1a). Willi Hennig (17) disagrees: new species should be recognized only upon splitting. This particular debate has been involved, touching upon both conceptual and empirical issues. For example, how can a gradually evolving lineage be divided into discrete species in an objective, nonarbitrary way? Are later organisms considered to belong to different species from their ancestors because they are sufficiently dissimilar or because they can no longer interbreed with them even if they coexisted? Can such extensive change take place in the absence of speciation?

I cannot attempt to answer fully all of these questions here. Instead, I must limit myself to the remark that, on Simpson's view, species and organisms are quite different sorts of things. An organism undergoes limited change, constrained by its largely unchanging genotype. A single species is capable of indefinite, open-ended development. Although the course of a species' development is constrained from generation to generation by its gene pool, this gene pool is indefinitely modifiable. However, if Eldredge an Gould are right, species are more like organisms than anyone has previously supposed. Both are finite and can undergo only limited change before ceasing to exist. Significant evolutionary change can take place only through a series of successive species, not within the confines of a single species. Species lineages, not species, are the things which evolve. On this view, Hennig's refusal to divide a single lineage into two or more species is preferable to Simpson's alternative.

No disagreement exists between Simpson and Hennig over the situation depicted in figure 18.1b, a single species splitting equally into two. Both agree that the ancestor species is extinct, having given rise to two new daughter species. However, this figure is drawn as if divergence always takes place upon speciation. When this diagram was interpreted as depicting the splitting of one organism into two, divergence was not presupposed. Two euglenae resulting from binary fission are two organisms and not one even though they may be phenotypically and genotypically identical. The same is true of species. Sometimes speciation takes place with no (or at least extremely minimal) divergence; e.g., sibling species are no less two species simply because they look alike. The assumption is, however, that in reproductively isolated species some divergence, at least in the mechanisms of reproduction, must have taken place, even if we cannot detect it. The role of similarity becomes controversial once again when speciation takes place and one species remains unchanged, while the other diverges from the parental type (see figure 18.1c). According to Hennig (17), when speciation occurs, the ancestor species must be considered extinct regardless of how similar it might be to one of its daughter species. Simpson (50) disagrees.

The factor which is causing the confusion in the preceding discussion is the role of similarity in the individuation of species. If species are classes defined by sets (or clusters) of traits, then similarity should be relevant. At one extreme, the pheneticists (54) argue that all that matters is phenetic similarity and dissimilarity, regardless of descent, reproduction, evolutionary cohesiveness, etc. Highly polytypic species such as dogs must be considered numerous different "species" because of the existence of so many reasonably discrete clusters. Sibling species must be considered a single "species" because they form a single cluster. At the other extreme, the Hennigians (commonly termed "cladists") concentrate solely on the splitting of phylogenetic lineages regardless of phenetic similarity. Polytypic species are single species because they form a single clade; sibling species are separate species because they form more than one clade. The evolutionists, represented by Simpson and Mayr, argue that somehow the two considerations must be balanced against each other.

However, on the historical entity interpretation, similarity is a red herring; it is not the issue at all. What really matters is how many organisms are involved and how much the internal organization of the species involved is disrupted. If speciation takes place when a small, peripheral isolate succeeds in bringing about a genetic revolution (see figure 18.1d), then the parent species can still be said to persist unchanged. It has not lost significant numbers of organisms, nor has its internal organization been affected much. One Hennigian, at least, has come to this conclusion for precisely these reasons (60). If, however, the species is split into two or more relatively large subgroups, then it is difficult to see how the ancestral species can still be said to exist, unless one of these subgroups succeeds in retaining the same organization and internal cohesion of the ancestral species. Incidentally, it would also be phenetically similar to the ancestral species, but that would be irrelevant.

Fusion can also take place at the level of species. The breaking down of reproductive isolation sufficient to permit two entire species to merge into one is extremely unlikely (see figure 18.2a). If it did occur, the consideration would be the same as those raised in connection with figure 18.1b. However, introgression and speciation by polyploidy are common (see figures 18.2b and 18.2c). In such cases, a few organisms belonging to separate species mate and produce fertile offspring. Contrary to popular opinion, the production of an occasional fertile hybrid is not enough for biologists to consider two species one. What matters is how extensive the introgression becomes—exactly the right consideration if species are historical entities. As Dobzhansky remarks, "What matters is not whether hybrids can be obtained but whether the Mendelian populations do or do not exchange genes, and if they do whether at a rate which destroys the adaptive equilibrium of the populations concerned'' (9, p. 586).

One final parallel between organisms and species warrants mentioning. Organisms are unique. When an organism ceases to exist, numerically that same organism cannot come into existence again. For example, if a baby were born today who was identical in every respect to Adolf Hitler, including genetic makeup, he still would not be an Adolf Hitler. He would be as distinct and separate a human being as ever existed because of his unique ''insertion into history,'' to use Vendler's propitious phrase (58; see also 57). But the same observation can be made with respect to species. If a species evolved which was identical to a species of extinct pterodactyl save origin, it would still be a new, distinct species. Darwin himself notes, ''When a species has once disappeared from the face of the earth, we have reason to believe that the same identical form never reappears'' (7, p. 313). Darwin presents this point as if it were a contingent state of affairs, when actually it is conceptual. Species are segments of the phylogenetic tree. Once a segment is terminated, it cannot reappear somewhere else in the phylogenetic tree. As Griffiths observes, the ''reference of an individual to a species is determined by its parentage, not by any morphological attribute'' (15, p. 102).

If species were actually spatiotemporally unrestricted classes, this state of affairs would be strange. If all atoms with atomic number 79 ceased to exist, gold would cease to exist, although a slot would remain open in the periodic table. Later when atoms with the appropriate atomic number were generated, they would be atoms of gold regardless of their origins. But in the typical case, to *be* a horse one must be *born* of horse. Obviously, whether one is a gradualist or saltationist, there must have been instances in which nonhorses (or borderline horses) gave rise to horses. The operative term is still ''gave rise to.'' But what of the science fiction examples so beloved to philosophers? What if a scientist made a creature from scratch identical in every respect to a human being including consciousness, emotionality, a feeling of personhood, etc.? Wouldn't it be included in *Homo sapiens*? It all depends. If all the scientist did was to make such a creature and then destroy it, it was never part of our species. However, if it proceeded to mate with human beings born in the usual way and to produce offspring, introducing its genes into the human gene pool, then it would become part of our species. The criterion is precisely the same one used in cases of introgression. In the evolutionary world view, unlike the Aristotelian world view, an organism can change its species while remaining numerically the same individual (see 19).

One might complain that being born of human beings and/or mating with human beings are biological criteria, possibly good enough for individuating *Homo sapiens*, but inadequate for the humanistic notion of a human being. We are a social species. An entity which played the role of a human being in a society would *be* a ''human being,'' even if it was not born of human beings or failed to mate with human beings. I'm not sure how one makes such decisions, but the conclusion is not totally incompatible with the position being presented in this chapter. Species as they are commonly thought of are not the only things which evolve. Higher levels of organization also exist. Entities can belong to the same cultural system or ecosystem without belonging to the same biological species. As Eugene Odum has put it, ''A human being, for example, is not only a hierarchical system composed of organs, cells, enzyme systems, and genes as subsystems, but is also a component of supraindividual hierarchical systems such as populations, cultural systems, and ecosystems'' (44, p. 1289). If pets or computers function as human beings, then from certain perspectives they might well count as human beings even though they are not included in the biological species *Homo sapiens*.

Biological and Philosophical Consequences

Empirical evidence is usually too malleable to be very decisive in conceptual revolutions. The observation of stellar parallax, the evolution of new species right before our eyes, the red shift, etc. are the sorts of things which are pointed to as empirical reasons for accepting new scientific theories. However, all reasonable people had accepted the relevant theories in the absence of such observations. Initial acceptance of fundamentally new ideas leans more heavily on the increased coherence which the view brings to our general world picture. If the conceptual shift from species being classes to species being historical entities is to be successful, it must eliminate longstanding anomalies both within and about biology. In this section, I set out some of the implications of viewing species as historical entities, beginning with those that are most strictly biological, and gradually working my way toward those that are more philosophical.

The role of type specimens in biological systematics puzzles philosophers and biologists alike. As R. A. Crowson remarks, ''The current convention that a single specimen, the Holotype, is the only satisfactory basic criterion for a species would be difficult to justify logically on any theory but Special Creation'' (5, p. 29). According to all three codes of biological nomenclature, a particular organism, part of an organism, or trace of an organism is selected as the type specimen for each species. In addition, each genus must have its type species, and so on. Whatever else one does with this type and for whatever resson, the name goes with the type.[4] The puzzling aspect of the type method on the class interpretation is that the type need not be typical. In fact, it can be a monster. The following discussion by J. M. Schopf is representative:

> It has been emphasized repeatedly, for the benefit of plant taxonomists, at least, that the nomenclatural type (holotype) of a species is not to be confused or implicated in anyone's concept of what is ''typical'' for a taxon. A nomenclatural type is simply *the specimen*, or other element, with which a name is permanently associated. This element need not be ''typical'' in any sense; for organisms with a complicated life cycle, it is obvious that no single specimen could physically rep-

> resent all the important characteristics, much less could it be taken to show many features near the mean of their range of variation. (see also 6, 39, 50, 51; 49, p. 1043)

Species are polymorphic. Should the type specimen for *Homo sapiens*, for instance, be male or female? Species are also polytypic. What skin color, blood type, etc. should the type specimen for *Homo sapiens* have? Given the sort of variability characteristic of biological species, no one specimen could possibly be "typical" in even a statistical sense (37, p. 369). On the class interpretation, one would expect at the very least for a type specimen to have many or most of the more important traits characteristic of its species (16, p. 465–466), but on the historical entity interpretation, no such similarity is required. Just as a heart, kidneys, and lungs are included in the same organism because they are part of the same ontogenetic whole, parents and their progeny are included in the same species because they are part of the same genealogical nexus, no matter how much they might differ phenotypically. The part/whole relation does not require similarity.

A taxonomist in the field sees a specimen of what he takes to be a new species. It may be the only specimen available or else perhaps one of a small sample which he gathers. The taxonomist could not possibly select a typical specimen, even if the notion made sense, because he has not begun to study the full range of the species' variation. He selects a specimen, any specimen, and names it. Thereafter, if he turns out to have been the first to name the species of which this specimen is part, that name will remain firmly attached to that species. A taxon has the name it has *in virtue of* the naming ceremony, not *in virtue of* any trait or traits it might have. If the way in which taxa are named sounds familiar, it should. It is the same way in which people are baptized.[5] They are named in the same way because they are the same sort of thing—historical entities (see Ghiselin, 13, 14).

But what, then, is the role of all those traits which taxonomists include in their monographs? For example, Article 13 of the Zoological Code of Nomenclature states that any name introduced after 1930 must be accompanied by a statement that "purports to give characteristics differentiating the taxon." Taxonomists distinguish between descriptions and diagnoses. A description is a lengthy characterization of the taxon, including reference to characteristics which are easily recognizable and comparable, to known variability within a population and from population to population, to various morphs, and to traits which can help in distinguishing sibling species. A diagnosis is a much shorter and selective list of traits chosen primarily to help differentiate a taxon from its nearest neighbors of the same rank. As important as the traits listed in diagnoses and descriptions may be for a variety of purposes, they are not definitions. Organisms could possess these traits and not be included in the taxon; conversely, organisms could lack one or more of these traits and be clear-cut instances of the taxon. They are, as the name implies, *descriptions*. As descriptions, they change through time as the entities which they describe change. Right now all specimens of *Cygnus olor* are white. No doubt the type specimen of this species of swan is also white. However, if a black variety were to arise, *Cygnus olor* would not on that account become a new species. Even if this variety were to become predominant, this species would remain the same species and the white type specimen would remain the type specimen. The species description would change but that is all. Organisms are not included in the same species *because* they are similar to the type specimen or to each other but *because* they are part of the same chunk of the genealogical nexus (Ghiselin, 13, 14).

On the class interpretation, the role of particular organisms as type specimens is anomalous. The role of lower taxa as types for higher taxa is even more anomalous. On the class interpretation, organisms are members of their taxa, while lower taxa are included in higher taxa (3). How could entities of two such decidedly different logical types play the same role? But on the historical entity interpretation, both organisms and taxa are of the same logical type. Just as organisms are part of their species, lower taxa are part of higher taxa. Once again, parts do not have to be similar, let alone typical, to be part of the same whole.

A second consequence of treating species as historical entities concerns the nature of biological laws. If species are actually spatiotemporally unrestricted classes, then they are the sorts of things which can function in laws. ''All swans are white,'' if true, might be a law of nature, and generations of philosophers have treated it as such. If statements of the form ''species *X* has the property *Y*'' were actually laws of nature, one might rightly expect biologists to be disturbed when they are proven false. To the contrary, biologists expect exceptions to exist. At any one time, a particular percentage of a species of crows will be non-black. No one expects this percentage to be universal or to remain fixed. Species may be classes, but they are not very important classes because their names function in no scientific laws. Given the traditional analyses of scientific laws, statements which refer to particular species do not count as scientific laws, as they should not if species are spatiotemporally localized individuals (20, 21).

Hence, if biologists expect to find any evolutionary laws, they must look at levels of organization higher than particular taxa. Formulations of evolutionary theory will no more make explicit reference to *Bos bos* than celestial mechanics will refer to Mars. Predictions about these entities should be derivable from the appropriate theories but no uneliminable reference can be made to them. In point of fact, no purported evolutionary laws refer to particular species. One example of such a law is the claim that in diploid sexually reproducing organisms, homozygotes are more specialized in their adaptive properties than heterozygotes (31, p. 397). Evolutionary theory deals with the rise of individual homeostasis as an evolutionary mode, the waxings and wanings of sexuality, the constancy or variability of extinction rates, and so on. People are dismayed to discover that evolutionists can make no specific predictions about the future of humankind *qua* humankind. Since that's all they are interested in, they conclude that evolutionary theory is not good for much. But dismissing evolutionary theory because it cannot be used to predict the percentage of people who will have blue eyes in the year 2000 is as misbegotten as dismissing celestial mechanics because it cannot be used to predict the physical make-up of Mars. Neither theory is designed to make such predictions.

The commonest objection raised by philosophers against evolutionary theory is that its subject matter—living creatures—is spatiotemporally localized (52, 53; see also 42). They exist here on earth and nowhere else. Even if the earth were the only place where life had arisen (and that is unlikely), this fact would not count in the least against the spatiotemporally unrestricted character of evolutionary theory. "Hitler" refers to a particular organism, a spatiotemporally localized individual. As such, Hitler is unique. But organisms are not. Things which biologists would recognize as organisms could develop (and probably have developed) elsewhere in the universe. "*Homo sapiens*" refers to a particular species, a spatiotemporally localized individual. As such it is unique, but species are not. Things which biologists would recognize as species could develop (and probably have developed) elsewhere in the universe. Evolutionary theory refers explicitly to organisms and species, not to Hitler and *Homo sapiens* (see 43, 48).

One advantage to biologists of the historical entity interpretation of species is that it frees them of any necessity of looking for any lawlike regularities at the level of particular species. Both "Richard Nixon has hair" and "most swans are white" may be true, but they are hardly laws of nature. If forces them to look for evolutionary laws at higher levels of analysis, at the level of *kinds* of species. It also can explain certain prevalent anomalies in philosophy. From the beginning, a completely satisfactory explication of the notion of a natural kind has eluded philosophers. One explanation for this failure is that the traditional examples of natural kinds were a mixed lot. The three commonest examples of natural kinds in the philosophical literature have been geometric figures, biological species, and the physical elements. By now it should be clear that all three are very different sorts of things. No wonder a general analysis, applicable equally to all of them, has eluded us.

Some of the implications of treating species as historical entities are more philosophical in nature. For example, one of Ludwig Wittgenstein's most famous (or infamous) contributions to philosophy is that of family resemblances, a notion which itself has a family resemblance to cluster concepts and multivariate analysis (64). Such notions have found their most fertile ground in ethics, aesthetics, and the social sciences. Hence, critics have been able to claim that defining a word in terms of statistical covariation of traits merely results from ignorance and informality of context. If and when these areas become more rigorous, cluster concepts will give way to concepts defined in the traditional way. The names of biological species have been the chief counter-example to these objections. Not only are the methods of contemporary taxonomists rigorous, explicit, objective, etc., but also good reasons can be given for the claim that the names of species can never be defined in classical terms. They are inherently cluster concepts (18). On the analysis presented in this chapter, advocates of cluster analysis lose their best example of a class term which is, nevertheless, a cluster concept. If "*Homo sapiens*" is or is not a cluster concept, it will be for the same reason that "Moses" is or (more likely) is not.

A second philosophical consequence of treating species as historical entities concerns the nature of scientific theories. Most contemporary philosophers view scientific theories as atemporal conceptual objects. A theory is a timeless set of axioms and that is that. Anyone who formulates a theory consisting of a particular set of axioms has formulated that theory period. Theories in this sense cannot change through time. Any change results in a new theory. Even if one decides to get reasonable and allow for some variation in axioms, one still must judge two versions of a theory to be versions of the "same" theory because of similarity of axioms. Actual causal connections are irrelevant. However, several philosophers have suggested that science might profitably be studied as an "evolutionary" phenomenon (4, 21, 27, 28, 29, 45, 46, 56). If one takes these claims seriously and accepts the analysis of biological species presented in this chapter, then it follows that whatever conceptual entities are supposed to be analogous to species must also be historical entities. Theories seem to be the most likely analog to species. Because biological species cannot be characterized intelligibly in terms of timeless essences, it follows that theories can have no essences either. Like species, theories must be individuated in terms of some sort of descent and cohesiveness, not similarity.

The relative roles of similarity and descent in individuating scientific theories go a long way in explaining the continuing battle between historians and philosophers of science. Philosophers individuate theories in terms of a set (or at least a cluster) of axioms. Historians tend to pay more attention to actual influence. For example, we all talk about contemporary Mendelian genetics. If theories are to be individuated in terms of a single set (or even cluster) of axioms, it is difficult to see the justification of such an appellation. Mendel's paper contained three statements which he took to be basic. Two of these statements were rapidly abandoned at the turn of the century when Mendel's so-called laws were rediscovered. The third has been modified since. If overlap in substantive claims is what makes two formulations versions of the "same" theory, then it is difficult to see the justification for interpreting all the various things which have gone under the title of "Mendelian genetics" versions of the same theory. Similar observations are appropriate for other theories as well, including Darwin's theory of evolution. The theory that was widely accepted in Darwin's day differed markedly from the one he originally set out. Modern theories of evolution differ from his just as markedly. Yet some are "Darwinian" and others not.

When presented with comparable problems, biologists resort to the type specimen. One organism is selected as the type. Any organism related to it in the appropriate ways belongs to its species, regardless of how aberrant the type specimen might turn out to be or how dissimilar other organisms may be. Males and females belong to the same species even though they might not look anything like each other. A soldier termite belongs in the same species with its fertile congeners even though it cannot mate with them. One possible interpretation of Kuhn's notion of an exemplar (27) is that it is designed to function as a type specimen. Even though scientific change is extremely complicated and at times diffuse, one still might be able to designate particular theories by reference to "concrete problem-solutions," as long as one realizes that these exemplars have a temporal index and need not be in any sense typical.[6] Viewing theories as sets (or clusters) of axioms does considerable damage to our intuitions about scientific theories. On this interpretation, most examples of scientific theories degenerate into unrelated formulations. Viewing scientific theories as historical entities also results in significant departures from our usual modes of conception. Perhaps scientific theories really cannot be interpreted as historical entities. If so, then this is just one more way in which conceptual evolution differs from biological evolution. The more these disanalogies accumulate, the more doubtful the entire analogy becomes.

Finally, and most controversially, treating species as historical entities has certain implications for those sciences which are limited to the study of single species. For instance, if enough scientists were interested, one might devote an entire science to the study of *Orycteropus afer*, the African aardvark. Students of aardvarkology might discover all sorts of truths about aardvarks; that it is nocturnal, eats ants and termites, gives birth to its young alive, etc. Because aardvarks are highly monotypic, aardvarkologists might be able to discover sets of traits possessed by all and only extant aardvarks. But could they discover the essence of aardvarks, the traits which aardvarks must have necessarily to be aardvarks? Could there be scientific laws which govern aardvarks necessarily and exclusively? When these questions are asked of aardvarks or any other nonhuman species, they sound frivolous, but they are exactly the questions that students of human nature treat with utmost seriousness. What is human nature and its laws?

Early in the history of learning theory, Edward L. Thorndike (55) claimed that learning performance in fishes, chickens, cats, dogs, and monkeys differed only quantitatively, not qualitatively. Recent work tends to contradict his claim (1). Regardless of who is right, why does it make a difference? Learning, like any other trait, has evolved. It may be universally distributed among all species of animals or limited to a few. It may be present in all organisms included in the same species or distributed less than universally. In either case, it may have evolved once or several times. If "learning" is defined in terms of its unique origin, if all instances of learning must be evolutionarily homologous, then "learning" is limited by definition to one segment of the phylogenetic tree. Any regularities which one discovers are necessarily descriptive. If, on the other hand, "learning" is defined so that it can apply to any organism (or machine) which behaves in appropriate ways, then it *may* be limited to one segment of the phylogenetic tree. It *need* not be. Any regularities which one discovers are at least candidates for laws of learning. What matters is whether the principles are generalizable. Learning may be species specific, but if learning theory is to be a genuine scientific theory, it cannot be limited *necessarily* to a single species the way that Freud's and Piaget's theories seem to be. As important as descriptions are in science, they are not theories.

If species are interpreted as historical entities, then particular organisms belong in a particular species because they are part of that genealogical nexus, not because they possess any essential traits. No species has an essence in this sense. Hence there is no such thing as human nature. There may be characteristics which all and only extant human beings possess, but this state of affairs is contingent, depending on the current evolutionary state of *Homo sapiens*. Just as not all crows are black (even potentially), it may well be the case that not all people are rational (even potentially). On the historical entity interpretation, retarded people are just as much instances of *Homo sapiens* as are their brighter congeners. The same can be said for women, blacks, homosexuals, and human fetuses. Some people may be incapable of speaking or understanding a genuine language; perhaps bees can. It makes no difference. Bees and people remain biologically distinct species. On other, nonbiological interpretations of the human species, problems arise (and have arisen) with all of the groups mentioned. Possibly women and blacks are human beings but do not "participate fully" in human nature. Homosexuals, retardates, and fetuses are somehow less than human. And if bees use language, then it seems we run the danger of considering them human. The biological interpretation has much to say in its favor, even from the humanistic point of view.

Notes

The research for this chapter was supported by NSF grant Soc 75 03535. I am indebted to the following people for reading and criticizing early versions of it: Michael Ghiselin, Stephen Gould, G. C. D. Griffiths, John Koethe, Ernst Mayr, Bella Selan, W. J. van der Steen, Gareth Nelson, Michael Perloff, Mark Ridley, Michael Ruse, Thomas Schopf, Paul Teller, Leigh Van Valen, Linda Wessels, Mary Williams, and William Wimsatt. Their advice and criticisms are much appreciated.

1. Once again I am excluding from the notion of class those "classes" defined by means of a spatiotemporal relation to a spatiotemporally localized individual. Needless to say, I am also excluding such constructions as "similar in origin" from the classes of similarities. I wish the need to state the obvious did not exist, but from past experience it does.

2. In population genetics the distinction between structurally similar genes forming a single lineage and those which do not is marked by the terms "identical" and "independent"; see (41), pp. 56–57.

3. Until recently even the most ardent group selectionists admitted that the circumstances under which selection can occur at the level of populations and/or entire species are so rare that group selection is unlikely to be a major force in the evolutionary process (30, 32, 33). Michael Wade (59), however, has presented a convincing argument to the effect that the apparent rarity of group selection may be the result of the assumptions commonly made in constructing mathematical models for group selection and not an accurate reflection of the actual state of nature. In his own research, the differential survival of entire populations has produced significant divergence.

4. The three major codes of biological nomenclature are (1) the International Code of Botanical Nomenclature, 1966, International Bureau for Plant Taxonomy and Nomenclature, Utrecht; (2) the International Code of Nomenclature of Bacteria, 1966, *International Journal of Systematic Bacteriology*, 16: 459–490; and (3) the *International Code of Zoological Nomenclature*, 1964, International Trust for Zoological Nomenclature, London. In special circumstances the priority rule is waived, usually because the earlier name is discovered only long after a later name has become firmly and widely established.

5. Although the position on the names of taxa argued for in this chapter might sound as if it supported S. Kripke's (26) analysis of general terms, it does not. Taxa names are very much like "rigid designators," as they should be if taxa are historical entities. However, Kripke's analysis is controversial because it applies to *general* terms. It is instructive to note that during the extensive discussion of the applicability of Kripke's notion of a rigid designator to such terms as "tiger," no one saw fit to see how those scientists most intimately concerned actually designated tigers. According to Putnam's principle of the linguistic division of labor (47), they should have. If they had, they would have found rules explicitly formulated in the various codes of nomenclature which were in perfect accord with Kripke's analysis—but for the wrong reason. That no one bothered tells us something about the foundations of conceptual analysis.

6. Kuhn himself (28) discusses taxa names such as "*Cygnus olor*" and the biological type specimen. Unfortunately, he thinks swans are swans because of the distribution of such traits as the color of feathers.

References

1. Bitterman, M. E. 1975. The comparative analysis of learning. *Science* 188: 699–709.

2. Boyden, A. 1954. The significance of asexual reproduction. *Systematic Zoology* 3: 26–37.

3. Buck, R. C., and D. L. Hull. 1966. The logical structure of the Linnaean hierarchy. *Systematic Zoology* 15: 97–111.

4. Burian, R. M. 1977. More than a marriage of convenience: On the inextricability of history and philosophy of science. *Philosophy of Science* 44: 1–42.

5. Crowson, R. A. 1970. *Classification and Biology*. New York, Atherton Press.

6. Davis, P. H., and V. H. Heywood. 1963. *Principles of Angiosperm Taxonomy*. Princeton, Van Nostrand.

7. Darwin, C. 1966. *On the Origin of Species*. Cambridge, Mass., Harvard University Press.

8. Dawkins, R. 1976. *The Selfish Gene*. New York and Oxford, Oxford University Press.

9. Dobzhansky, T. 1951. Mendelian populations and their evolution. In L. C. Dunn (ed.), *Genetics in the 20th Century*. New York, Macmillan, pp. 573–589.

10. Ehrlich, P. R., and P. H. Raven. 1969. Differentiation of populations. *Science* 165: 1228–1231.

11. Eldredge, N., and S. J. Gould. 1972. Punctuated equilibria: An alternative to phyletic gradualism. In T. J. M. Schopf (ed.), *Models in Paleobiology*. San Francisco, Freeman, Cooper and Company, pp. 82–115.

12. Ghiselin, M. T. 1966. On psychologism in the logic of taxonomic controversies. *Systematic Zoology* 15: 207–215.

13. ———. 1969. *The Triumph of the Darwinian Method*. Berkeley and London, University of California Press.

14. ———. 1974. A radical solution to the species problem. *Systematic Zoology* 23: 536–544.

15. Griffiths, G. C. D. 1974. On the foundations of biological systematics. *Acta Biotheoretica* 23: 85–131.

16. Heise, H., and M. P. Starr. 1968. Nomenifers: Are they christened or classified? *Systematic Zoology* 17: 458–467.

17. Hennig, W. 1966. *Phylogenetic Systematics*. Urbana, Illinois, University of Illinois Press.

18. Hull, D. L. 1965, 1966. The effect of essentialism on taxonomy. *British Journal for the Philosophy of Science* 15: 314–326; 16: 1–18.

19. ———. 1968. The conflict between spontaneous generation and Aristotle's metaphysics. *Proceedings of the Seventh Inter-American Congress of Philosophy*. Québec City, Les Presses de l'Université Laval, 2: 245–250.

20. ———. 1974. *Philosophy of Biological Science*. Englewood Cliffs, Prentice-Hall.

21. ———. 1975. Central subjects and historical narratives. *History and Theory* 14: 253–274.

22. ———. 1976. Are species really individuals? *Systematic Zoology* 25: 174–191.

23. ———. 1976. The ontological status of biological species. In R. Butts and J. Hintikka (eds.), *Boston Studies in the Philosophy of Science*, vol. 32, Dordrecht, D. Reidel, pp. 347–358.

24. Huxley, T. H. 1889. Biology. *Encyclopedia Britannica*.

25. Janzen, Daniel. 1977. What are dandelions and aphids? *American Naturalist* 111: 586–589.

26. Kripke, S. S. 1972. Naming and necessity. In D. Davidson and H. Harman (eds.), *Semantics and Natural Language*. Dordrecht, Holland, D. Reidel, pp. 253–355.

27. Kuhn, T. S. 1969. *The Structure of Scientific Revolutions*. Chicago, University of Chicago Press, 2nd ed.

28. ———. 1974. Second Thoughts on Paradigms. In F. Suppe (ed.), *The Structure of Scientific Theory*. Urbana, Illinois, University of Illinois Press.

29. Laudan, L. 1977. *Progress and Its Problems*. Berkeley and London, University of California Press.

30. Levins, R. 1968. *Evolution in Changing Environments*. Princeton, Princeton University Press.

31. Lewontin, R. C. 1961. Evolution and the theory of games. *Journal of Theoretical Biology* 1: 382–403.

32. ———. 1970. The units of selection. *Annual Review of Ecology and Systematics* 1: 1–18.

33. ———. 1974. *The Genetic Basis of Evolutionary Change*. New York, Columbia University Press.

34. Löther, R. 1972. *Die Beherrschung der Mannigfaltigkeit*. Jena, Gustav Fisher.

35. Mayr, E. 1957 (ed.). *The Species Problem*. Washington, D.C., American Association for the Advancement of Science, Publication Number 50.

36. ———. 1959. Isolation as an evolutionary factor. *Proceedings of the American Philosophical Society* 103: 221–230.

37. ———. 1963. *Animal Species and Evolution*. Cambridge, Mass., Belknap Press of Harvard University Press.

38. ———. 1976. Is the species a class or an individual? *Systematic Zoology* 25: 192.

39. Mayr, E., E. G. Linsley, and R. L. Usinger. 1953. *Methods and Principles of Systematic Zoology*. New York, McGraw-Hill Book Company.

40. Meglitsch, P. A. 1954. On the nature of species. *Systematic Zoology* 3: 49–65.

41. Mettler, L. E., and T. G. Gregg. 1969. *Population Genetics and Evolution*. Englewood Cliffs, Prentice-Hall.

42. Monod, J. L. 1975. On the molecular theory of evolution. In R. Harré (ed.), *Problems of Scientific Revolution*. Oxford, Clarendon Press, pp. 11–24.

43. Munson, R. 1975. Is biology a provincial science? *Philosophy of Science* 42: 428–447.

44. Odum, E. P. 1977. The emergence of ecology as a new integrative discipline. *Science* 195: 1289–1293.

45. Popper, K. R. 1972. *Objective Knowledge*. Oxford, Clarendon Press.

46. ———. 1975. The rationality of scientific revolutions. In R. Harré (ed.), *Problems of Scientific Revolution*. Oxford, Clarendon Press, pp. 72–101.

47. Putnam, H. 1974. The meaning of meaning. In K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, vii. Minneapolis, University of Minnesota Press, pp. 131–193.

48. Ruse, M. J. 1973. *The Philosophy of Biology*. London, Hutchinson University Library.

49. Schopf, J. M. 1960. Emphasis on holotype. *Science* 131: 1043.

50. Simpson, G. G. 1945. The principles of classification and a classification of mammals. *Bulletin of the American Museum of Natural History* 85: 1–350.

51. ———. 1961. *Principles of Animal Taxonomy*. New York, Columbia University Press.

52. Smart, J. J. C. 1963. *Philosophy and Scientific Realism*. London, Routledge and Kegan Paul.

53. ———. 1968. *Between Science and Philosophy*. New York, Random House.

54. Sneath, P. H. A., and R. R. Sokal. 1973. *Numerical Taxonomy*. San Francisco, W. H. Freeman and Company.

55. Thorndike, E. L. 1911. *Animal Intelligence*. New York, Macmillan.

56. Toulmin, S. 1972. *Human Understanding*. Princeton, Princeton University Press.

57. Van Fraassen, Bas. 1972. Probabilities and the problem of individuation. In S. A. Luckenbach (ed.), *Probabilities, Problems and Paradoxes*. Encino, Calif., Dickinson Publishing Co., pp. 121–138.

58. Vendler, Z. 1976. On the possibility of possible worlds. *Canadian Journal of Philosophy* 5: 57–72.

59. Wade, M. J. 1978. A critical review of the models of group selection. *Quarterly Review of Biology* 53: 101–114.

60. Wiley, E. O. 1978. The evolutionary species concept reconsidered. *Systematic Zoology* 27: 17–26.

61. Williams, G. C. 1966. *Adaptation and Natural Selection*. Princeton, Princeton University Press.

62. Wilson, E. O. 1975. *Sociobiology: The New Synthesis*. Cambridge, Mass., Belknap Press of Harvard University Press.

63. Wynne-Edwards, V. C. 1962. *Animal Dispersion in Relation to Social Behaviour*. Edinburgh and London, Oliver and Boyd.

64. Wittgenstein, L. 1953. *Philosophical Investigations*. New York, Macmillan.

19 Choosing Among Alternative ''Phylogenetic'' Species Concepts

David A. Baum and Michael J. Donoghue

Several different ''phylogenetic'' species concepts have been proposed, and we consider how to choose among them. There appear to be two main approaches. ''Character-based'' concepts define species on the possession of characters, whereas ''history-based'' concepts are based on historical relatedness. Under the latter view, characters may be used to discover species in practice but they are not viewed as defining attributes of species. To illustrate the distinction we discuss a character-based approach utilizing ''diagnostic'' characters and a history-based approach using genetic coalescence. We argue that the choice between character- and history-based concepts is primarily determined by one's understanding of systematics. If the goal of systematics is simply to describe the hierarchical distribution of characters (''pattern cladistics''), a character-based definition of species is required. In contrast, if systematics is concerned with inferring the evolutionary relationships of organisms (''evolutionary phylogenetics''), a history-based definition of species is needed. We hold the view that phylogenetic systematics is concerned with evolutionary history and therefore we maintain that a phylogenetic species concept should be history-based.

The rise of phylogenetic systematics resulted in dissatisfaction with prevailing species concepts. In particular, the biological species concept (e.g., Mayr 1942) was called into question because it emphasized a property (the capacity to interbreed) that is not necessarily a good guide to relationships (Rosen 1979; Cracraft 1983; Donoghue 1985). In response to the perceived incompatibility of this and other species concepts (e.g., the ''ecological,'' ''cohesion,'' and ''recognition'' concepts) with the principles of phylogenetic systematics, a number of alternative approaches have been proposed. However, despite being motivated by a concern with defining the species category in a manner compatible with phylogenetic systematics (broadly construed), there are significant differences among so-called ''phylogenetic'' species concepts. Our aim is to characterize these different concepts and discuss how a choice might be made among them.

It appears to us that the most fundamental division among different phylogenetic species concepts is that some define species on the basis of characters, whereas others

From *Systematic Botany* 20, no. 4 (October–December 1995), 560–573. Reprinted by permission of the American Society of Plant Taxonomists.

define species in terms of historical relationships or ancestry. Under ''character-based'' concepts, an organism is a member of a given species if and only if it possesses some character (i.e., an observable organismal attribute) or combination of characters. Generally, the origins of these characters are ignored (e.g., whether they are ancestral or derived) as is the actual genealogy of the organisms in question. In contrast, ''history-based'' concepts consider an organism a member of a given species if and only if it is historically related to other organisms in the species. It is important to appreciate that although history-based concepts view characters as irrelevant to the definition of species, characters generally are needed to recognize species and assign organisms to species in practice. Thus, under both history-based and character-based concepts, characters provide the main source of evidence in species delimitation. However, whereas history-based concepts view characters as (fallible) evidence of species existence, character-based concepts view characters as defining attributes of species.

A first step in deciding among ''phylogenetic'' species concepts is to understand clearly the distinction between history- and character-based concepts. This entails a careful evaluation of the consequences of adopting one approach or the other. To do this we will describe and contrast one character-based and one history-based species concept. We will focus on the most completely explicated character-based approach, that formulated by Cracraft (1983, 1989) and extended by Nixon and Wheeler (1990, 1992) and Davis and Nixon (1992). Similarly, we will focus on one of the many history-based approaches, that developed by Baum and Shaw (1995). It should be stressed, however, that there are other character- and history-based concepts besides these two, and the fact that we focus attention on two concepts does not mean we advocate these over the alternatives. Rather, we use them simply to illustrate the two basic approaches to defining phylogenetic species, and to raise general concerns pertinent to the choice between them. This choice, we will argue in the second half of the paper, is guided primarily by one's view of the aim of systematics.

Character-Based Approaches

The character-based concept we will consider defines a species as ''the smallest aggregation of populations (sexual) or lineages (asexual) diagnosable by a unique combination of character states in comparable individuals (semaphoronts)'' (Nixon and Wheeler 1990, p. 218). This concept is called ''the phylogenetic species concept'' or PSC by its proponents, but to avoid confusion with other phylogenetic species concepts (e.g., Rosen 1979; Mishler and Donoghue 1982; Donoghue 1985; Mishler and Brandon 1987) we will refer to it as the ''diagnostic approach.''

The diagnostic approach is clearly character-based, inasmuch as the definition rests on the possession of characters (specifically ''diagnostic'' characters), with considerations of organismal history or relatedness playing a subordinate role. This dependence on characters can be seen by considering a hypothetical example. Consider a situation in which four populations of *Fritillaria* exist, two of which are composed entirely of red-flowered individuals and the other two of white flowered individuals. Perianth color is the only attribute that distinguishes the populations. In this situation two diagnostic species exist, one comprising the two white-flowered populations and the other comprising the red-flowered populations. As this example shows, species limits are determined, and organisms are assigned to species, based solely upon the distribution of characters, without reference to evolutionary history. However, although character-based, the diagnostic approach requires additional information aside from characters, as shown by further examples.

Imagine that *Fritillaria* were found growing in a fifth locality, but at this site both red and white perianths occurred (figure 19.1). This finding is subject to two alternative interpretations: 1) the new locality constitutes a single population and perianth color is a polymorphic trait [i.e., not a "character" in the sense of Nixon and Wheeler (1990)]; hence, all five populations are members of the same diagnostic species (figure 19.1A), or 2) the newly discovered site is not a single population but two sympatric populations, one composed of the red-flowered species and the other of the white-flowered species (figure 19.1B). Clearly, the choice between these alternatives rests on what is meant by the term "population."

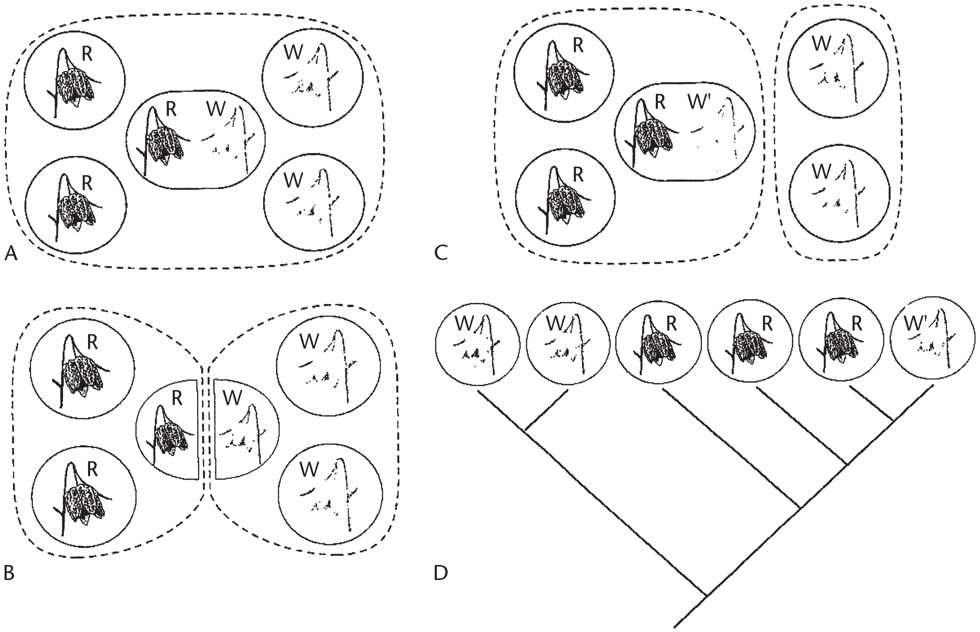


Figure 19.1

It should be obvious that a "population" is not simply a group of organisms living in one place, because this would result in distantly related organisms (e.g., oak trees and squirrels) being assigned to the same population and thus species. Traditionally, population definitions rest on interbreeding or reproductive cohesion (e.g., Dobzhansky 1950). Davis and Nixon (1992, p. 430) adopted this view stating that, when considering the local units aggregated into species, populations are "the arenas in which most genetic recombination occurs." Thus, application of the diagnostic approach to sexual organisms implies knowledge of patterns of gene flow in nature.

This concept of population clearly applies only to sexual organisms, and a different approach must be taken for asexuals. Nixon and Wheeler (1990) used "lineage" in place of "population" when applying their species definition to asexuals. By "lineage," they mean a group of organisms that has a unique combination of characters, whether or not they comprise all the descendants of an ancestral organism. Thus, despite the connotation of history implied by the term "lineage" the diagnostic approach as applied to asexuals is also character-based. Delimiting "lineages" is the same as delimiting diagnostic species, which in turn is the same as finding any discrete character variation: "when unique character combinations occur in asexual or clonal forms, these forms should be recognized as distinct species" (Nixon and Wheeler 1990, p. 219).

The reason we bring up the issue of the treatment of asexual organisms is not because we have a problem with a species concept that applies only to sexuals (or conversely to asexuals), and not because of worries about there being "too many" species. Rather, we wish to clarify that although minimal "diagnostic" groups can be delimited in principle in both sexual and asexual organisms, species status is achieved in different ways under the different circumstances and, therefore, "species" are not comparable entities (e.g., Frost and Hillis 1990). In sexual organisms breeding relations are critical whereas in asexuals all that is needed is the possession of similar characters. What unites these both as "species" is not some underlying biological property, but simply the fact that they pass some operational test ("diagnosability").

Returning now to the hypothetical *Fritillaria* example, if it is decided that the fifth site is indeed a single population that is polymorphic for flower color, then, at least under early versions of the diagnostic approach (Cracraft 1983, 1989; Nixon and Wheeler 1990), one would recognize only a single species. However, Davis and Nixon (1992) proposed a modification of the concept of "character." They retained the view of a diagnostic character as a fixed attribute but noted that "by 'fixed' we do not mean that the character is necessarily observed as monomorphic but that it occurs in all individuals of the lineage, in either its original or in a transformed state" (p. 424). Thus, in the *Fritillaria* example, two species would exist if, for example, the white flowers in the polymorphic population were derived from the red flowers in that population rather than being derived from white flowered individuals of the other "species" (figure

19.1C). This is an issue of homology assessment, which can be resolved by determining the historical relations among the character-states (figure 19.1D). Although this may be difficult to unravel in practice (especially in cases involving reversal or parallel loss of a state), in principle it provides a criterion for determining whether an attribute is diagnostic. However, there are complications that Davis and Nixon (1992) have not adequately addressed.

To illustrate one such problem, imagine two populations differing only in the alleles they manifest at a particular allozyme locus. Population 1 has alleles a and b, whereas population 2 has alleles c and d. Assume that allele d was derived from c which was derived from b which was, in turn, derived from a. If the approach advocated by Davis and Nixon (1992) is applied consistently, population 2 would be judged to be fixed for the diagnostic character (c, d) because it contains all organisms with either the original state (c) or the derived state (d). In contrast, population 1 is not fixed for any diagnostic character, because it contains no character that does not also occur (in the derived state) in population 2. If diagnostic species must have fixed characters, then is population 1 a species? In the spirit of Nixon and Wheeler (1990), population 1 should be considered a diagnostic species, but to do so seems contrary to the historical conception of characters promoted by Davis and Nixon (1992). If one followed Nixon and Wheeler (1990) and considered both populations diagnosable species then, in effect, one is applying two definitions of "fixation." Population 2 is "fixed" because it has a unique historical character, whereas population 1 is "fixed" because the organisms in it bear characters absent from population 2. Again, as in the cases of sexual vs. asexual "species," such "species" share no substantive properties; they are equivalent only in the sense of having passed an operational test.

So far our discussion of the diagnostic approach has focused on its application at a point in time. However, Nixon and Wheeler (1992) explore its implications for time extended lineages. For them, character-fixation (i.e., the extinction of an ancestral trait) is synonymous with speciation and therefore all character fixation entails "speciation," regardless of proximity to lineage-branching events. In other words, "speciation" occurs in unbranching lineages each time the last organism bearing an ancestral trait dies. When branching is involved, "speciation" is not linked with the branching event *per se* but occurs as soon as the two (or more?) lineages become diagnosably distinct (figure 19.4). This can occur through fixation of a derived trait in the speciating population (Nixon and Wheeler, 1992), or alternatively, a population can speciate when it is already fixed for a derived trait and that trait goes extinct in all other lineages (figure 19.4).

The time extended model of speciation proposed by Nixon and Wheeler (1992) is consistent with the diagnostic approach, but has several worrisome implications (in addition to effectively ignoring the distinction between anagenesis and cladogenesis). For example, if one considers a "trait" to be any heritable feature of an organism,

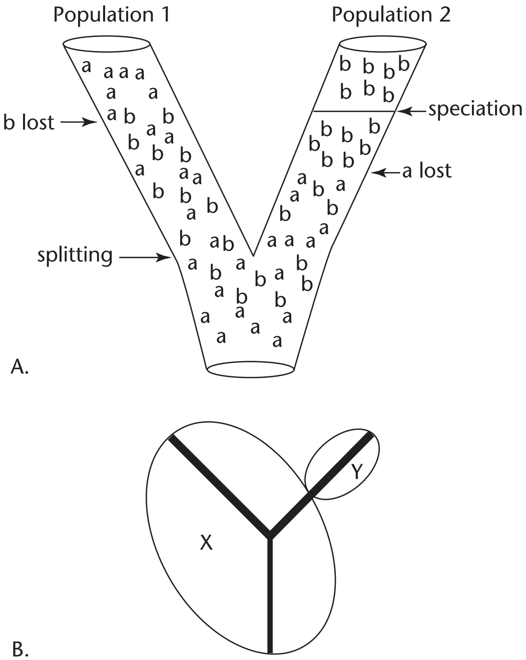


Figure 19.2
Diagnostic species through time. A. The events leading to the formation of a descendant species Y, from an ancestral species, X, are shown. The ancestral lineage, comprising a single population, is polymorphic for an ancestral trait, a, and a derived trait, b. The population splits into two isolated populations (1 and 2). Population 2 then loses a an some time later, population 1 loses b. Speciation of population 2 as species Y occurs with loss of the derived trait in population 1 and is not coincident with character fixation in population 2. B. The temporal boundaries of diagnostic species X and Y on the population tree.

then each organism has a potentially huge number of traits. This means that each organismal death is likely to cause at least one ancestral trait to become extinct and, hence, one derived trait to become fixed (the expected number will depend upon the number of traits per genome and the population size). Thus, in many plausible demographic situations, Nixon and Wheeler's (1992) approach will imply ''speciation'' more than once per generation. While it is important to recognize conflicts with prevailing views, such as those we have just described, such conflicts are not by themselves grounds for rejecting this or any other ''phylogenetic'' species concept. We could choose to revamp existing speciation theory instead. Rather, we must compare this approach to alternative species concepts and then choose among them based on consistency with the perceived goals of systematics.

History-Based Approaches

A variety of history-based definitions of species are possible and several of these have been proposed. For example, it is clear that Hennig (1966) considered genealogical relationships to be the main concern in defining species (and other taxa), with characters being viewed merely as the evidence by which one could infer relationships (e.g., pp. 30, 79–80). It is note-worthy, for example, that in his frequently reproduced figures 19.4 and 19.6, both depicting speciation, no characters are shown. Likewise, several recent attempts to formulate phylogenetic species concepts granted primacy to history over characters (e.g., Mishler and Donoghue 1982; Donoghue 1985; Mishler and Brandon 1987; de Queiroz and Donoghue 1988, 1990; Frost and Hillis 1990).

In order to highlight the basic issues that any history-based approach must confront, we have chosen to focus on just one such concept, the gene coalescence view outlined recently by Baum and Shaw (1995). In doing so, our purpose is not to promote their "genealogical species concept" as the solution. In fact, whereas the two of us agree completely on the need for a history-based species concept (see below), we do not agree entirely on which of these concepts is best.

The central aim of a history-based species concept is to define species based on historical relationships. However, the notion of historical relationship needs clarification. Especially critical here is the need to identify concepts that are sensitive to the fact that within populations of biparental organisms relationships are reticulate whereas between higher taxa they are divergent (Hennig 1966). There must be a boundary between these two types of relationship, and it is at this boundary that history-based species concepts have generally attempted to locate species. An obstacle that needs to be overcome is to develop a history-based concept of "relationship" that can potentially apply both within and among populations.

The concept of monophyly, as currently defined by many systematists (a single ancestral species and all of the species descended from it), cannot logically apply to species themselves or to entities below that level (Hennig 1966; de Queiroz and Donoghue 1988; McKitrick and Zink 1988; Wheeler and Nixon 1990). For this reason, de Queiroz and Donoghue (1988) suggested an expansion of the concept of monophyly to allow entities other than "species" to be ancestors, including individual organisms or breeding pairs (a possibility hinted at by Hennig 1966, e.g., p. 209; also see Donoghue 1985; Mishler and Brandon 1987). In principle, this broader definition of monophyly permits species to be monophyletic. However, because this concept of monophyly extends down to the level of mating pairs and their offspring, species concepts based on monophyly have tended to add "ranking" criteria such as the ability to interbreed (e.g., Mishler and Brandon 1987). If one wishes to define species without such non-historical criteria, however, it is necessary to base the definition on a historical attribute other than monophyly.

The historical property we will focus on here is "exclusivity," where an exclusive group of organisms is one whose members are more closely related to each other than they are to any organisms outside the group (de Queiroz and Donoghue 1990; Baum 1992; Baum and Shaw 1995). Before continuing to discuss the implementation of exclusivity in a definition of species it is necessary to clarify the origin of the concept, because there has been some confusion on this point.

As well as presenting his well known definition of monophyly (see above). Hennig (1966) also stated another definition of the concept: "a group of species in which every species is more closely related to every other species than to any species that is classified outside the group" (p. 73). Apparently he recognized that when a phylogeny is strictly divergent, the two definitions of monophyly would always apply to the same groups, and in view of this synonymy he evidently saw no need to coin a new term for the property described by the alternative definition. In broadening the concept of monophyly to include descent from any common ancestor, de Queiroz and Donoghue (1988, 1990) noted that a strict correspondence between the alternative aspects of monophyly broke down. That is, the descendants of a particular ancestor may not form a group all of whom are each others closest relatives (e.g, within a population of sexually reproducing organisms). For this reason they suggested that the term "exclusivity" be used for the property analogous to Hennig's alternative version of monophyly (de Queiroz and Donoghue 1990). It should be noted, however, that the concept of "exclusivity" is not identical to either of Hennig's definitions of monophyly, because it may apply to groups of organisms rather than only to groups of species. Furthermore, whereas monophyly (in the standard usage) cannot logically apply to species (McKitrick and Zink 1988; Wheeler and Nixon 1990), exclusivity can (see below). Thus, even if the broadened definition of monophyly (de Queiroz and Donoghue 1988) were rejected (which would, incidentally, require some other term to be developed for the general phenomenon of descent from a common ancestor), the concept of exclusivity is still a valuable one, distinct from any developed by Hennig.

Under the species concept of Baum and Shaw (1995), species are viewed as basal, exclusive taxa; that is, taxa containing within them no subgroups that are themselves exclusive. In order to decide whether a group is exclusive the degree of relatedness of its constituent organisms, and between these organisms and other organisms outside the group, must be evaluated. The novel feature of this genealogical species concept is that, rather than being based on descent from an ancestral organism, relatedness is viewed in terms of the genealogical descent of the genome as a whole. This can be accomplished within the conceptual framework of the recently developed branch of population genetics named "coalescent theory" [see Hudson (1990) for an introduction to the field, and Maddison (1995) for a discussion of its relevance to phylogenetic systematics]. To illustrate this approach it is easiest to go through a hypothetical example.

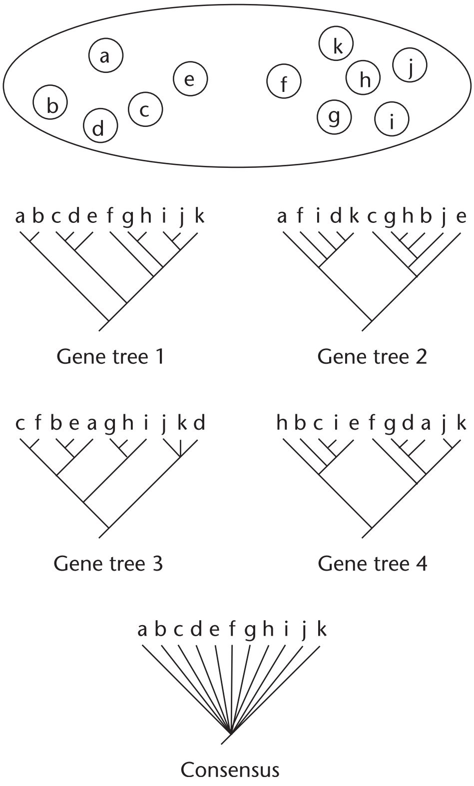


Figure 19.3
The relationship among phylogenies of different unlinked genes from 11 organisms (a–k) within a single panmictic population (note only one allele of each organism is considered for each locus). Four gene trees are shown and their consensus. No groups of organisms are clades on all four gene trees.

Imagine 11 sexual organisms in the same panmictic population (figure 19.3). The copies of a homologous gene (i.e., a piece of DNA inherited from a common ancestral gene and small enough not to have recombined) present in each of the 11 organisms will have a tree-like history (the gene-tree). Looking, for example, at gene 1 in figure 19.3, the alleles possessed by organisms A and B trace back to a common ancestral gene (i.e., ''coalesce'') more recently than either coalesces with C. Another gene, unlinked to the first, will have its own independent gene tree, which is likely to show a pattern of coalescence that is somewhat different from that of the first gene. For example, in figure 19.3, gene 2 shows more recent coalescence between organisms B and C than between either of these and A. In such a panmictic population, there should be no groups of organisms that form a clade in all of the genetrees (Avise and Ball 1990). Under the concept of exclusivity, as applied by Baum and Shaw (1995), there are, thus, no exclusive groups within the population.

Imagine now that this population had split very recently into two genetically isolated populations. The pattern of coalescence will not be any different from that found in the panmictic population, namely, the organisms of both populations will together constitute a single exclusive group without any exclusive subgroups. However, if these two populations continue to be genetically isolated, many of the gene lineages present at the time of the split will become extinct in one or the other population (or both). Eventually, if the populations remain isolated long enough, all copies of any gene present in one population will coalesce with each other before coalescing with copies in the other population (figure 19.4; see Avise and Ball 1990). At this point the two populations each constitute an exclusive group of organisms and, assuming that they contain no exclusive subgroups, are "genealogical species" (Baum and Shaw 1995).

Extrapolating from this example, we may insert the coalescent view of relatedness into the concept of exclusivity and define a species as: a basal group of organisms all of whose genes coalesce more recently with each other than with those of any organisms outside the group. Notice that this concept is history-based because species are defined solely on genealogical history rather than on characters. As explained earlier, this certainly does not prevent characters from being used to infer history, and reference to characters will almost certainly be necessary in practice. However, whether a group of organism is a species is determined by the genealogical history of their genes, and gene trees exist regardless of whether that history can be reconstructed by reference to characters.

Many systematists are uncomfortable with a species definition that precludes knowing with certainty whether a group of organisms is a species. However, the act of describing a species can be viewed, and often has been, as the formulation of an hypothesis that a group of organisms has some special property (such as genetic isolation, or even independent creation). A species concept then is seen as guiding taxonomists as to the criteria that are relevant in testing a species hypothesis. However, it need not provide a prescribed set of operations by which "species" are discovered (Frost and Hillis 1990). The genealogical approach to defining species provides a concept of what a species is [contrary to the claim of Frost and Kluge (1994)] and, thus, suggests methods (e.g., gene tree analysis) by which particular species hypotheses can be tested rigorously.

As we have shown, the genealogical concept defines species based solely on the historical relationships of their constituent organisms [contrary to the implication in O'Hara (1992) that all species concepts must be prospective]. However, the genealogical species concept does have some implications that conflict with taxonomic tradi-

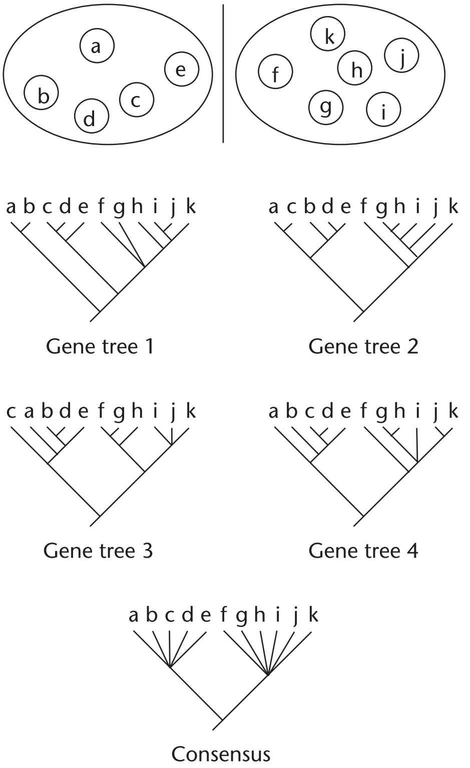


Figure 19.4

The relationship among phylogenies of different unlinked genes from 11 organisms in two populations that have been genetically isolated for a long time. Four gene trees are shown and their consensus. Organisms in population 1 (a–e) and those in population 2 (f–k) always appear as distinct clades.

tion, of which three are worth briefly summarizing. We refer the reader to Baum and Shaw (1995) for discussion of other implications of the coalescent approach.

If the two populations shown speciating in figure 19.4 were unequal in size, coalescent theory predicts that the smaller population would become exclusive before the larger. Therefore, there will be a period of time during which the smaller population is a genealogical species, but the organisms of the larger population are not members of any genealogical species [see figure 5 in Baum and Shaw (in press)]. The larger population in this example has no history that is not also shared with the small population. However, to reflect the likelihood that its descendants become a distinct exclusive group, it and analogous groups of organisms may be termed ‘‘metaspecies’’ (Donoghue 1985; de Queiroz and Donoghue 1988). This poses a problem due to the traditional requirement that all organisms be ascribable to a species. However, it may well be that history-based concepts of taxa (at any level) will conflict with existing taxonomic conventions (de Queiroz and Gauthier 1992).

Another issue raised by the coalescent approach, which we think will apply to other history-based concepts, is "fuzziness." Because different genes coalesce at different rates the boundary between reticulate and divergent genealogy is not sharp. We believe that this fuzziness is an inherent feature of such systems rather than a result of the particular definitions of reticulation and divergence advocated by Baum and Shaw (1995). Thus any concept aiming to place species limits at the reticulate/divergent boundary must accept a certain fuzziness. However, it should be noted that, provided a single historical criterion is applied (e.g., exclusivity), fuzziness reflects the way the world is rather than a lack of conceptual clarity. Thus, whereas fuzziness may complicate the act of delimiting species in practice, it need not undermine the theoretical utility of a history-based species concept.

Finally, it should be noted that the definition of genealogical species given above applies to the organisms living at one point in time. Baum and Shaw (1995) suggested two alternatives for dealing with this fact. First, basal exclusive groups can be equated with species, meaning that species would not be seen as persisting through time (analogous to semaphoronts; Hennig 1966). Alternatively, basal exclusive groups could be equated with the temporal cross-sections of species, which would therefore be equated with time-extended lineages. Future discussions of history-based species definitions (and other types of species definition) will need to address the issue of temporal extent.

Choosing Between a Character-Based and History-Based Approach

We have described both a character-based and a history-based species concept and have shown that these approaches are quite distinct and that they each appear to be internally consistent. They differ in the ways they conflict with current ideas in evolutionary biology and taxonomy, but this, we would argue, should not be the basis on which to choose between them. Similarly, the choice should not be achieved by comparing the species delimited under each concept with the "species" that a "good taxonomist" recognizes. Instead, we think that the choice between alternative species concepts must be discussed in the context of the overall aims and rationale of systematics. Thus, it is necessary to characterize alternative versions of phylogenetic systematics and examine their compatibility with character-based and history-based species concepts.

As we see it, among the systematists who consider themselves to be descendants of Hennig, there are currently two poles, which de Queiroz and Donoghue (1990) termed "cladistics" and "phylogenetic systematics." These designations are, however, potentially confusing (e.g., because the methods referred to as "cladistics" are used by proponents of both views) so well will use modifiers that we hope will minimize misunderstanding: "pattern cladistics" and "evolutionary phylogenetics." Under the pattern cladistic view, systematics is concerned with providing a theory-neutral description of the hierarchic distribution of characters, whereas under the evolutionary phylogenetic view it is concerned with reconstructing descent relationships (de Queiroz 1992; de Queiroz and Gauthier 1990). We believe that the choice between the pattern cladistic and evolutionary phylogenetic views is critical in choosing between alternative phylogenetic species concepts.

Species from a Pattern Cladistic Perspective

The pattern cladistic view [perhaps best exemplified by Patterson (1988) and Nelson (1989)] argues for the primacy of "observation" (characters) over "theory" (evolutionary history). Thus, cladograms are viewed as summaries of character distributions rather than depictions of evolutionary history. Following from this, taxa are seen as being properties of characters rather than the reverse (Nelson 1989), and monophyletic higher taxa are implicitly or explicitly defined as groups of species having shared characters. It is, therefore, internally consistent to define species on the basis of characters. There are, however, two alternative character-based species concepts that are compatible with the pattern cladistic view of systematics, depending upon whether one assumes that there is a lower boundary below which cladistic methods are invalid. Denying such a boundary leads to species being viewed as operational taxa composed of those individual organisms that, in a given analysis, are identical for all the characters under consideration (Vrana and Wheeler 1992). However, most proponents of the cladistic view assume that there is a lower bound below which characters are not distributed hierarchically [i.e., the boundary between phylogenetic and tokogenetic relationships (Hennig 1966)] and, thus, they consider species to be the least inclusive groups that show hierarchic character distributions. This perspective seems to have led to the diagnostic approach discussed above, perhaps using the following reasoning. 1) Organisms are inappropriate as terminals in cladistic analyses (at least for sexuals) and, therefore, some collection of organisms ("population" or "lineage") must be used instead. 2) Only populations/lineages possessing fixed differences can be assumed to be hierarchically related. 3) Therefore, populations/lineages can be aggregated together until each aggregation has a fixed difference from all other such aggregations (e.g., Davis and Manos 1991; Davis and Nixon 1992). 4) Because these aggregations cannot be broken up using the evidence at hand, they are appropriate basal taxa, that is, species.

It seems that a character-based approach is consistent with the pattern cladistic view of systematics, but would a history-based species concept also be acceptable? Clearly, the answer is "no." A history-based definition, such as the genealogical species concept, makes a number of assumptions; for example, it assumes that evolution occurred and that we can use evolutionary theory to devise methods for reconstructing phylogenetic history. This conflicts with the pattern cladists' attempt to avoid assumptions and use only "theory-neutral" methodologies. The incompatibility of history-based concepts and pattern cladistics is amply demonstrated by the fact that history-based approaches imply the existence of species that cannot be discovered by reference to characters, which is unthinkable under pattern cladistics. Furthermore in pattern cladistics, taxa, and hence species, are defined based on operations performed on the data at hand (i.e., characters) and therefore, if they are to be delimited consistently, it is important that only one "discovery procedure" (parsimony) be admitted (Nelson 1989). In contrast, history-based species concepts permit evidence other than characters to enter into decisions about species delimitation. For example, information on the vagility of organisms and biogeography might be brought to bear in assessing whether an individual variant in a population originated in that population or represented gene flow from another population (i.e., distinguishing between the alternatives in figure 19.1). Furthermore, even when only characters are used, diverse methods of analysis can contribute to the evaluation of a species hypothesis (e.g., maximum likelihood).

Species from an Evolutionary Phylogenetic Perspective

Under the evolutionary phylogenetic view [as exemplified by de Queiroz and Donoghue (1990), de Queiroz and Gauthier (1990), and de Queiroz (1992)] the goal is to discover historical entities that exist in the real world (i.e., have causes or are effects; Ereshefsky 1992) as inferred within an external theoretical framework including, but not limited to, current evolutionary theory. Taxa (including species) are viewed as monophyletic or exclusive groups of organisms and thus, by virtue of ancestry and descent and extinction of lineages, they exist independently of the occurrence of synapomorphies. Characters, under this view, are seen as a basis for hypothesizing taxon status and testing such hypotheses. Exactly how species and other taxa are to be defined historically remains to be seen, but the point here is that history-based species concepts are in principle compatible with an evolutionary phylogenetic approach to systematics.

Is a diagnostic or other character-based species concept compatible with phylogenetics? We think the answer is "no." Under an evolutionary phylogenetic perspective, systematists are concerned with describing evolutionary history and, therefore, it follows that the entities defined are those that gain their existence by virtue of history. Character-based approaches define species based on the distribution of characters among organisms (and/or populations) rather than on historical relationships and therefore must be rejected.

It could perhaps be argued that since the distribution of characters among organism is a product of evolutionary history, species concepts such as the diagnostic approach are in some sense history-based and therefore compatible with evolutionary phyloge-

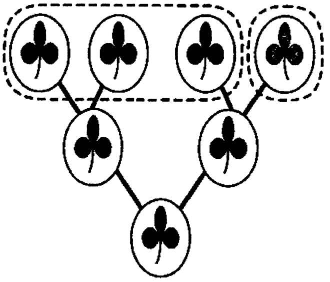


Figure 19.5

A hypothetical example to illustrate the non-historical nature of diagnostic species. An ancestral *Trifolium* population has undergone a series of cladogenetic events. One of the four extant populations has become fixed for the presence of a light v-shaped mark on the leaves. Two diagnostic species must be delimited among the extant populations, one comprising individuals with marked and the other with unmarked leaves.

netics. This argument is, however, fallacious because it is characters that give diagnostic species their existence, not history. Nonetheless, if it were the case that the diagnostic criterion succeeded infallibly in delimiting historically meaningful groups, then this point would be a mere philosophical quibble. If, on the other hand, diagnostic species can be shown to be non-historical groups in some cases, then an evolutionary phylogeneticist would have practical as well philosophical grounds for rejecting the diagnostic approach. For this reason it is important to evaluate the claim by Davis and Nixon (1992, p. 429) that: "if every phylogenetic species exhibits a unique combination of characters, each is an extended genealogical population, all of whose constituent individuals are historically related." This can be achieved by considering two examples.

Imagine an ancestral population of *Trifolium* with unmarked leaves, which gives rise to a number of isolated daughter populations through a series of cladogenetic events that are not marked by any discrete character evolution (figure 19.5). Note that the previous statement is possible only from an evolutionary phylogenetic perspective because, under pattern cladistics, cladogenesis does not occur without character evolution (Nixon and Wheeler 1992). Suppose that in one of the descendant populations a mutation occurs for the presence of a v-shaped mark on the leaves and that this mutation goes to fixation in this population (figure 19.5). Applying the diagnostic criterion described above, two *Trifolium* species would be recognized among the extant populations, one comprising all the unmarked clover populations (of which three are shown in figure 19.5) and the other comprising only the population with v-marked leaves.

In claiming that all individuals in diagnostic species are "historically related," Davis and Nixon (1992) are arguing that both the marked and unmarked clover "species" are composed of organisms that are "historically related" to each other. However, while it may be true that the members of the marked clover species are historically related (more information would be needed to apply the coalescent approach discussed above), it is clear that individuals with unmarked leaves are not historically related. The unmarked populations have no common history that is not also shared with the marked population and thus they constitute a group that is analogous to a paraphyletic higher taxon. Phylogenetic systematists should therefore agree that the unmarked clover "species" is not composed of "historically related" organisms under any substantive definition of that phrase.

A second example serves to emphasize further the possible non-historical nature of diagnostic species. Imagine an ancestral population of *Senecio* containing individuals with both rayed and rayless capitula (figure 19.6). This splits into two similarly polymorphic populations, but these daughter populations give rise to populations all of which are fixed for either the rayed or rayless form. Applying the diagnostic definition to the extant populations, two diagnostic species would have to be recognized, one composed entirely of rayed and the other of rayless individuals (figure 19.6). However, neither the rayed nor rayless species are composed of historically related individuals, each being analogous to a polyphyletic higher taxon (further information would be needed to determine which groups of populations, if any, are species using the coales-

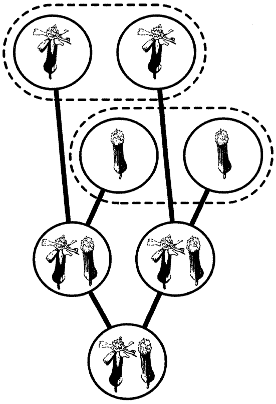


Figure 19.6

A hypothetical example to illustrate the non-historical nature of diagnostic species. An ancestral *Senecio* population polymorphic for the presence of ray florets gave rise to two daughter populations each, likewise polymorphic. These in turn gave rise to the four extant populations, all of which have become fixed for either rayed or rayless forms. Two diagnostic species must be delimited among the extant populations, one comprising individuals with rayed and the other with rayless capitula.

cent approach presented above). It should be noted that recognition of two diagnostic species rests on the distribution of characters in the current populations and ignores information on when those characters became fixed. Thus, even if we had full knowledge that lineage sorting had occurred as depicted in figure 19.6, we would nonetheless have to recognize two species.

These two examples illustrate that, notwithstanding claims to the contrary, the diagnostic approach can lead to the recognition of species that lack historical meaning. This should not be surprising because there are many biological situations in which character distributions fail to track history. As well as the retention of ancestral characters and lineage sorting (illustrated with our hypothetical examples), strong local adaptation and introgressive hybridization can lead to groups that are historically unrelated but nonetheless manifest diagnostic characters. When such conflicts arise between history and characters, the diagnostic approach would give primacy to characters and, as a result, groups with fundamentally different historical structure (e.g., the "paraphyletic" and "monophyletic" *Trifolium*) are not conceptually discriminated. Since evolutionary phylogenetics is above all interested in reconstructing history, this view of systematics is incompatible with the diagnostic approach.

How to Choose a Species Concept

It should be clear from the foregoing arguments that we think the choice between history- and character-based species concepts rests critically on whether one takes a pattern cladistic or evolutionary phylogenetic view of systematics. Pattern cladistics demands a character-based concept (e.g., the diagnostic approach), whereas evolutionary phylogenetics demands a history-based concept (e.g., the coalescent approach).

We hold an evolutionary phylogenetic view, and therefore favor a history-based concept of species. Faced with the fact that we cannot know evolutionary history with certainty, we think that our concepts should at least be designed to inspire us to learn as much as possible about that history. The alternative reaction to this uncertainty—the one taken by pattern cladists—is to retreat to the false certainty of operation-alism. Under this view we "know" only what we observe, and furthermore only those things that can be "observed directly" are admitted to exist at all. Retreats such as this have happened before in the history of systematics (e.g., phenetics), and in other branches of science, but these have ultimately seemed unsatisfactory (Hull 1968). We think that the reason for this is that scientists are interested in more than surface appearances, and instead want to learn how the world works. The decision not to look beyond the immediately observable is, in our view, fundamentally a decision not to engage in the scientific enterprise. We think that operational definitions of key terms such as species, by discouraging the attempt to penetrate surface appearances, impede our ability to learn about the world. They are simply not bold enough. Physicist have not been content to define electrons as clicks emitted by a Geiger-counter and neither should systematists be satisfied with defining species as things discovered using a particular procedure.

Notwithstanding our own preference for evolutionary phylogenetics, the main point of our paper is simply that one's general philosophy of systematics is the major consideration guiding the choice between alternative "phylogenetic" species concepts. In fact, we think the choice among species-concepts is made more-or-less automatically by deciding on the overall rationale for systematics. It is quite obvious, however, that many find it difficult to make that choice.

We suspect that many systematists hold no particular view of the philosophy of systematics, or hold a composite view to the effect that characters and history tend not to conflict in practice, so there is no need to decide which is more important. However, this inclination to "sit on the fence" in the debate between pattern cladistics and evolutionary phylogenetics certainly does not imply that such a position is tenable in the long run. In fact, because historical groups and character-based groups may often fail to coincide in nature, an intermediate or mixed position will serve only to perpetuate confusion about the ontology of taxa. The debate over alternative species concepts, while perhaps unseemly at times, is ultimately beneficial if it leads to a clarification of alternative philosophies of systematics and if systematists are compelled to confront the broader question of what they are trying to achieve.

Acknowledgments

We are grateful to John Freudenstein, Kevin de Queiroz, Kerry Shaw, and an anonymous reviewer for helpful comments and discussion, to Jerrold Davis for arranging the symposium, and to Gerald Gastony for agreeing to act as editor of this paper.

Literature Cited

Avise, J. C. and R. M. Ball. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology* 7: 45–67.

Baum, D. A. 1992. Phylogenetic species concepts. *Trends in Ecology and Evolution* 7: 1–2.

——— and K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pp. 289–303 in *Experimental and molecular approaches to plant biosystematics*, eds. P. C. Hoch and A. G. Stephenson. St. Louis: Missouri Botanical Garden.

Cracraft, J. 1983. Species concepts and speciation analysis. *Current Ornithology* 1: 159–187.

———. 1989. Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. Pp. 28–59 in *Speciation and its consequences*, eds. D. Otte and J. A. Endler. Sunderland, Massachusetts: Sinauer Associates, Inc.

Davis, J. I. and P. S. Manos. 1991. Isozyme variation and species delimitation in *Puccinellia nuttalliana* complex (Poaceae): An application of the phylogenetic species concept. *Systematic Botany* 16: 431–445.

——— and K. C. Nixon. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology* 41: 421–435.

de Queiroz, K. 1992. Phylogenetic definitions and taxonomic philosophy. *Biology and Philosophy* 7: 295–313.

——— and M. J. Donoghue. 1988. Phylogenetic systematics and the species problem. *Cladistics* 4: 317–338.

——— and ———. 1990. Phylogenetic systematics or Nelson's version of cladistics? *Cladistics* 6: 61–75.

——— and J. Gauthier. 1990. Phylogeny as a central principle in taxonomy: Phylogenetic definitions of taxon names. *Systematic Zoology* 39: 307–322.

——— and ———. 1992. Phylogenetic taxonomy. *Annual Review of Ecology and Systematics* 23: 449–480.

Dobzhansky, T. 1950. Mendelian populations and their evolution. *American Naturalist* 84: 401–418.

Donoghue, M. J. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist* 88: 172–181.

Ereshefsky, M. 1992. Eliminative pluralism. *Philosophy of Science* 59: 671–690.

Frost, D. R. and D. M. Hillis. 1990. Species in concept and practice: Herpetological applications. *Herpetologica* 46: 87–104.

——— and A. C. Kluge. 1994. A consideration of epistemology in systematic biology, with special reference to species. *Cladistics* 10: 259–294.

Hennig, W. 1966. *Phylogenetic systematics*. Urbana: Univ. of Illinois Press.

Hull, D. L. 1968. The operational imperative: Sense and nonsense in operationism. *Systematic Zoology* 17: 438–457.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* 7: 1–44.

Maddison, W. 1995. Phylogenetic histories within and among species. Pp. 273–287 in *Experimental and molecular approaches to plant biosystematics*, eds. P. C. Hoch and A. G. Stephenson. St. Louis: Misouri Botanical Garden.

Mayr, E. 1942. *Systematics and the origin of species from the viewpoint of a zoologist*. New York: Columbia Univ. Press.

McKitrick, M. C. and R. M. Zink. 1988. Species concepts in ornithology. *Condor* 90: 1–14.

Mishler, B. D. and R. N. Brandon. 1987. Individuality, pluralism and the phylogenetic species concept. *Biology and Philosophy* 2: 397–414.

——— and M. J. Donoghue. 1982. Species concepts: A case for pluralism. *Systematic Zoology* 31: 491–503.

Nelson, G. 1989. Cladistics and evolutionary models. *Cladistics* 5: 275–289.

Nixon, K. C. and Q. D. Wheeler. 1990. An amplification of the phylogenetic species concept. *Cladistics* 6: 211–223.

——— and ———. 1992. Extinction and the origin of species. Pp. 119–143 in *Extinction and phylogeny*, eds. M. J. Novacek and Q. D. Wheeler. New York: Columbia Univ. Press.

O'Hara, R. J. 1992. Systematic generalization, historical fate, and the species problem. *Systematic Biology* 42: 231–246.

Patterson, C. 1988. The impact of evolutionary theories on systematics. Pp. 59–91 in *Prospects in systematics*. Systematics Association Special Volume, No. 36. Oxford: Oxford Univ. Press.

Rosen, D. E. 1979. Fishes from the uplands and intermontane basins of Guatemala: Revisionary studies and comparative geography. *Bulletin of the American Museum of Natural History* 162: 267–375.

Vrana, P. and W. Wheeler. 1992. Individual organisms as terminal entities: Laying the species problem to rest. *Cladistics* 8: 67–72.

Wheeler, Q. D. and K. C. Nixon. 1990. Another way of *looking at* the species problem: A reply to de Queiroz and Donoghue. *Cladistics* 6: 77–81.

X Phylogenetic Inference

20 Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading

Joseph Felsenstein

For some simple three- and four-species cases involving a character with two states, it is determined under what conditions several methods of phylogenetic inference will fail to converge to the true phylogeny as more and more data are accumulated. The methods are the Camin–Sokal parsimony method, the compatibility method, and Farris's unrooted Wagner tree parsimony method. In all cases the conditions for this failure (which is the failure to be statistically consistent) are essentially that parallel changes exceed informative, nonparallel changes. It is possible for these methods to be inconsistent even when change is improbable a priori, provided that evolutionary rates in different lineages are sufficiently unequal. It is by extension of this approach that we may provide a sound methodology for evaluating methods of phylogenetic inference.

Parsimony or minimum evolution methods were first introduced into phylogenetic inference by Camin and Sokal (1965). This class of methods for inferring an evolutionary tree from discrete-character data involves making a reconstruction of the changes in a given set of characters on a given tree, counting the smallest number of times that a given kind of event need have happened, and using this as the measure of the adequacy of the evolutionary tree. (Alternatively, one can compute the weighted sum of the numbers of times several different kinds of events have occurred.) One attempts to find that evolutionary tree which requires the fewest of these evolutionary events to explain the observed data. Camin and Sokal treated the case of irreversible changes along a character state tree, minimizing the number of changes of character states required. A number of other parsimony methods have since appeared in the systematic literature (Kluge and Farris, 1969; Farris, 1969, 1970, 1972, 1977; Farris, Kluge, and Eckhardt, 1970) and parsimony methods have also found widespread use in studies of molecular evolution (Fitch and Margoliash, 1967, 1970; Dayhoff and Eck, 1968; see also Fitch, 1973). Cavalli-Sforza and Edwards (1967; Edwards and Cavalli-Sforza, 1964) earlier formulated a minimum evolution method for continuous-character data.

From *Systematic Zoology* 27, no. 4 (December 1978): 401–410. © 1978 Joseph Felsenstein. Reproduced by permission of Taylor & Francis, Inc., http://www.taylorandfrancis.com

An alternative methodology for phylogenetic inference is the compatibility method, introduced by Le Quesne (1969, 1972). He suggested that phylogenetic inference be based on finding the largest possible set of characters which could simultaneously have all states be uniquely derived on the same tree. The estimate of the phylogeny is then taken to be that tree. While Le Quesne's specific suggestions as to how this might be done have been criticized by Farris (1969), his general approach, which is based on Camin and Sokal's (1965) concept of the compatibility of two characters, has been made rigorous and extended in a series of papers by G. F. Estabrook, C. S. Johnson, Jr., and F. R. McMorris (Estabrook, 1972; Estabrook, Johnson, and McMorris, 1975, 1976a, 1976b; Estabrook and Landrum, 1975).

There has been relatively little examination of the properties of parsimony or compatibility methods as methods of statistical inference. Farris (1973, 1977) has shown that a number of different parsimony methods produce maximum likelihood estimates of an "evolutionary hypothesis" consisting of a phylogeny along with the reconstructed states of the characters in a large number of ancestral populations. However, when the object is to estimate only the phylogeny, the Camin–Sokal method has not been proven to give a maximum likelihood estimate except in the case when the probabilities of change in the character states are known to be small (Felsenstein, 1973).

For a given probabilistic model of evolution, one can construct a maximum likelihood estimate of the phylogeny, given the observed data on a set of discrete characters. Phylogenies constructed by the proper maximum likelihood method typically have the property of *consistency*. A statistical estimation method has the property of consistency when the estimate of a quantity is certain to converge to its true value as more and more data are accumulated. The purpose of this paper is to show that parsimony methods (as exemplified by the criterion of Camin and Sokal and by Farris's unrooted tree method) as well as compatibility methods do not possess the property of consistency in all cases. This is done by constructing a particular three-species case in which lack of consistency can be proven, a case in which parallel evolution is relatively probable. In finding such a case, we have thereby also shown that Farris's (1973) maximum likelihood estimate of the "evolutionary hypothesis" can give an inconsistent estimate of the phylogeny, since it always gives the same estimate as a parsimony method. Although it had been suspected that Farris's estimate of the phylogeny might be inconsistent, it was previously known only that it was not the same as direct maximum likelihood estimation of the phylogeny (Felsenstein, 1973), and no actual proof of its inconsistency had been made.

The result may be regarded as warning us of the weakness of parsimony and compatibility methods. Alternatively, the conditions which must hold in order to have lack of consistency may be regarded as so extreme that the result may be taken to be a validation of parsimony or compatibility approaches. Readers must decide for themselves. In either case the conclusion reached will have the merit of being based on an examination of the properties of phylogenetic methods when considered as methods of statistical inference. Systematists may be tempted to reject this sort of attempt to evaluate phylogenetic methods by the criteria of statistical inference, particularly in view of the oversimplified models of evolution used here. It would seem difficult to take such a reaction seriously if unaccompanied by an attempt to erect a more adequate set of criteria, or to use the present criteria to examine more realistic models of evolution.

To show that a parsimony or compatibility method does not yield a consistent estimate of the phylogeny, it is not sufficient simply to show that it does not yield a maximum likelihood estimate. There are many examples known in statistics of consistent estimation methods which are not maximum likelihood estimates. For example, in samples drawn independently from a normal distribution, the maximum likelihood estimate of the mean of the underlying normal distribution is the sample mean. But the sample median is also a consistent estimator of the true mean. As more and more points are collected, it too will approach the true mean. By analogy to this case it might be argued that, although parsimony and compatibility estimates of the phylogeny are not maximum likelihood estimates, they do provide consistent estimates of the phylogeny. While this will often be the case, we shall see that this conjecture is not always true.

The Example

The example involves characters each of which has two states, 0 and 1. The ancestral state in each character is 0 and the derived state is 1. It is possible for the state of a population to change from 0 to 1, but not to revert from state 1 to state 0. Suppose that we have observed three species A, B, and C and that the (unknown) true phylogeny is as given in figure 20.1. Once a character is in state 1 at the beginning of a segment of the tree, it will not change thereafter, so that all we need to know for each segment is the probability that a character which is in state 0 at the beginning of the segment will have changed to state 1 by the end of the segment. These probabilities are assumed to be the same for all characters in this particular case; they are the quantities P, Q, and R shown in figure 20.1 next to the segments. In this particular case, the probabilities of change are assumed to be the same in segments II and IV of the tree, and the same in segments III and V. This is done purely to make the algebra easier: this assumption could be relaxed somewhat without altering the qualitative conclusions. It is important to realize that the constancy of P, Q, and R from character to character, and the differences between them from segment to segment, amount to strong assumptions about the biological situation. The differences in the probability of change may be due to the segments' being of different length in time (so that the tip species are not contemporaneous). Alternatively, they may be due to differences in the rate of evolution per unit time, differences from segment to segment of a sort which affect all characters.

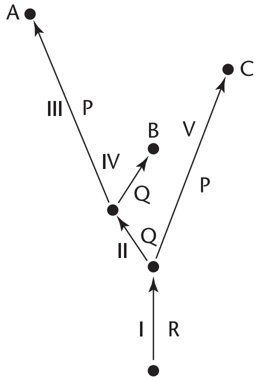


Figure 20.1
An evolutionary tree with three tip species. The segments of the tree are numbered I through V, and next to each is shown the probability of change from state 0 to state 1 in the segment.

This amounts to the assumption that there are true differences in the overall rates of evolution of different lineages.

In figure 20.1 the segments of the tree are also numbered with Roman numerals. Knowing the probability of 0 → 1 change in each segment, we can easily obtain the probabilities of each of the possible combinations of states in the tips. For example, for the three tip species to be in states 1, 1, and 0 respectively, there must have been no change from state 0 in segments I and V. There may have been a 0 → 1 change in segment II, or else no change in that segment but 0 → 1 changes in both segments III and IV. The probability of observing states 1, 1, and 0 is thus

$$P_{110} = (1 - R)[Q + (1 - Q)PQ](1 - P) \tag{1}$$

Similarly, we can compute the probabilities of all eight possible configurations of character states:

$$P_{000} = (1 - P)^2(1 - Q)^2(1 - R) \tag{2a}$$

$$P_{001} = P(1 - P)(1 - Q)^2(1 - R) \tag{2b}$$

$$P_{010} = (1 - P)^2Q(1 - Q)(1 - R) \tag{2c}$$

$$P_{011} = P(1 - P)Q(1 - Q)(1 - R) \tag{2d}$$

$$P_{100} = P(1 - P)(1 - Q)^2(1 - R) \tag{2e}$$

$$P_{101} = P^2(1 - Q)^2(1 - R) \tag{2f}$$

$$P_{110} = (1 - P)[Q + (1 - Q)PQ](1 - R) \tag{2g}$$

$$P_{111} = PQ[P(1 - Q) + 1][1 - R] + R \tag{2h}$$

Results of the Camin–Sokal Parsimony Method

If we examine N characters in these three species, we can count how many of the characters are in each of the eight possible combinations: $000, 001, 010, \ldots, 111$. Let us call the resulting numbers of characters $n_{000}, n_{001}, \ldots, n_{111}$. We can use these numbers to discover what will be the result of applying the Camin–Sokal parsimony method to these data. When a character has the configuration 000, then no matter which phylogeny we propose, no changes of character state will be required to explain the evolution of this character along that phylogeny. There are four other configurations of the data which will require only one character state change to be assumed, no matter what phylogeny is postulated. These are 001, 010, and 100, which require one character state change on the segment of the evolutionary tree leading to a single species, as well as 111, which requires a single change at the root of the tree.

The remaining three configurations, 110, 101, and 011, will require different numbers of changes of state on different phylogenies. Let us represent the three possible bifurcating phylogenies as (AB)C, A(BC), and (AC)B, placing parentheses around monophyletic groups. On the phylogeny (AB)C, the configuration 110 requires only one change while the others require two changes. If we let

$$S = n_{001} + n_{010} + n_{100} + n_{111} + 2(n_{110} + n_{101} + n_{011}), \tag{3}$$

then (AB)C requires $S - n_{110}$ changes of state to be assumed. By similar logic, A(BC) requires $S - n_{011}$ changes, and (AC)B requires $S - n_{101}$ changes. Which tree we estimate depends on which requires us to assume the fewest changes of character state. We can immediately see that the Camin–Sokal parsimony method will estimate the correct phylogeny as (AB)C if and only if $n_{110} \geq n_{101}, n_{011}$. When n_{011} is the greatest of these three numbers A(BC) will be the estimate, and when n_{101} is the greatest the estimate will be (AC)B. When there are ties for the greatest of n_{110}, n_{101}, and n_{011}, there will be two or more possible estimates.

Inconsistency of the Result

We assume that the N characters have evolved independently of one another, and have been chosen for study without regard to the configuration of their character states in these three species. Each character may be regarded as falling independently into one of the eight configurations $000, \ldots, 111$ with probabilities $P_{000}, \ldots, P_{111}$. So the n_{ijk} are drawn from a multinomial distribution with these probabilities.

In such a case, an elementary application of the Strong Law of Large Numbers (e.g., Feller, 1957:243–244) tells us that as we let $N \to \infty$, $n_{ijk}/N \to P_{ijk}$ for all configurations ijk. In particular, this implies that as we score more and more characters, n_{110} will ultimately become larger and remain larger than either n_{101} or n_{011} if and only if

$P_{110} > P_{101}, P_{011}$. Whichever of these three probabilities is largest determines which of the three bifurcating phylogenies is certain to be the ultimate estimate as we accumulate more and more characters. Thus the condition for the Camin–Sokal estimate to have the property of consistency is simple: that P_{110} be greater than or equal to both P_{101} and P_{011}. Note in particular that if this condition does not hold, the consequences are striking: if, say $P_{101} > P_{110}, P_{011}$, then as we accumulate more and more information the Camin–Sokal parsimony method is *increasingly certain to give the wrong answer*, in this case (AC)B.

We now examine the conditions on P and Q which are required in order to have inconsistency of the Camin–Sokal parsimony methods. The three quantities P_{110}, P_{101}, and P_{011} are given by the expressions (2g), (2f), and (2d). Note that all of these quantities contain a common factor of $(1 - R)$. Provided that $R < 1$ (which we assume), this factor can be dropped. The condition $P_{110} \geq P_{011}$ then becomes

$$(1 - P)[Q + (1 - Q)PQ] \geq P(1 - P)Q(1 - Q) \tag{4}$$

which simplifies to

$$Q(1 - P) \geq 0. \tag{5}$$

This will always hold, so in the present case it will always be true that $P_{110} \geq P_{011}$. Now we need only inquire whether $P_{110} \geq P_{101}$. This is the same as asking whether

$$(1 - P)[Q + (1 - Q)PQ] \geq P^2(1 - Q)^2, \tag{6}$$

which is equivalent to requiring that

$$0 \geq P^2(1 - Q) + PQ^2 - Q. \tag{7}$$

Let us view this as a quadratic equation in P whose coefficients depend on Q. Since $1 - Q > 0$ (which we assume), the quadratic in (7) has a minimum at $P = -Q/(1 - Q)$. Since this is never positive, the positive values of P for which (7) is satisfied are those values of P below the point where the quadratic function is zero:

$$P \leq P_1 = (-Q^2 + [Q^4 + 4Q(1 - Q)]^{1/2})/2(1 - Q) \tag{8}$$

P_1 is always a real number, so no complications arise. Figure 20.2 shows P_1 plotted for values of Q between 0 and 1. P_1 rises from 0 to 1 as Q goes from 0 to 1. Above the P_1 curve is the region of values of P for which $P_{110} < P_{101}$.

This is the region in which the Camin–Sokal parsimony method is guaranteed to converge to the wrong estimate of the tree as we accumulate more and more data. Note that for every possible value of Q there is a range of values of P in which we will encounter this unpleasant behavior. A similar statement holds if we rearrange (7) to obtain limits on the values of Q as a function of P, so that for every value of P there is a range of Q values in which this unpleasant behavior occurs. Note that for small Q, the condition (8) is closely approximated by

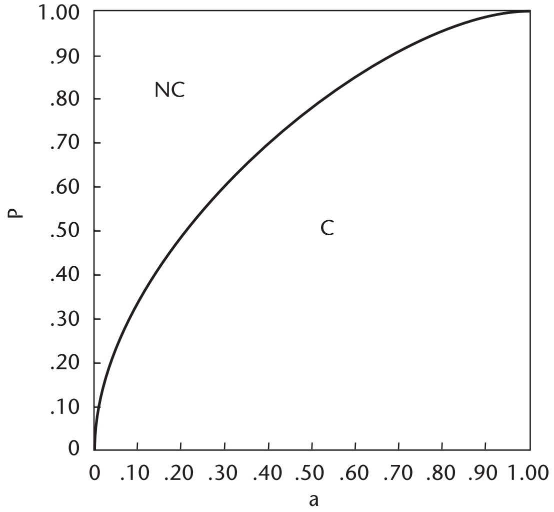


Figure 20.2

Values of P and Q for which the Camin-Sokal method fails to be consistent in the present case. C denotes the region of consistency, NC the region of inconsistency. Their boundary is the curve relating P_1 to Q.

$$P \leq Q^{1/2}, \tag{9}$$

and for Q near 1 it is closely approximated by

$$P \geq 1 - (1 - Q)^2. \tag{10}$$

The effect of (8) is that the Camin–Sokal method will tend to fail when there is a sufficient disproportion between P and Q, which is the same as requiring that there be a sufficiently great disproportion between the lengths of the long and the short segments of the tree in figure 20.1.

In a previous paper (Felsenstein, 1973), I showed that for sufficiently small probabilities of evolutionary change, the Camin–Sokal method yields a correct maximum likelihood estimate of the phylogeny, and hence would be consistent. This might appear to be contradicted by (8) and (9), since these show that the Camin–Sokal method can be inconsistent even when P and Q are small. But my earlier proof involved holding the lengths (in time) of the segments of the tree constant while letting the rate of change in the characters become small. This is equivalent to holding the ratio of P to Q constant while letting both approach zero. As will be apparent from dividing both sides of (9) by Q, when this is done the values of P and Q enter the region of consistency for sufficiently small values of P and Q, no matter what the (constant) ratio P/Q. So in this sense, the Camin–Sokal method works for sufficiently small rates of character state change.

Compatibility Methods

It is a convenient fact that precisely the same three-species example also allows us to find conditions in which the compatibility methods yield an inconsistent estimate of the phylogeny. While the original approach of Le Quesne assumed that the direction of character state change was unknown, and could not be applied to a three-species case, the extensions of the compatibility approach by Estabrook, Johnson, and McMorris do allow us to make inferences in a three-species case when the direction of change on the character-state trees is known. For example, if two binary characters have the states $(1,1)$, $(1,0)$, and $(0,1)$ respectively in the three species, then it is impossible for the transition $0 \rightarrow 1$ to have taken place only once in each character on a branching phylogeny.

A pairwise consideration of all of the eight possible outcomes of the data will show that the outcomes 110, 101, and 011 are mutually incompatible, but that all other combinations are compatible. If we are trying to find the phylogenies suggested by the largest possible set of mutually compatible characters, these will include (AB)C if and only if $n_{110} \geq n_{101}, n_{011}$. Thus, the compatibility method for rooted binary character-state trees will give the same estimate as the Camin–Sokal method in the three-species case. We thus can apply all of the above conditions for inconsistency of the Camin–Sokal method to the compatibility approach. This allows the conclusion that consistency is not a general property of the compatibility methods, but must be proven for specific probability models of evolution if it is desired.

Unrooted Wagner Trees

One of the most widely used parsimony methods has been Farris's (1970) method of inferring unrooted evolutionary trees under the assumption that character-state changes are reversible. The consistency of this method can be investigated by an extension of the present approach to a four-species case. This is necessary because there is only one possible unrooted tree in the three-species case, rendering it trivial. Figure 20.3 shows an unrooted tree with four species, A, B, C, and D. In order to more closely approximate the evolutionary model which underlies Farris's method, we assume that although the characters were originally in state 0, they have the same probability of reversion $1 \rightarrow 0$ once they are in state 1, as they have of origination $0 \rightarrow 1$ of state 1 when they are in state 0. Thus each segment of the evolutionary tree is characterized by a probability of character state change which applies equally to both forward change $0 \rightarrow 1$ and reversion $1 \rightarrow 0$. Once again, we assume for simplicity that characters are independently sampled and all have the same probabilities. There are 16 possible data outcomes, 0000 through 1111. Once again, the outcome of applying Farris's parsimony method will depend only on the numbers of characters $n_{0000}, \ldots, n_{1111}$ having each outcome.

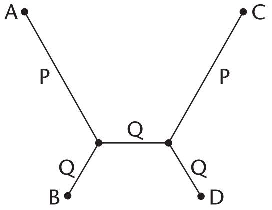


Figure 20.3
True unknown phylogeny (with root omitted) used to find cases in which unrooted Wagner tree parsimony methods will be inconsistent.

It is easy to show, along the same lines as before, that whether the unrooted tree obtained is of form (AB)(CD), (AC)(BD), or (AD)(BC) is determined by which of the three numbers $n_{1100} + n_{0011}$, $n_{1010} + n_{0101}$, and $n_{1001} + n_{0110}$ is largest. It is not difficult to demonstrate that the exact placement of the root of the true tree will affect only the relative probabilities of obtaining 1100 and 0011, but will leave the total probability $P_{1100} + P_{0011}$ unchanged, and similarly for $P_{1010} + P_{0101}$ and $P_{1001} + P_{0110}$. Therefore, we need not specify the placement of the root on the (unknown) true tree to compute the probabilities which determine the outcome of this parsimony method. Suppose that the true phylogeny is one whose unrooted form is given in figure 20.3. We may as well assume that the root is at the left-hand end of the central segment, and that all characters start there in state 0, as these assumptions do not affect $P_{1100} + P_{0011}$ and the other relevant probabilities.

Considering the two possible character states at the right-hand end of the central segment, we find that

$$\begin{aligned} P_{1100} + P_{0011} = {} & PQ[1-Q)^2(1-P) + Q^2P] \\ & + (1-P)(1-Q)[Q(1-Q)(1-P) + Q(1-Q)P] \end{aligned} \tag{11}$$

with analogous expressions for the other two relevant probabilities:

$$\begin{aligned} P_{1010} + P_{0101} = {} & P(1-Q)[Q^2(1-P) + (1-Q)^2P] \\ & + (1-P)Q[Q(1-Q)P + Q(1-Q)(1-P)] \end{aligned} \tag{12}$$

and

$$\begin{aligned} P_{1001} + P_{0110} = {} & P(1-Q)[Q(1-Q)P + Q(1-Q)(1-P)] \\ & + Q(1-P)[(1-Q)^2P + Q^2(1-P)]. \end{aligned} \tag{13}$$

After some elementary but tedious algebra it can be shown from (12) and (13) that provided that $Q \leq \frac{1}{2}$, which we assume,

$$P_{1010} + P_{0101} \geq P_{1001} + P_{0110}. \tag{14}$$

This establishes that when the true tree is as shown in figure 20.3, our estimate of the unrooted tree topology may converge to either (AB)(CD) or to (AC)(BD), but never to (AD)(BC) as we collect more and more characters. So to establish the consistency of the estimation of unrooted tree topology, we need only enquire whether

$$P_{1100} + P_{0011} \geq P_{1010} + P_{0101}, \tag{15}$$

which will be the condition for consistency. Using (11) and (12) we find after further tedious algebra that (15) is simply,

$$2P^2Q - P^2 + 2Q^3 - 3Q^2 + Q \geq 0 \tag{16}$$

which is

$$(2Q - 1)(P^2 + Q(Q - 1)) \geq 0. \tag{17}$$

Since $Q \leq \frac{1}{2}$, (17) is simply

$$P^2 \leq Q(1 - Q), \tag{18}$$

a considerably simpler condition than (8). Note that when Q is small, (18) reduces to (9). Thus, all the statements about consistency in the Camin–Sokal case when P and Q are small are also correct in the case of unrooted Wagner trees.

Discussion

We have seen that there are circumstances under which three different estimation methods are not statistically consistent, these being the Camin–Sokal parsimony method, the Estabrook–Johnson–McMorris compatibility method, and Farris's parsimony method for estimating unrooted Wagner trees. For small values of P and Q, the condition for inconsistency amounts to requiring that simultaneous changes on two long segments of the tree be more probable a priori than one change on a short segment. This amounts to requiring that parallelism of changes be more probable than unique and unreversed change in an informative part of the tree (e.g., that simultaneous changes in segments III and V of the tree in figure 20.1 be more probable than a single change in segment II). This certainly seems like a reasonably intuitive condition for inconsistency. The advantage of the argument presented here lies not in leading to a particularly surprising conclusion, one that will cause abandonment of these parsimony and compatibility methods, but as a formal investigation of one of the statistical properties of phylogenetic inference methods.

The models employed here certainly have severe limitations: it will hardly ever be the case that we sample characters independently, with all of the characters following the same probability model of evolutionary change. Extending this analysis to more realistic evolutionary models will certainly be difficult. Yet the task must be undertaken:

if inconsistency of a parsimony or compatibility technique is suspected, it does little good simply to point out that the evolutionary models employed here do not apply to the type of data being encountered in practice. That amounts to a confession of ignorance rather than validation of the inference method in question.

Likelihood Methods

Methods of phylogenetic inference which entirely avoid the problem of statistical inconsistency are already known. Maximum likelihood estimation of the phylogeny is one of them. I have outlined elsewhere (Felsenstein, 1973) how this may be done. In the three species cases maximum likelihood estimation methods can easily be developed. The likelihood of a tree will simply be

$$L = \prod_{ijk} P_{ijk}^{n_{ijk}}, \tag{19}$$

where P_{ijk} is the probability of data configuration ijk and n_{ijk} is the number of characters having that configuration. Estimation is carried out by maximizing (19) over the unknown parameters of the evolutionary model (such as P and Q in equations [2]). This is done for each tree topology, and the final estimate consists of the topology and the evolutionary parameters which yield the highest likelihood. (Note that despite the connotations of the term, the likelihood of a tree is *not* the probability that it is the correct tree.) When there are larger numbers of species, the number of possible data configurations (the number of terms n_{ijk}) in each character becomes so large that it is impractical to use equation (19). I have presented elsewhere (Felsenstein, 1973) an algorithm for evaluating the likelihood of a tree which avoids this difficulty.

Maximum likelihood estimates are not desirable in themselves, but because they have desirable statistical properties such as consistency and asymptotic efficiency. In the case of discrete multistate characters under the sorts of evolutionary model considered here, it can be shown quite generally that the maximum likelihood estimation procedure has the property of consistency. In particular, in the case of the tree shown in figure 20.1, it will be a consistent method whatever the values of P, Q, and R.

The reader familiar with the paper of Farris (1973), which establishes a general correspondence between parsimony methods and maximum likelihood methods may be puzzled at this stage: if parsimony methods are maximum likelihood methods, why have the two been described here as separate methods? Why is one sometimes not consistent while the other is always consistent? This paradox is resolved once one recalls that the maximum likelihood methods used by Farris are different from those described in Felsenstein (1973) and here. Farris used the maximum likelihood method to estimate not only the parameters of the evolutionary tree, but also the states of the characters in a large number of ancestral populations. When this latter kind of maximum likelihood estimate is made, the number of parameters being estimated rises without limit as more characters are examined.

From the point of view of estimating the phylogeny, these extra parameters are ''nuisance'' parameters. As a result of their presence, the ratio between the number of data items and the number of parameters does not increase indefinitely as more characters are added. It is in situations such as this that maximum likelihood methods are particularly prone to lack of consistency, as I have previously pointed out (Felsenstein, 1973). Indeed, the present results establish that there are conditions under which Farris's likelihood method (giving the same results as a parsimony method) fails to be consistent.

Perspective

The weakness of the maximum likelihood approach is that it requires us to have a probabilistic model of character evolution which we can believe. The uncertainties of interpretation of characters in systematics are so great that this will hardly ever by the case. We might prefer to have methods which, while not statistically optimal for any one evolutionary model, were robust in that they had reasonable statistical properties such as consistency for a wide variety of evolutionary models. The present results establish that parsimony and compatibility methods can fail to be consistent if parallelism is expected to occur frequently. This helps establish that they do not yield maximum likelihood estimates. However, they pass the test of consistency when parallelism is rare. This leaves them as viable candidates for robust methods. Establishing that robustness (or disproving it) by examining a wider range of models is a daunting task, but it must be undertaken. If phylogenetic inference is to be a science, we must consider its methods guilty until proven innocent.

Acknowledgments

I wish to thank a reviewer for penetrating comments which were of great assistance. This work was supported by Department of Energy Contract No. EY-76-S-06-2225 5 with the University of Washington.

References

Camin, J. H., and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. *Evolution* 19: 311–326.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 32: 550–570 (also in *Amer. J. Human Genetics* 19: 233–257).

Dayhoff, M. O., and R. V. Eck. 1968. *Atlas of protein sequence and structure 1967–1968*. National Biomedical Research Foundation, Silver Spring, Maryland.

Edwards, A. W. F., and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. *In* Heywood, V. H., and J. McNeill (eds.), *Phenetic and phylogenetic classification*. Systematics Association Publication No. 6, London. pp. 67–76.

Estabrook, G. F. 1972. Cladistic methodology: a discussion of the theoretical basis for the induction of evolutionary history. *Ann. Rev. Ecol. Syst.* 3: 427–456.

Estabrook, G. F., C. S. Johnson, Jr., and F. R. McMorris. 1975. An idealized concept of the true cladistic character. *Math. Biosci.* 23: 263–272.

Estabrook, G. F., and L. Landrum. 1975. A simple test for the possible simultaneous evolutionary divergence of two amino acid positions. *Taxon* 24: 609–613.

Estabrook, G. F., C. S. Johnson, Jr., and F. R. McMorris. 1976a. A mathematical foundation for the analysis of character compatibility. *Math. Biosci.* 29: 181–187.

Estabrook, G. F., C. S. Johnson, Jr., and F. R. McMorris. 1976b. An algebraic analysis of cladistic characters. *Discrete Math.* 16: 141–147.

Farris, J. S. 1969. A successive approximations approach to character weighting. *Syst. Zool.* 18: 374–385.

Farris, J. S. 1970. Methods for computing Wagner trees. *Syst. Zool.* 19: 83–92.

Farris, J. S., A. G. Kluge, and M. J. Eckhardt. 1970. A numerical approach to phylogenetic systematics. *Syst. Zool.* 19: 172–189.

Farris, J. S. 1972. Estimating phylogenetic trees from distance matrices. *Amer. Nat.* 106: 645–668.

Farris, J. S. 1973. On the use of the parsimony criterion for inferring evolutionary trees. *Syst. Zool.* 22: 250–256.

Farris, J. S. 1977a. Phylogenetic analysis under Dollo's Law. *Syst. Zool.* 26: 77–88.

Farris, J. S. 1977b. Some further comments on Le Quesne's methods. *Syst. Zool.* 26: 220–223.

Feller, W. 1957. *An introduction to probability theory and its applications*. Volume I. Second edition. Wiley, New York.

Felsenstein, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.* 22: 240–249.

Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees. *Science* 155: 279–284.

Fitch, W. M. 1973. Aspects of molecular evolution. *Ann. Rev. Gen.* 7: 343–380.

Kluge, A. G., and J. S. Farris. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18: 1–32.

Le Quesne, W. J. 1969. A method of selection of characters in numerical taxonomy. *Syst. Zool.* 18: 201–205.

Le Quesne, W. J. 1972. Further studies based on the uniquely derived character concept. *Syst. Zool.* 21: 281–288.

Morris, F. R. 1977. On the compatibility of binary qualitative taxonomic characters. *Bull. Math. Biol.* 39: 133–138.

21 The Logical Basis of Phylogenetic Analysis

James Farris

Phylogeneticists hold that the study of phylogeny ought to be an empirical science, that putative synapomorphies provide evidence on genealogical relationship, and that (aside possibly from direct observation of descent) those synapomorphies constitute the only available evidence on genealogy. Opponents of phylogenetic systematics maintain variously that genealogies cannot (aside from direct observation) be studied empirically, that synapomorphies are not evidence of kinship because of the possibility of homoplasy, or that raw similarities also provide evidence on genealogy. Most phylogeneticists recognize that inferring genealogy rests on the principle of parsimony—that is, choosing genealogical hypotheses so as to minimize requirements for ad hoc hypotheses of homoplasy. But other criteria as well have been proposed for phylogenetic analysis, and some workers believe that parsimony is unnecessary for that purpose. Others contend that that principle is not truly "parsimonious," or that its application depends crucially on the false supposition that homoplasy is rare in evolution. Authors of all these criticisms have in common the view that phylogenetic systematics as it is now practiced may be dismissed as futile or at best defective. Phylogeneticists must refute that view, but accomplishing that goal seems complicated both by the apparent multiplicity of phylogenetic methods and by the diversity of the objections. I shall show here that the complexity of this problem is superficial. An analysis of parsimony will not only provide a resolution of the objections to that criterion, but will supply as well an understanding of the relationship of genealogical hypotheses to evidence, and with it a means of deciding among methods of phylogenetic inference.

Ad Hoc Hypotheses

I share Popper's disdain for arguing definitions as such, but is it important to make intended meanings clear, and so I shall first dismiss terminological objections to the

From N. Platnick and V. Funk (eds.), *Advances in Cladistics: Proceedings of the Second Meeting of the Willi Hennig Society* (New York: Columbia University Press, 1982), 7–36.

parsimony criterion. These all come to the idea that parsimonious phylogenetic reconstructions are so primarily by misnomer: the word might equally well refer to any of several other qualities. The meanings of "parsimony" would surely take volumes to discuss, but doing so would be quite pointless. Whether the word is used in the same way by all has no bearing on whether the phylogenetic usage names a desirable quality. I shall use the term in the sense I have already mentioned: most parsimonious genealogical hypotheses are those that minimize requirements for ad hoc hypotheses of homoplasy. If minimizing ad hoc hypotheses is not the only connotation of "parsimony" in general usage, at least it is scarcely novel. Both Hennig (1966) and Wiley (1975) have advanced ideas closely related to my usage. Hennig defends phylogenetic analysis on grounds of his auxiliary principle, which states that homology should be presumed in the absence of evidence to the contrary. This amounts to the precept that homoplasy should not be postulated beyond necessity, that is to say, parsimony. Wiley discusses parsimony in a Popperian context, characterizing most parsimonious genealogies as those that are least falsified on available evidence. In his treatment, contradictory character distributions provide putative falsifiers of genealogies. As I shall discuss below, any such falsifier engenders a requirement for an ad hoc hypothesis of homoplasy to defend the genealogy. Wiley's concept is then equivalent to mine.

Cartmill (1981) has effectively objected to that last equivalence, claiming that neither phylogenetic analysis nor parsimony can be scientific in Popper's sense. His argument is superficially technical, but his principal conclusion is in fact based on a terminological confusion, and so I shall discuss his ideas here.

Cartmill cites Gaffney (1979) to the effect that character distributions are falsifiers of genealogical hypotheses, and that it is possible that every conceivable genealogy will be falsified at least once. From the first of these admissions he "deduces" that Gaffney must have relied on the "theorem" that any genealogy contradicted by a character distribution is false. Cartmill then reasons: Some genealogy must be true. "Gaffney's" theorem, together with a falsifier for every genealogy, implies that every genealogy is false. Therefore, Gaffney's claim that character distributions are falsifiers is false.

Cartmill's argument rests directly and entirely on a misrepresentation of the Popperian meaning of "falsifier": a test statement that, if true, allows a hypothesis to be rejected. There is a great difference between "falsify" in Popper's sense and "prove false." The relationship between a theory and its falsifiers is purely logical; Popper never claimed that proof of falsity could literally be achieved empirically. "Observing" a falsifier of a theory does not prove that the theory is false; it simply implies that either the theory or the observation is erroneous. It is then seen that the only implication that can be derived from falsification of every genealogy is that some of the falsifiers are errors—homoplasies. It is thus seen as well that Cartmill's "syllogism" is nothing other than an equivocation.

So much for the claim that characters cannot be Popperian falsifiers, but is phylogenetic parsimony Popperian? Cartmill admits that phylogeneticists hold that the least falsified genealogy is to be preferred. The reason for this preference is that each falsifier of any accepted genealogy imposes a requirement for an ad hoc hypothesis to dispose of the falsifier. According to Popper—as Cartmill also cites—ad hoc hypotheses must be minimized in scientific investigation. Cartmill never attempts to argue that conflicts between characters and genealogies do not require hypotheses of homoplasy, and so none of his claims can serve to question the connection between parsimony and Popper's ideas.

Parsimony and Synapomorphy

The objection that parsimony requires rarity of homoplasy in evolution is usually taken to be just that: a criticism of parsimony. It might seem that the problem posed by that objection could be avoided simply by using some other criterion for phylogenetic analysis. Some quite nonphylogenetic proposals, such as grouping according to raw similarity, have been made along those lines, and I shall discuss those eventually. Of more immediate interest is the question whether grouping by putative synapomorphy can do without the parsimony criterion.

Watrous and Wheeler (1981) suggest that parsimony is needed only when characters conflict, with the implication that a set of congruent characters can be analyzed without invoking ad hoc hypotheses of homoplasy. A similar idea would appear to underlie advocacy by Estabrook and others (reviewed by Farris and Kluge 1979) of techniques (''clique'' methods) that ''resolve'' character conflicts by discarding as many characters as necessary so that those surviving (the clique) are mutually congruent. The surviving characters are then used to construct a tree. Proponents of such methods maintain that the tree so arrived at rests on a basis different from parsimony.

The character selection process itself may well have a distinctive premise, a possibility that I shall discuss below. To claim that the interpretation of the characters selected rests on a basis other than parsimony, however, seems not to be defensible. The tree constructed from a suite of congruent characters by a clique method is chosen to avoid homoplasy in any of those characters, the possibility of doing so being assured by the selection. (Selection aside, Watrous and Wheeler proceed likewise.) It seems accurate, then, to describe that construction as minimizing requirements for ad hoc hypotheses of homoplasy for the characters within the congruent suite, but, more particularly, there seems to be no other sensible rationale for the construction. No one seems to have suggested any such principle, aside from the obvious: that if the characters were free of homoplasy (were ''true'' as it is often put), then the tree would follow. But the characters comprising a congruent suite are hardly observed to be free of homoplasy. At the most it might be said that the selected characters seem to suggest no genealogy other than the obvious one.

Of course it is what data suggest, or how they do it, that is at issue. If a suite of congruent characters is interpreted by avoiding unnecessary postulates of homoplasy, then the interpretation embodies parsimony. But the only apparent motivation for concentrating just on congruent characters is to avoid reliance on parsimony. That avoidance would seem sensible only on the supposition that parsimony is ill founded, and the only apparent reason for that supposition is the charge that parsimony depends crucially on unrealistic assumptions about nature. If that charge means anything at all, it must mean that taking conditions of nature realistically into account would lead to preference for a less parsimonious arrangement over a more parsimonious one. But if that charge were correct, then it would be—to say the least—less than obvious why the implications of those natural conditions would be expected to change simply because any characters incongruent with those chosen had been ignored.

If avoiding ad hoc hypotheses of homoplasy is unjustified, then neither Watrous and Wheeler nor clique advocates are entitled to the inferences on phylogeny that they draw, but the significance of parsimony for Hennigian methods is much more general than that. Watrous and Wheeler probably thought that they had no need for anything so questionable as parsimony, because they were simply applying Hennig's well-established principle of grouping according to synapomorphy. Just how did that principle come to be well established? It is usually explained by taking note of the logical relationship between monophyletic groups and true synapomorphies, but that leaves open the question of how genealogies are related to observed features. It might well be questioned whether the logical construct can legitimately be extended into a principle to guide interpretation of available characters. That question has in fact often been raised, and almost always in the form of the suggestion that putative synapomorphies are not evidence of kinship because they might well be homoplasies. Hennig's (1966) own reply to that objection was his auxiliary principle, which, as I have already observed, is a formulation of the parsimony criterion.

Hennig's defense of the synapomorphy principle by recourse to parsimony is not accidental, but necessary. The analytic relationship of correct synapomorphies to phylogeny is just that a property that evolved once and is never lost must characterize a monophyletic group. Synapomorphies are converted into a genealogy, that is, by identifying the tree that allows a unique origin for each derived condition. A phylogeny based on observed features is parsimonious to the degree that it avoids requirements for homoplasies—multiple origins of like features. Secondary plesiomorphies aside, a plesiomorphic trait will already have a single origin at the root of the putative tree, so that the effect of parsimony is precisely to provide unique derivations wherever possible. (Secondary plesiomorphies, being a kind of apomorphy, are treated likewise.) Grouping by synapomorphy would thus have to behave like parsimony, but further, the latter applies to actual traits, whereas the logic of true synapomorphies does not. Superficially, the use of the synapomorphy principle in phylogenetic inference seems to be just a consequence of the logical connection between true synapomorphies and genealogies, but it cannot be just that, as the condition of that logic—that the traits are indeed synapomorphies—need not be met. Grouping by putative synapomorphy is instead a consequence of the parsimony criterion.

Abundance of Homoplasy

There are two main varieties of the position that use of the parsimony criterion depends crucially on the supposition that homoplasy is rare in evolution. In the first, the observation that requirements for homoplasy are minimized is taken as prima facie evidence that the supposition is needed. In the second, the claim is advanced in conjunction with some more elaborate, often statistical, argument. The conclusion from the first kind of reasoning is quite general, while that from the second is necessarily limited by the premises of the argument employed. If the first kind of criticism were correct, there would be little point to considering arguments of the second sort. I shall thus first point out why the first type of objection rests on a fallacy.

To evaluate the claim that an inference procedure that minimizes something must ipso facto presuppose that the quality minimized is rare, it is useful to consider a common application of statistics. In normal regression analysis, a regression line is calculated from a sample of points so as to minimize residual variation around the line, and the residual variation is then used to estimate the parametric residual variance. Plainly the choice of line has the effect of minimizing the estimate of the residual variance, but one rarely hears this procedure criticized as presupposing that the parametric residual variance is small. Indeed, it is known from normal statistical theory that the least squares line is the best point estimate of the parametric regression line, whether the residual variance is small or not. The argument that the parsimony criterion must presume rarity of homoplasy just because it minimizes required homoplasy is thus at best incomplete. That reasoning presumes a general connection between minimization and supposition of minimality, but it is now plain that no such general connection exists. Any successful criticism of phylogenetic parsimony would have to include more specific premises.

The same conclusions can readily be reached in a specifically phylogenetic context. Suppose that for three terminal taxa A, B, C, there are ten putative synapomorphies of A + B and one putative apomorphy shared by B and C. We assume for simplicity of discussion that the characters are independent and all of equal weight, and that attempts to find evidence to support changes in the data have already failed. Parsimony then leads to the preference for ((A, B), C) over alternative groupings. We will be interested in whether abundance of homoplasy leads to preference for some other grouping. If it does not, then the claim that parsimony presupposes rarity of homoplasy is at best not generally true.

It is plain that the grouping ((B, C), A) is genealogically correct if the one B + C character is, in fact, a synapomorphy, and that ((A, B), C) is instead correct if the A + B characters are synapomorphies. Truth of the latter grouping does not require, however, that all ten of the putative synapomorphies of A + B be accurate homologies. If just one of those characters were truly a synapomorphy while all the other characters in the data were in fact parallelisms, the genealogy would necessarily be ((A, B), C). That A and B share a common ancestor unique to them, in other words, does not logically require that every feature shared by A and B was inherited from that ancestor. In the extreme, if all the characters were parallelisms, this would not imply that ((A, B), C) is genealogically false. Under those circumstances the data would simply leave the question of the truth of that (or any other) grouping entirely open.

The relationship between characters and genealogies thus shows a kind of asymmetry. Genealogy ((A, B), C) requires that the B + C character be homoplasious, but requires nothing at all concerning the A + B characters. The genealogy can be true whether the conforming characters are homoplasious or not. One kind of objection to phylogenetic parsimony runs that ad hoc hypotheses are indeed to be minimized, but this does not mean minimizing homoplasies, because a genealogy also requires ad hoc hypotheses of homology concerning the characters that conform to it. It is seen that such is not the case. Only characters conflicting with a genealogy lead to requirements for ad hoc hypotheses, and so the only ad hoc hypotheses needed to defend a genealogy are hypotheses of homoplasy.

The sensitivity of inference by parsimony to rarity of homoplasy is readily deduced from these observations. If homoplasy is indeed rare, it is quite likely with these characters that ((A, B), C) is the correct genealogy. In order for that grouping to be false, it would be required at least that all ten of the A + B characters be homoplasious. As these characters are supposed to be independent, the coincidental occurrence of homoplasy in all ten should be quite unlikely. Suppose, then, that homoplasy is so abundant that only one of the characters escapes its effects. That one character might equally well be any of the eleven in the data, and if it is any one of the ten A + B characters the parsimonious grouping is correct. That grouping is thus a much better bet than is ((B, C), A). At the extreme, as has already been seen, if homoplasy is universal, the characters imply nothing about the genealogy. In that case the parsimonious grouping is no better founded than is any other, but then neither is it any worse founded.

It seems that no degree of abundance of homoplasy is by itself sufficient to defend choice of a less parsimonious genealogy over a more parsimonious one. That abundance can diminish only the strength of preference for the parsimonious arrangement; it can never shift the preference to a different scheme. In this the relationship of abundance of homoplasy to choice of genealogical hypothesis is quite like that between residual variance and choice of regression line. Large residual variance expands the confidence interval about the line, or weakens the degree to which the least squares line is to be preferred over nearby lines, but it cannot by itself lead to selection of some other line that fits the data even worse.

Stochastic Models

The supposition of abundance of homoplasy by itself offers no grounds for preferring unparsimonious arrangements, but it is easy enough to arrive at that preference by resorting to other premises. Felsenstein (1973, 1978, 1979) objects to parsimony on statistical grounds. He suggests (as others have) that genealogies ought to be inferred by statistical estimation procedures. In his approach he devises stochastic models of evolution, then applies the principle of maximum likelihood, choosing the genealogy that would assign highest probability to the observed data if the model were true. With the models that he has investigated, it develops that the maximum likelihood tree is most parsimonious when rates of change of characters are very small, under which circumstances the models would also predict very little homoplasy. He concludes from this that parsimony requires rarity of homoplasy. In his 1978 paper he discusses a model according to which both parsimony and clique methods would be certain to yield an incorrect genealogy if a large enough random sample of independent characters were obtained. He contends that maximum likelihood estimation under the same conditions would yield the correct tree, as that estimate possesses the statistical property termed consistency. That last is the logical property that if an indefinitely great number of independent characters were sampled at random from the distribution specified by the model, then the estimate would converge to the parameter of the model, the hypothetical true tree.

Felsenstein does not try to defend his models as realistic. His attitude on their purpose seems to be instead that "if a method behaves poorly in this simple model framework, this calls into question its use on real characters" (1981, p. 184), or perhaps (1978, p. 409), "If phylogenetic inference is to be a science, we must consider its methods guilty until proven innocent." The first is preposterous except on the supposition that reasoning from false premises cannot lead to false conclusions. As for the second: to the extent that these models are intended seriously, they comprise empirical claims on evolution. If science required proof concerning empirical claims in order to draw conclusions, no kind of science would be possible.

Felsenstein nonetheless apparently believed that he had demonstrated that practical application of parsimony requires rarity of homoplasy, but in fact such is hardly the case. The dependence of parsimony on rarity of homoplasy is in Felsenstein's analysis a consequence of his models. These models, as he is well aware, comprise "strong assumptions about the biological situation" (1978, p. 403). If those assumptions do not apply to real cases, then so far as Felsenstein can show, the criticism of parsimony need not apply to real cases either. But, again, Felsenstein does not maintain that the assumptions of his models are realistic. He has not shown that abundance of homoplasy implies preference for unparsimonious genealogies. Instead he has shown at most that if homoplasy were abundant, and if in addition the conditions of his models prevailed in nature, then one should prefer unparsimonious schemes. We have already seen that abundance of homoplasy by itself does not justify departure from parsimony. If Felsenstein's argument offers any reason for that departure, then that reason would have to rely on the supposition that his models apply to nature. An ironic result indeed. The original criticism of parsimony was that it required an unrealistic assumption about nature. It now seems instead that unparsimonious methods require such assumptions, whereas parsimony does not.

Felsenstein's arguments from consistency and maximum likelihood have a related drawback. Consistency is a logical relationship between an estimation method and a probability model. In the hypothetical case imagined by Felsenstein, his method would have obtained the right answer, but whether the method would work in practice depends on whether the model is accurate. If it is not, then the consistency of the estimator under the model implies nothing about the accuracy of the inferred tree. The status of a procedure as a maximum likelihood estimator is also bound to the probability model. If the model is false, the ability of a procedure to find the most likely tree under the model implies nothing of how likely the chosen tree might actually be. Likewise the conclusion that parsimony would arrive at the wrong tree depends on the model, and so the hypothetical analysis implies little about the practical accuracy of parsimony.

One might say, of course, that the model illustrates a potential weakness of parsimony: That criterion will fail if the conditions of the model should happen to be met. And how are we to know that this will not happen? This seems in fact to be the intended substance of Felsenstein's remarks. While admitting that his premises are unrealistic, he rejects realism as a criticism of his attack on parsimony, claiming that an objection based on realism "amounts to a confession of ignorance rather than a validation of the inference method in question" (1978, p. 408). A derivation that implies nothing about reality is not much of an improvement on ignorance, of course. To the extent that Felsenstein has a point, then, it seems to be just that parsimony is invalid because we cannot be certain that it will not lead to errors of inference. But there is nothing distinctive about Felsenstein's model in that regard. One may always concoct fantastic circumstances under which scientific conclusions might prove incorrect. It is hardly necessary to resort to mathematical manipulations in order to produce such fears. One need only imagine that his characters have evolved in just the right way to lead him to a false conclusion. Or, with Descartes, that his perceptions and reason have been systematically and maliciously distorted by a demon. None of these possibilities can be disproved, but it hardly matters. There is likewise no reason for accepting any of them, and so collectively they amount to no more than the abstract possibility that a conclusion might be wrong. No phylogeneticist—or any scientist—would dispute that anyway, and so such "objections" are entirely empty. That thinking provides no means of improving either conclusions or methods, but instead offers, if anything, a rejection of all conclusions that cannot be established with certainty. If that attitude were taken seriously, no scientific conclusions whatever could be drawn.

Explanatory Power

A number of authors, myself among them (Farris, 1973, 1977, 1978), have used statistical models to defend parsimony, using, of course, different models from Felsenstein's. Felsenstein has objected to such derivations on grounds of statistical consistency: as before, parsimonious reconstructions are not consistent under his models. That is no more than an equivocation, as the models differ, and consistency is a relationship between method and model. But I do not mean by this to defend those favorable derivations, for my own models, if perhaps not quite so fantastic as Felsenstein's, are nonetheless like the latter in comprising uncorroborated (and no doubt false) claims on evolution. If reasoning from unsubstantiated suppositions cannot legitimately question parsimony, then neither can it properly bolster that criterion. The modeling approach to phylogenetic inference was wrong from the start, for it rests on the idea that to study phylogeny at all, one must first know in great detail how evolution has proceeded. That cannot very well be the way in which scientific knowledge is obtained. What we know of evolution must have been learned by other means. Those means, I suggest, can be no other than that phylogenetic theories are chosen, just as any scientific theory, for their ability to explain available observation. I shall thus concentrate on evaluating proposed methods of phylogenetic analysis on that basis.

That ad hoc hypotheses are to be avoided whenever possible in scientific investigation is, so far as I am aware, not seriously controversial. That course is explicitly recommended by Popper, For example. No one seems inclined to maintain that ad hoc hypotheses are desirable in themselves; at most they are by-products of conclusions held worthy on other grounds. Nonetheless, I suspect that much of the criticism of the phylogenetic parsimony criterion arises from a failure to appreciate the reasons why ad hoc hypotheses must be avoided. Avoiding them is no less than essential to science itself. Science requires that choice among theories be decided by evidence, and since the effect of an ad hoc hypothesis is precisely to dispose of an observation freely, there could be no effective connection between theory and observation, and the concept of evidence would be meaningless. The requirement that a hypothesis of kinship minimize ad hoc hypotheses of homoplasy is thus no more escapable than the general requirement that any theory should conform to observation; indeed, the one derives from the other.

There are a number of properties commonly held to characterize a theory that gives a satisfactory account of observation. The theory must first of all provide a description of what is known, else it would serve little purpose. As Sober (1975) puts it, theories serve to make experience redundant. But not all descriptions are equally useful. Good theories describe in terms of a coherent framework, so that experience becomes comprehensible; in short they are explanatory. Explanations in turn may be judged on their ability to cover observations with few boundary conditions, that is, with little extrinsic information. Sober has characterized theories satisfying this goal as most informative, or simplest. All these criteria are interrelated in the case of phylogenetic inference, so that they effectively yield a single criterion of analysis. These connections have already been recognized, and I shall summarize them only briefly.

I have elsewhere (Farris 1979, 1980, 1982) already analyzed the descriptive power of hierarchic schemes. I showed that most parsimonious classifications are descriptively most informative in that they allow character data to be summarized as efficiently as possible. That conclusion has aroused some opposition, as syncretistic taxonomists had been inclined to suppose that grouping according to (possibly weighted) raw similarity gave hierarchies of greatest descriptive power. There seems to be no reason for taking that view seriously, however, as no attempt has been made to derive clustering by raw similarity from the aim of effective description of character information.

In my treatment I found that a hierarchic classification provides an informative or efficient description of the distribution of a feature to the degree that the feature need occur in the diagnoses of few taxa. The utility of efficient descriptions is precisely that they minimize redundancy. As I have observed before (particularly Farris, 1980), the presence of a feature in the diagnosis of a taxon corresponds to the evolutionary interpretation that the feature arose in the stem species of that taxon. There is thus a direct equivalence between the descriptive utility of a phylogenetic taxon and the genealogical explanation of the common possession of features by members of that group. Sober (1975) has stressed the importance of informativeness of theories, and has developed a characterization of informativeness in terms of simplicity. It is no surprise to find that simplicity is related to parsimony, and Beatty and Fink (1979) have lucidly discussed the connection in terms of Sober's ideas. Sober (1982) has likewise concluded that phylogenetic parsimony corresponds to simplicity (efficiency, informativeness) of explanation.

In choosing among theories of relationship on the basis of explanatory power, we wish naturally to identify the genealogy that explains as much of available observation as possible. In general, deciding the relative explanatory power of competing theories can be a complex task, but it is simplified in the present case by the fact that genealogies provide only a single kind of explanation. A genealogy does not explain by itself why one group acquires a new feature while its sister group retains the ancestral trait, nor does it offer any explanation of why seemingly identical features arise independently in distantly related lineages. (Either sort of phenomenon might, of course, be explained by a more complex evolutionary theory.) A genealogy is able to explain observed points of similarity among organisms just when it can account for them as identical by virtue of inheritance from a common ancestor. Any feature shared by organisms is so by reason of common descent, or else it is a homoplasy. The explanatory power of a genealogy is consequently measured by the degree to which it can avoid postulating homoplasies.

It is necessary in applying that last observation to distinguish between homoplasies postulated by the genealogy and those concluded for other reasons. A structure common to two organisms might be thought to be a homoplasy on grounds extrinsic to the genealogy. Such a conclusion would amount to specifying that the structure is not a point of heritable similarity. A genealogy would not explain such a similarity, but that would be no grounds for criticizing the genealogy. Rather, the extrinsic conclusion would make the feature irrelevant to evaluating genealogies by effectively stipulating that there is nothing to be explained. The same would hold true for any trait that is known not to be heritable, such as purely phyenotypic variations. The explanatory power of a genealogy is consequently diminished only when the hypothesis of kinship requires ad hoc hypotheses of homoplasy.

By analogy with the abundance of homoplasy argument, it might be objected that seeking a genealogical explanation of similarities is pointless, inasmuch as most similarities are likely to be homoplasies anyway. If homoplasy were universal, that point might well hold. It seems unlikely, however, that homoplasy is universal. It is seldom maintained that segmented appendages have arisen independently in each species of insect. Universality of homoplasy would imply in the extreme that organisms do not generally resemble their parents, a proposition that seems at best contrary to experience. That the character distributions of organisms generally correspond to a hierarchic pattern, furthermore, seems comprehensible only on the view that the character patterns reflect a hierarchy of inheritance. Indeed, the recognized organic hierarchy was one of the chief line of evidence for Darwin's theory of descent with modification. The idea that homoplasy is abundant is not usually intended in such extreme form, of course. Usually it is meant just to suggest that there is room for doubt concerning whether a shared feature is a homology or a homoplasy. Under those circumstances, however, genealogies retain explanatory power. More to the point, the explanatory power of alternative genealogies is still related to their requirements for homoplasies. Suppose that one genealogy can explain a particular point of similarity in terms of inheritance, while a second hypothesis of kinship cannot do so. If that point of similarity is, in fact, a homoplasy, the similarity is irrelevant to evaluating genealogical hypotheses, as has already been seen. If the similarity is, instead, a homology, then only the first genealogy can explain it. If there is any chance that the similarity is homologous, the first genealogy is to be preferred.

There is nothing unusual in the relationship of genealogical hypotheses to characters: scientific theories are generally chosen to conform to data. But it is seldom possible to guarantee that observations are free of errors, and it is no criticism of a theory if it turns out that some of the observations that conform to it are susceptible to error. If a theory does not conform to some observation, however, then the mere suspicion that the observation might be erroneous is not logically adequate to save the theory. Instead the data must be dismissed outright by recourse to an ad hoc hypothesis. Establishing that an observation is erroneous, on the other hand, simply makes it irrelevant to evaluating the theory. The relationship between the explanatory power of genealogies and their requirements for ad hoc hypotheses is likewise characteristic of theories in general. Any observation relevant to evaluating a theory will either conform—and so be explained—or fail to do so, in which case an ad hoc hypothesis is needed to defend the theory. It is generally true that a theory explains relevant observations to the degree that it can be defended against them without recourse to ad hoc hypotheses.

Independence of Hypotheses

Identifying the more explanatory of two alternative hypotheses of kinship is accomplished by finding the total of ad hoc hypotheses of homoplasy required by each. Reckoning those totals will generally involve summing over both separate characters and over observed similarities within characters. Only required ad hoc hypotheses diminish the explanatory power of a putative genealogy. It is thus important to ensure that the homoplasies combined in such totals are logically independent, since otherwise their number need not reflect required ad hoc hypotheses. If two characters were logically or functionally related so that homoplasy in one would imply homoplasy in the other, then homoplasy in both would be implied by a single ad hoc hypothesis. The "other" homoplasy does not require a further hypothesis, as it is subsumed by the relationship between the characters. This is the principle underlying such common observations as that only independent lines of evidence should be used in evaluating genealogies, and that there is no point to using both number of tarsal segments and twice that number as characters. Phylogeneticists seldom attempt to use logically related characters as separate sources of evidence (although an example of this mistake is discussed by Riggins and Farris 1983), and so it seems unnecessary to discuss this point further here.

A different sort of interdependence among homoplasies may arise in considering similarities within a single character. Suppose that 20 of the terminal taxa considered show a feature X, and that a putative genealogy distributes these taxa into two distantly related groups A and B of 10 terminals each. There are 100 distinct two-taxon comparisons of members of A with members of B, and each of those similarities in X considered in isolation comprises a (pairwise) homoplasy. Those homoplasies do not constitute independent required hypotheses, however. The genealogy does not require that similarities in X within either group be homoplasies; it is consistent with identity by descent of X within each group. If X is identical by descent in any two members of A, and also in any two members of B, then the A-B similarities are all homoplasies if any one of them is. The genealogy thus requires but a single ad hoc hypothesis of homoplasy. Of course the numbers in the groups do not matter; the same conclusion would follow if they were 15 and 5, or 19 and 1.

Similar reasoning can be extended to more complex examples, but the problem can be analyzed more simply. If a genealogy is consistent with a single origin of a feature, then it can explain all similarities in that feature as identical by descent. A point of similarity in a feature is then required to be a homoplasy only when the feature is required to originate more than once on the genealogy. A hypothesis of homoplasy logically independent of others is thus required precisely when a genealogy requires an additional origin of a feature. The number of logically independent ad hoc hypotheses of homoplasy in a feature required by a genealogy is then just one less than the number of times the feature is required to originate independently. (The lack of a structure might, of course, be a feature; ''origins'' should be interpreted broadly, to include losses.)

Lengths

That last observation reduces to the rule that genealogies with greatest explanatory power are just those that minimize the (possibly weighted) total of required independent origins of known features. There is another way of putting that characterization, in terms of length. Each required origin of a feature can be assigned (although not necessarily uniquely) to a particular branch of a putative tree, and the weighted total of the origins in a branch can be regarded as that branch's length. If such lengths are summed over branches of the tree, the result is the total of required origins, or the length of the tree. Early work on automatic techniques of parsimony analysis (particularly the Wagner method formulation of Kluge and Farris 1969) used the length conception of parsimony. That formulation has turned out to be technically very useful and has facilitated considerable progress in methods of analysis. (Basic principles are described by Farris 1970; for some applications of greatly improved procedures see Mickevich 1978, 1980; Schuh and Farris 1981; Mickevich and Farris 1981). Nonetheless, its use was in a way unfortunate, for the length terminology has probably caused more misunderstanding than has any other single aspect of parsimony methods.

The length measure used by Kluge and Farris is coincidentally a familiar mathematical measure of distance in abstract spaces, the Manhattan metric. Once ideas have been reduced to formulas, it is easy to forget where the formulas came from, and to devise new methods with no logical basis simply by modifying formulas directly. Phylogenetic reconstructions typically infer the features of hypothesized ancestors, so that the length of a branch lying between two nodes of a tree can be regarded as the distance between two points in the space of possible combinations of features. If one notes only that length is to be minimized, then he might just as well seek trees of minimum Euclidean length, or indeed of minimum length in any of the other uncountably many possible measures of abstract distance. But even that does not exhaust the possibilities. Numerical values—lengths of a sort—can be calculated for branches without regard to the possible features of nodes, by fitting the tree directly to a matrix of pairwise distances between terminal taxa (such methods are reviewed by Farris 1981). Such trees, too, might be selected to minimize ''length,'' and this might be done for any of the huge number of ways of arriving at a matrix of pairwise distances.

The analogy through length has allowed methods such as this to become confused with parsimony analysis, and that confusion has played a role in specious criticisms of phylogenetic methods. Felsenstein (1981)—one of the main proponents of the idea that ''parsimony'' might mean almost anything—for example, attributes ''parsimony'' to one such method that had been used by Edwards and Cavalli-Sforza (1964). Rohlf and Sokal (1981) used a procedure for fitting branch lengths to a distance matrix to analyze the data of Schuh and Polhemus (1980), then criticized the parsimony analysis of the latter authors on the grounds that the distance-fitted tree is ''shorter.'' As Schuh and Farris (1981) pointed out, the length that Rohlf and Sokal attribute to their tree is quite meaningless, inasmuch as it is smaller than the number of origins of features required to account for the data (for related discussion see Farris 1981).

That will serve as a general commentary on this class of methods, which are too numerous in their possibilities to discuss here individually. The lengths arrived at by such calculations are generally incapable of any interpretation in terms of origins of features, and the evaluation of trees by such lengths consequently has nothing to do with the phylogenetic parsimony criterion. What is worse, the trees produced by these methods frequently differ in their grouping from parsimonious genealogies, and to that extent the use of these procedures amounts to throwing away explanatory power.

Pairwise Homoplasies

The situation is somewhat different with some types of comparative data, such as matrices of immunological distance, in which no characters are directly observed. I have emphasized before (Farris 1981) that the parsimony criterion cannot be directly applied to such cases, and so I shall not consider them here. (The paper just cited offers other bases for evaluating methods of distance analysis). Some analogies with distance analytic methods, however, can be related to the present discussion.

In fitting a tree to distances, branch lengths are used to determine a matrix of pairwise tree-derived distances between terminal taxa. The derived distance between a pair is just the sum of the lengths of the branches that lie on the path connecting the two taxa on the tree. Evaluation of the fit of the tree to observed distances is based on conformity of the derived to the observed distance values, this being measured by, say, the sum of the unsigned differences between the corresponding elements of the matrices (other measures are discussed by Farris 1981). Parsimony analyses can also be used to produce derived distances, the patristic differences of Farris (1967). I had earlier (Farris 1967) termed the departure of patristic from observed differences (pairwise) homoplasies, and from this, as well as by analogy with distance analytic procedures, one might be tempted to evaluate genealogies according to the total of those pairwise homoplasies (such a suggestion has been made by D. Swofford). The drawback of doing so is already clear from earlier discussion: the pairwise homoplasies are not independent.

A more extreme problem of interpretation of pairwise homoplasies arises in some methods for analyzing electrophoretic data. Suppose that each of three terminal taxa A, B, and C is fixed for a different allele at some locus, and that these taxa are related through an unresolved tree with common ancestor X. There are a number of ways of calculating distance between gene frequency distributions (see Farris 1981). To fix ideas, suppose that the Manhattan distance on frequencies is used. The distance between any two terminals is then 2. The ancestor X might plausibly be assigned frequency 1/3 for each observed allele. In that case the distance between X and any terminal is 4/3, the patristic difference between any two terminals is 8/3, and the corresponding pairwise homoplasy is 2/3. That implies that there is homoplasious similarity between any two terminals, but the conclusion is nonsense, inasmuch as there is no similarity between them at all. The three terminals simply have three entirely different conditions of the locus.

The details of that example depend on how gene frequencies are assigned to X, but no assignment can bring all the pairwise "homoplasies" to 0 simultaneously. In part this observation reflects the difficulties inherent in any attempt to utilize distances between gene frequency distributions as evidence in phylogenetic analysis (discussed in further detail by Farris 1981). Of greater interest for present purposes is what the example reveals about alleles as characters. The algebraic reason for the existence of those spurious homoplasies is that the distance coefficient treats shared 0 frequencies as points of similarity. Two taxa are assessed as similar in that both lack some allele, whereas in fact they simply possess different alleles. It is clear that those shared absences offer no independent assessment of the resemblances among the taxa, as the 0 frequency in any one allele is a necessary consequence of the fixation of any other. This problem then results from treating dependent quantities as if they were independent.

That difficulty is not limited to analysis of frequency data. Mickevich and Johnson (1976) introduced a method in which frequencies are transformed into a two-state coding: any frequency above a cutting point is coded as 1 (presence), any other as 0 (absence). The standard Wagner method is then used to find a tree minimizing required origins of states for the coded data. This procedure obviates many of the difficulties of analyzing frequency data through distance measures, but it still suffers from dependence of variables. Fixation of one allele will necessarily control the codes of others at the same locus. The number of state origins for the coding thus need not indicate the number of independent hypotheses of homoplasy for a genealogy, and this procedure should not then be regarded as a parsimony method. The problem of interdependence can, however, be avoided by choosing a better means of coding. Mickevich and Mitter (1981) and Mitter and Mickevich (1982) have made impressive progress in developing coding methods for analyzing electrophoretic data.

Covering Assumptions

Inasmuch as the aim is to minimize ad hoc hypotheses, it might seem that one could do better still by posing single hypotheses to cover several separate cases of homoplasy. Any putative genealogy might on that reasoning be defended against any character by concocting some premise to imply that all similarities in that character are homoplasies—or against any set of characters by dismissing evidence in general. I shall refer to such mass dismissals of evidence as covering assumptions.

The danger of using covering assumptions can be readily seen through a consideration of usual scientific practice. Suppose that an experiment is designed to evaluate a theory on the basis of readings from several instruments, and that some of the readings do not conform to the theory. If the nonconforming observations are only a few of the many readings made, the theory may seem to offer a generally satisfactory explanation; it is less so to the degree that such observations are abundant. Even then attempts may be made to salvage the theory. If the offending readings all come from the same instrument and so are logically related, they might be dismissed through the premise that the instrument is defective. (If it is found to be defective, so much the better.) But if no connection can be found among the nonconforming readings, the claim that they are coincidentally erroneous would have to be viewed with suspicion. Even the best theories seldom conform to every relevant observation, and so theories are well founded to the degree that nonconforming observations are rare. If contradictory observations could be dismissed as uninformative without regard to their abundance, the link between theory and observation would be tenuous at best.

Of course this is generally recognized, and attempts to defend theories by doing away with entire masses of evidence are typically rationalized by postulating mechanisms to account for what would otherwise be coincidental departures of observation from expectation. The legitimacy of that procedure depends crucially on validity of the postulates used. If the postulated mechanisms can themselves be corroborated by other sources of evidence, their use to defend the original theory is justified, and indeed they constitute improvements or extensions of the original theory. But if such mechanisms cannot be defended on extrinsic grounds, then they amount to no more than ad hoc excuses for the failure of the theory. Logically (albeit not rhetorically), they have no more force than the flat assertion that all nonconforming observations must be erroneous because the theory is true. Covering assumptions must be forbidden in scientific study, not only because they are ad hoc, but more particularly because they provide false license to dismiss any amount of evidence whatever.

The reason for prohibiting covering assumptions might be encapsulated by the observation that their use would allow theories to be chosen without regard to explanatory power. This effect can be seen directly in phylogenetic application. If twenty terminals share a particular feature, a genealogy consistent with a single origin of that feature explains those similarities fully. A hypothesis of kinship that broke those terminals into two separate groups of ten would not explain all the similarities among taxa, but it would still explain similarities within those groups. A tree that divided the same terminals into four separate groups of five would explain still less of observed similarities, but would still retain some explanatory power, while a scheme that required twenty separate origins would leave the observed similarities entirely unexplained. Some ad hoc rationale might be used to combine three or nineteen logically independent hypotheses of homoplasy into a "single" hypothesis. The possibility of that combination might be interpreted to mean that all these genealogies but the first conform equally well with observation. If such a course were followed, the differences in explanatory power among the last three hypotheses of kinship would play no role in choosing among them.

Almost any method that led to departure from parsimony might be suspected of involving a covering assumption. One might presume that the various length measures discussed before arise from some underlying premises that would amount to assumptions about the nature of evolution. But inasmuch as those premises, supposing that they exist, have never been made explicit, there is no real possibility of evaluating them as theories, and it is more immediately useful to view those methods as resulting simply from misunderstanding of the explanatory relationship between genealogies and characters.

Felsenstein's maximum likelihood methods offer fine examples of reliance on covering assumptions. The stochastic models would—if they were realistic—explain why seemingly independent characters would depart systematically from a parsimonious arrangement, hence would justify preference for unparsimonious schemes. Likewise, that neither Felsenstein nor anyone else maintains that those models can be corroborated makes it clear that in practice that justification would be entirely specious. But most of these methods have never been advocated for practical application, anyway. Felsenstein's own recent efforts center on likelihood interpretations of procedures that had already been advocated on other grounds, as I shall discuss later. It is of more practical interest to analyze methods that have been proposed more or less seriously.

Irreversibility

Some techniques have been proposed as restricted "parsimony" methods. In these the number of origins of features is minimized, subject to the condition that some kinds of origins be rare or forbidden. Commonly, methods of this sort embody some version of the idea that evolution of individual characters is irreversible. In the method of Camin and Sokal (1965) secondary plesiomorphies are supposed not to occur, and so are excluded from reconstruction, the tree being chosen to minimize parallelisms. In the "Dollo" method of Farris (1977), origins of structures are supposed to be unique—structures once lost cannot be regained—and the tree is chosen to minimize secondary plesiomorphies.

Any of these methods might yield the same genealogy as would be obtained without the restriction, but none of them needs to do so, and in general applying the restriction will increase the number of hypotheses of homoplasy needed to defend the conclusion. Since there is no particular limit to that increase, using the restriction amounts simply to dismissing en mass any evidence that might otherwise seem to vitiate the conclusion. The motivation for doing so seems often to be more a matter of technical convenience than of conviction of the propriety of the restriction. That seems particularly to apply to the Camin–Sokal method, as it was one of the earliest techniques to be implemented as a computer algorithm. The reason for my own (Farris 1977) development of the Dollo method likewise had little to do with the realism of the assumption. That study was intended primarily to show logical flaws in Le Quesne's (1974) earlier attempt to analyze the same problem.

In a serious study, defending conclusions that depended crucially on use of a restricted method would require defending the restriction itself. I would not claim that the supposition of irreversible character evolution could never be supported by extrinsic evidence. I would suggest, however, that what acceptance that idea has gained has been based mostly on generalizations derived from hypotheses of kinship. The common notion that evolution generally proceeds from many, similar, parts to fewer, differentiated, parts, for example, seems to have been arrived at by induction from putative lineages. If the putative phylogenies used to draw such conclusions had been arrived at by presupposing irreversibility, then the conclusion would have no legitimate empirical support. If the idea of irreversibility is supported at all, then, it must have been derived from analyses that did not depend crucially on its truth. The evidence for a directed evolutionary trend, then, would be that the postulated trend conforms to a pattern of kinship that is in turn supported by other evidence—that is, that itself conforms to other characters. If it were known that evolution is irreversible, application of that knowledge might lead to genealogical inferences that otherwise might seem unparsimonious. But in fact no such thing is known, and the attempt to apply an empirically supported claim of irreversibility as a criticism of parsimony leads to a peculiar difficulty. Any body of characters might be made to appear consistent with the postulate of irreversibility. It is always technically possible to construct a tree so that all homoplasies take the form of parallelisms. It might seem from this that character information could never challenge the theory. But if the evidence for irreversibility was originally based on character distributions, then it would be quite unwarranted to analyze further cases so as to force them into conformity with irreversibility. The effect of doing so would be precisely to confer on irreversibility the status of an empirical conclusion that cannot be questioned by evidence—a contradiction in terms. In order to avoid that fallacy, it is necessary to allow that character information may support a conclusion of reversal. Whenever a putative reversal offers a more complete (that is, as already seen, more parsimonious) explanation of observed similarities than does a reconstruction enforcing irreversibility, irreversibility must be discounted. (In that particular case trends might still be accepted as rough descriptive generalizations.)

A proponent of irreversibility might nevertheless insist that when an analysis that does not presuppose irreversibility gives a different result from another that does use that premise, then the conclusion of the former depends crucially on the supposition that reversal is possible. Moreover, as the procedure just outlined will always discount irreversibility when parsimony requires, there is no way of rejecting the possibility of reversal. That possibility might seem, then, to be an ad hoc hypothesis, so that a conclusion of reversal actually requires more ad hoc hypotheses than would be suggested just by counting independent origins. But even if possibility of reversal did constitute an ad hoc hypothesis, it would certainly not be an additional independent hypothesis, for it is entailed by the particular hypothesis of reversal postulated. That observation, in fact, contains the key to the defect of the whole objection. If a particular conclusion of reversal could be legitimately criticized as presupposing the possibility of reversal, then any scientific conclusion whatever could be dismissed as requiring the supposition of its own possibility. The argument outlined is seen in that light to be simply another rationalization for discarding evidence.

It is clear that the reasoning outlined effectively views irreversibility and the possibility of reversal as competing theories. The charge that possibility of reversal cannot be rejected by parsimony analysis would be pertinent only as a criticism of a way of testing an empirical claim. But that view is itself suspicious. Irreversibility is certainly an empirical claim, and, furthermore, it is plainly testable in principle, inasmuch as it prohibits something, namely reversals. The possibility of reversal, on the other hand, can hardly be by itself an empirical claim in the same sense (although the claim that particular reversals have occurred might be), as it does not prohibit anything. One might think that admitting that reversal might occur, if it is not itself directly an empirical contention, nonetheless implies one, in that using a method that can discard irreversibility for parsimony would necessarily yield conclusions of reversal. But in fact it is quite possible for a parsimonious reconstruction to lack requirements for reversal.

(The contrary, of course, is also possible, and is often observed. But that is a consequence of the idea in conjunction with particular observations, not of the idea itself.) While irreversibility and the possibility of reversal seem superficially to be simply alternative theories, then, they are in fact not the same kind of idea. The first is a theory that forbids conclusions that might otherwise seem supported by observation, and, when confronted with such cases, can be saved only by ad hoc supposition. The second is simply an attitude. The possibility that irreversibility (or any theory) is false must be considered in order to test the theory. No kind of empirical science would be possible without such attitudes.

Polymorphism

Because of their reliance on covering assumptions to justify otherwise unnecessary ad hoc hypotheses of homoplasy, the Camin–Sokal and Dollo techniques should not be regarded as proper parsimony methods, prior usage notwithstanding. The situation may be different, however, for another restricted procedure. In the chromosome inversion model of Farris (1978), it is presumed that each of two alternative inversion types originated uniquely. Inversion types may nonetheless show incongruence with a genealogy through independent fixations from polymorphic ancestral populations, and the tree is chosen to minimize such fixations. The accuracy of the premise of unique origin might, of course, be questioned, but the idea is accepted by specialists on grounds extrinsic to genealogical hypotheses, and I shall not attempt to dispute its validity here. A further observation in this connection, however, seems worthwhile.

As this model presumes a unique origin for each inversion type, it might seem that similarity between organisms would on this premise be due to inheritance regardless of the genealogy postulated, so that the relationship between parsimony and genealogical explanation would no longer hold. The inherited similarity covered by the premise, however, holds only for chromosomes of individuals. Resemblance between populations fixed for the same inversion may still be explained by inheritance, or else the coincidental result of independent fixations. As it is populations that are grouped in postulating a genealogy, it is still possible to compare alternative genealogical hypotheses on explanatory power. There is in fact nothing unusual in this conclusion. It is generally true that features used to arrange taxa are characteristics of populations, rather than of individuals. The observation that deer have antlers is just a contracted way of stating that normal, adult, male deer in breeding condition possess those structures. The females, young, and deformed are not given a separate place in the system by reason of lacking the characteristic. The same principle underlies Hennig's emphasis of the idea that holomorphs rather than specimens are classified. Mickevich and Mitter (1981) arrive at the same concept in developing their greatly improved methods for analyzing electrophoretic data. They concentrate on recognizing suites of alleles as features of populations, rather than attempting to use single alleles—traits of individuals—as characters.

Phenetic Clustering

Clustering by raw similarity (phenetic clustering) has sometimes been advocated as a means of making genealogical inferences, typically with the justifying assumption that rates of evolutionary change (or divergence) are nearly enough constant so that degree of raw similarity reflects recency of common ancestry. The method is most often used with comparative biochemical data, but it has been recommended for morphological data as well (for example, by Colless 1970).

Constancy of rate is rather a different theory from irreversibility of evolution, but many of the comments made earlier apply here as well. Phenetic clustering might coincidentally produce a parsimonious scheme, but it certainly need not do so, and again there is no limit in principle to the number of otherwise unnecessary requirements for hypotheses of homoplasy that this method might impose. The assumption is certainly an empirical claim, and advocates of the method usually defend it by producing evidence for rate constancy. (Colless is an exception; he shows no inclination to resort to evidence.) That evidence typically takes the form of correlations between observed raw similarities and putative recency of common ancestry. Those last naturally depend on hypotheses of kinship, and this raises the familiar dilemma. If the genealogies used as evidence depended crucially on rate constancy, there would be no evidence. Supposing, then, that they do not, the evidence must consist of agreement between the theory and arrangements that conform to character distributions. Just as before, if the premise of rate constancy is used to justify unparsimonious conclusions, the effect is to consider rate constancy as empirical and irrefutable at one. Likewise parsimony analysis might be accused of presupposing that rates can vary, but discussion of that idea would precisely parallel what has already been said in connection with irreversibility.

A molecular evolutionist is quite happy with the generally good correlation that is observed between raw similarity and putative recency of common ancestry; for him it substantiates the molecular "clock." But as I have emphasized before (Farris 1981), such correlations are not enough to justify clustering by raw similarity. The correlations reported show considerable scatter. The implication of this, accepting the usual interpretation of the general correlation, is that rates of divergence vary somewhat. Even if it is often true, then, that genealogically most closely related taxa are also mutually most similar, there are evidently exceptions. Those exceptions could not be identified if genealogy were inferred by presupposing that raw similarity reflects kinship. To make accurate inferences in such cases—to discover what the cases are—it is necessary to use a method that can discount raw similarity as indicative of kinship if the data seem to require doing so—if doing so is required to achieve a more complete explanation of observed features. By analogy with the discussion of irreversibility, the same conclusion would be reached just by requiring that the relationship of raw similarity to kinship be vulnerable to evidence. It seems, then, that a correlation between raw similarity and kinship—even if it often holds—can provide no legitimate grounds for accepting unparsimonious inferences.

I commented before on the distinction between an ad hoc covering assumption and a corroborated improved theory able to account systematically for observations that would otherwise seem coincidental departures from its predecessor. This distinction suggests a further defect in the attempt to defend phenetic clustering on grounds of a correlation between raw similarity and kinship. In a legitimate extension of theory, the old coincidences are not dismissed as such, but explained by the extension. The process, that is, expands explanatory power, rather than discarding it. Suppose that clustering by raw similarity in some case requires otherwise unnecessary hypotheses of homoplasy, and that the conclusion is defended on grounds of a theoretical relationship between raw similarity and recency of common ancestry. If this is not ad hoc, then the theory must offer an explanation of the putative homoplasies. It is far from clear, however, that it can do so. Homoplasies, as already observed, are not explained by the inferred genealogy, from the standpoint of which the shared features that they represent are so only coincidentally. Inasmuch as raw similarities are calculated from features, it seems curious that they could either explain or be explained by a scheme that left the feature themselves unaccounted for. In order for a relationship between raw similarity and kinship to explain homoplasies, furthermore, it would seem necessary to suppose that that relationship rests on some real mechanism. That mechanism would have to have the property that organisms would come to possess features in common for reasons other than inheritance, and in just such a way as to maintain the correlation between raw similarity and recency of common ancestry. As no known natural process appears to have this property, it would seem that use of a postulated correlation between raw similarity and kinship to defend clustering by raw similarity rests necessarily on an ad hoc covering assumption.

A related conclusion can be reached by another route. Phenetic clustering ignores considerations of parsimony and so effectively proceeds by freely introducing whatever hypotheses of homoplasy are needed to derive a result conforming to the rate constancy premise. The procedure would be highly questionable on statistical grounds alone, then, if homoplasy were supposed to be rare. The method then requires the assumption that homoplasy is abundant, and indeed its proponents are prominent in criticizing parsimony as requiring rarity of homoplasy. The premise that homoplasy is abundant, however, poses a problem for clustering by raw similarity as well. That method infers recency of common ancestry of two taxa from the fraction of characters in which the two are similar. If homoplasy were rare (and rates constant), that would be superficially reasonable. Similarity between two lineages would decrease in clocklike fashion as ancestral similarities were lost. But if homoplasy is abundant, many of the similarities between two taxa are likely to be homoplasious, in which case they need indicate nothing about how recently the pair diverged. Two populations having only a remote common origin, and so (if rates were constant) very little homologous similarity, might have many recently acquired homoplasies, and so be judged to be of recent common ancestry. It is easy enough to identify conditions under which inferences might still be valid. If pairwise homoplasies were all the same, or nearly so, homoplasy would not alter the relative degrees of raw similarity among taxa, and then (if rates of change were constant) the method would still work. Phenetic clustering effectively presumes, then, that the variance of pairwise homoplasies is small. Keeping that variance small would be the task of the hypothetical mechanism just discussed.

While phenetic clustering does not consider homoplasies as such, it does select a tree by finding a constant-rate (ultrametric) model that conforms to observed raw similarities as closely as possible. If rates of evolution were constant, homologous similarities would conform to the constant rate model, so that departure from the model would be due to variation in pairwise homoplasies. The phenetic clustering procedure most commonly applied for genealogical inference, UPGMA, has precisely the effect of minimizing the variance of pairwise departures of observed from ultrametric similarities (Farris 1969b). Phenetic clustering and parsimony analysis are similar, then, in the sense that each minimizes a criterion. But whereas abundance of homoplasy need not imply error by parsimonious inference, large variations in pairwise homoplasies would certainly vitiate the conclusion of phenetic clustering. Phenetic clustering, unlike parsimony, depends crucially on minimality in nature of the quantity that it minimizes. Clustering by raw similarity possesses the very sort of defect that its proponents had incorrectly claimed as a weakness of phylogenetic analysis.

Cliques

Clique methods rely on parsimony to interpret suites of congruent characters, but their trees require homoplasy for characters outside the selected clique, and often the clique tree will be quite unparsimonious for those characters. In practice the excluded characters are often numerous, so that basing the inferred genealogy just on the clique imposes a considerable loss of explanatory power. These methods are then prime suspects for reliance on a covering assumption, but for a long time it was not clear from the clique literature what that assumption was supposed to be.

Le Quesne (1972) offered an approximate method (later made exact by Meacham 1981) for finding the probability that a suite of characters would all be congruent if features were distributed independently and at random among taxa. He suggested selecting the clique with the lowest such probability, and other proponents also commonly refer to cliques as "least likely." It is possible that this idea is intended as a justification of clique methods. If so, the justifying reasoning amounts to no more than misunderstanding of statistics. If a clique were evaluated just on its probability under a random model, the evaluation would be bound to the model. In that case the covering assumption of cliques would be that characters—being randomly distributed—have no relationship to genealogy. Perhaps it was intended that low probability under a null model would lend credence to an alternative, genealogical interpretation of the clique, but that idea, too, rests on a fallacy. Observing that a clique (or anything else) has low probability under a model might provide statistical grounds for rejecting the model, but it does not by itself offer any basis for choosing any particular alternative hypothesis. Once a model has been rejected, the probabilities it assigns to events necessarily become irrelevant. In this case rejecting the null model is uninformative, as no one interested in making phylogenetic inferences would have taken it seriously anyway. The statistical reason for accepting a new hypothesis is that it assigns much higher probability to observation than does the old. In normal statistics, a large enough difference between sample means serves as grounds for rejecting the hypothesis that the two samples were drawn from populations with the same parametric mean. If an alternative hypothesis is chosen so that it assigns maximum probability to the observed difference, the new theory conforms best with observation. But one hardly proceeds by choosing observations so as to minimize their nominal probability under the original hypothesis, let alone using such observations as the basis for choosing a new theory. Making statistical genealogical inferences from characters that had been used to reject the hypothesis of randomness would likewise require choosing a genealogy that would assign maximum probability to available characters—a maximum likelihood tree. I have already commented on the difficulties of applying that approach in practice, but this case is far worse. No model other than the rejected one of randomness is provided, and so neither are any grounds whatever for accepting the tree from the "least likely" clique as a genealogical inference best conforming to observation. (Felsenstein has made much the same point.)

As none of the ideas just discussed provides any legitimate rational for clique methods, those procedures must rest on an undisclosed assumption, if indeed they rest on anything at all. It is not difficult to discern what that assumption would have to be. Cliques are usually chosen to comprise as many mutually congruent characters as possible, and any characters that must be discarded to achieve this are simply counted as excluded. If the genealogy corresponding (by parsimony) to the clique is accepted, each of the excluded characters will require at least one hypothesis of homoplasy, but the number required may well vary among those characters. As characters are counted just as excluded or not, the number of hypotheses of homoplasy required by excluded characters plays no role in the analysis: similarities in those characters are dismissed en masse. The covering assumption involved is thus like the archetypical one discussed before. Ad hoc hypotheses of a sort are counted, but the counts do not reflect simple hypotheses of homoplasy. Instead any and all similarities in each excluded character are discounted by recourse to a "single" covering assumption. Excluding a character amounts to treating all similarities in it as irrelevant to assessing kinship. Those similarities could all be logically irrelevant only if they were all homoplasies. The covering assumption utilized is, then, that excluding a character—concluding that it shows some homoplasy—implies that all points of similarity in that character are homoplasies.

As discussed before, the collective dismissal of similarities in a character would be justified if the multiple required origins of features were not logically independent. It is readily seen, however, that such is not the case. The conclusion that endothermy has evolved independently in mammals and in birds does not imply that each species of bird or mammal has independently achieved that condition. Such being the case, it is likewise clear, from earlier discussion, that use of such a covering assumption leads to loss of explanatory power. As before, a single requirement of homoplasy may leave many of the similarities in a character explained, while a large enough number of required homoplasies will leave the same similarities entirely unexplained. Counting characters as simply excluded or not produces an evaluation oblivious to that distinction.

An attempt might be made to defend clique methods by advancing their covering assumption as an empirical claim on evolution, although of course doing so would raise the same sort of difficulties already discussed for irreversibility and rate constancy. Clique advocates have not tried to take that course—perhaps for fear of inviting ridicule. Unlike superficially tenable premises such as rate constancy, the clique assumption implies a theory that no one would take seriously as a realistic possibility. A different approach to rationalizing cliques has been taken by Felsenstein, though. He has proposed two stochastic models (Felsenstein 1979, 1981) under which he derives cliques as maximum likelihood estimates of genealogy. Both of these operate, just as would be expected from the clique assumption, by supplying principles that would excuse dismissing characters as units. In the 1979 paper this effect is achieved by introducing the possibility of a carefully selected type of error. Any character incongruent with the accepted tree is characterized as erroneous, and this is taken to mean that the character has been so completely misinterpreted that it is uninformative on genealogy. In the other model, incongruent characters are instead regarded as having changed so frequently in evolution as to be unrelated to genealogical grouping. Felsenstein himself emphasized that his models are inconsistent with the observed frequency of incongruence among characters. Realism thus plays no role in these justifications, which seem aimed instead at defending the clique assumption just by translating it into statistical terminology. Both rationalizations, consequently, have the same faults as the clique assumption itself.

The error idea rests on a misrepresentation of how systematists recognize characters. At one time it was believed that the eyes of octopi and of vertebrates were the same. That was certainly an error, for the two organs differ in both structure and ontogeny. But that the mistake was made does not mean that the sameness attributed to the eyes of rats and of mice is likewise a misrepresentation. Concluding a homoplasy in a feature may well invite renewed inspection and possible reinterpretation. But that reinspection certainly need not lead to dismissing all the agreements between taxa that the original feature had been intended to summarize. While it is reasonable to attribute some homoplasies to errors, it does not follow from this that those errors will turn out to have universal effects. It is seen from this that Felsenstein's use of the idea of wholesale error as a defense of the clique assumption amounts to no more than stating the desired conclusion as a premise. The defect of cliques is just that they treat every conclusion of homoplasy as if it implied universal homoplasy. Felsenstein attributes homomplasies to errors, but bolsters cliques only by supposing that any conclusion of error implies universal error. Neither implication is valid, and so either is merely an ad hoc rationalization for dismissing relevant evidence.

Much the same applies to Felsenstein's second argument. Dismissing an incongruent character on the grounds that it must have changed very frequently clearly depends on discounting the possibility that it changed only a few times. As Felsenstein (1981, p. 183) puts it, the clique method is suitable when "it is known that a few characters have very high rates of change, and the rest very low rates, but it is not known which characters are the ones having high rates." He does not disclose, however, how one comes to know the rate of change of a character without a prior phylogenetic analysis. Nor does he explain how one would apply that undisclosed method to gain the knowledge that his method calls for, without in the process incidentally learning which characters had the high rates. Nor, again, does he offer any pretense of a reason why rates should restrict themselves to be either very high or else very low—or why rapidly changing characters ought to be "few." In the absence of such explanation, it is seen that the covering assumption that one conclusion of homoplasy implies universal homoplasy has once again been "defended" simply by restating it as the entirely equivalent—and equally unsubstantiated—premise that any feature that originates more than once must have done so a very great number of times.

Conclusion

Advocacy of nonphylogenetic methods has consistently been based on the charge that parsimony depends on unrealistic assumptions. That allegation has never been supported by substantial argument. It has been instead motivated by the dependency of other approaches on false suppositions: Proponents of other views have tried to bolster their position through the pretense that no means of phylogenetic analysis can be realistic.

Parsimony analysis is realistic, but not because it makes just the right suppositions on the course of evolution. Rather, it consists exactly of avoiding uncorroborated suppositions whenever possible. To a devotee of supposition, to be sure, parsimony seems to presume very much indeed: that evolution is not irreversible, that rates of evolution are not constant, that all characters do not evolve according to identical stochastic processes, that one conclusion of homoplasy does not imply others. But parsimony does not suppose in advance that those possibilities are false—only that they are not already established. The use of parsimony depends just on the view that the truth of those—and any other—theories of evolution is an open question, subject to empirical investigation.

The dichotomy between parsimony and supposition is just that; parsimony offers no barrier to evolutionary theories as such. Rate constancy—or any other supposition—seems to be in conflict with parsimony in the abstract, as it seems to offer a different basis for making genealogical inferences. But it would conflict with parsimony in application only in conjunction with observation, if maintaining the supposition required discarding a parsimonious—explanatory—interpretation of evidence. In that case, however, the same evidence would serve to question the supposition, which could then be defended only by presupposing its truth, or—entirely equivalently—simply dismissing the evidence. But if parsimonious interpretation of evidence did not refute the supposition, then the latter would become a corroborated theory. Parsimony does not require that no such theories will be corroborated, but offers a means for that corroboration, provided evidence allows it. Unlike prior supposition, empirically supported evolutionary theories can offer no criticism of parsimony, for those theories could have become corroborated just to the extent that they require few dismissals of evidence. The insistence by proponents of suppositions that parsimony is unrealistic, it is then seen, is merely a subterfuge. Ostensibly the objection is to parsimony, but in fact the complaint is that some cherished idea does not conform to evidence.

I return finally to the questions raised at the beginning. Phylogenetic analysis is necessarily based on parsimony, both because it is precisely the criterion that leads to grouping according to putative synapomorphy, and because empirical investigation is impossible without avoiding ad hoc hypotheses. Only synapomorphy provides evidence of kinship, for the attempt to use raw similarity as evidence would necessarily either rest on uncorroborated—and so nonevidential—supposition, or else could lead to no conclusion conflicting with synapomorphy. And phylogenetic analysis is most certainly empirical, for in applying the parsimony criterion, it chooses among alternative hypotheses of relationship on the basis of nothing other than their explanatory power. Differing as it thus does from all other approaches, phylogenetic systematics alone provide a logical basis for the empirical study of the relationships among organisms.

References

Beatty, J., and W. L. Fink. 1979. Review of *Simplicity*. *Syst. Zool.* 28: 643–651.

Camin, J. H., and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. *Evolution* 19: 311–326.

Cartmill, M. 1981. Hypothesis testing and phylogeny reconstruction. *Zeit. Zool. Syst. Evolut.-forsch.* 19: 73–95.

Colless, D. H. 1970. The phenogram as an estimate of phylogeny. *Syst. Zool.* 19: 352–362.

Edwards, A. W. F., and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. In V. H. Heywood and J. McNeill, eds., *Phenetic and Phylogenetic Classification*. London, Systematics Association, pp. 67–76.

Farris, J. S. 1966. Estimation of conservatism of characters by constancy within biological populations. *Evolution* 20: 587–591.

———. 1967. The meaning of relationship and taxonomic procedure. *Syst. Zool.* 16: 44–51.

———. 1969a. On the cophenetic correlation coefficient. *Syst. Zool.* 18: 279–285.

———. 1969b. A successive approximations approach to character weighting. *Syst. Zool.* 18: 374–385.

———. 1970. Methods for computing Wagner trees. *Syst. Zool.* 19: 83–92.

———. 1973. On the use of the parsimony criterion for inferring evolutionary trees. *Syst. Zool.* 22: 250–256.

———. 1977. Phylogenetic analysis under Dollo's law. *Syst. Zool.* 26: 77–88.

———. 1978. Inferring phylogenetic trees from chromosome inversion data. *Syst. Zool.* 27: 275–284.

———. 1979. The information content of the phylogenetic system. *Syst. Zool.* 28: 483–519.

———. 1980. The efficient diagnoses of the phylogenetic system. *Syst. Zool.* 29: 386–401.

———. 1981. Distance data in phylogenetic analysis. In *Advances in Cladistics: Proceedings of the First Meeting of the Willi Hennig Society*, ed. V. A. Funk and D. R. Brooks. New York Botanical Garden.

———. 1982. Simplicity and informativeness in systematics and phylogeny. *Syst. Zool.* 31: 413–444.

Farris, J. S., and A. G. Kluge. 1979. A botanical clique. *Syst. Zool.* 28: 400–411.

Farris, J. S., A. G. Kluge, and M. F. Mickevich. 1982. Phylogenetic analysis, the monothetic group method, and myobatrachid frogs. *Syst. Zool.* 31: 317–327.

Felsenstein, J. 1973. Maximum likelihood and minimum steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.* 22: 240–249.

———. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27: 401–410.

———. 1979. Alternative methods of phylogenetic inference and their interrelationship. *Syst. Zool.* 28: 49–62.

———. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. Jour. Linn. Soc.* 16: 183–196.

Funk, V. A., and D. R. Brooks. 1981. National Science Foundation workshop on the theory and application of cladistic methodology. Organized by T. Duncan and T. Stuessy. University of California, Berkeley, March 22–28, 1981. *Syst. Zool.* 30: 491–498.

Gaffney, E. S. 1979. An introduction to the logic of phylogeny reconstruction. In J. Cracraft and N. Eldredge, eds., *Phylogenetic Analysis and Paleontology*. New York, Columbia University Press, pp. 79–112.

Hennig, W. 1966. *Phylogenetic Systematics*, Urbana, Ill., University of Illinois Press.

Kluge, A. G., and J. S. Farris. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18: 1–32.

Le Quesne, W. J. 1969. A method of selection of characters in numerical taxonomy. *Syst. Zool.* 18: 201–205.

———. 1972. Further studies based on the uniquely derived character concept. *Syst. Zool.* 21: 281–288.

———. 1974. The uniquely evolved character concept and its cladistic application. *Syst. Zool.* 23: 513–517.

Meacham, C. A. 1981. A probability measure for character compatibility. *Math. Biosci.* 57: 1–8.

Mickevich, M. F. 1978. Taxonomic congruence. *Syst. Zool.* 27: 143–158.

———. 1980. Taxonomic congruence: Rohlf and Sokal's misunderstanding. *Syst. Zool.* 29: 162–176.

Mickevich, M. F., and M. S. Johnson. 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. *Syst. Zool.* 25: 260–270.

Mickevich, M. F., and C. Mitter. 1981. Treating polymorphic characters in systematics: A phylogenetic treatment of electrophoretic data. In V. A. Funk and D. R. Brooks, eds., *Advances in cladistics: Proceedings of the First Meeting of the Willi Hennig Society*. New York Botanical Garden, pp. 45–60.

Mickevich, M. F., and J. S. Farris. 1981. The implications of congruence in *Menidia*. *Syst. Zool.* 30: 351–370.

Nei, M. 1972. Genetic distance between populations. *Amer. Nat.* 106: 283–292.

Riggins, R., and J. S. Farris. 1983. Cladistics and the roots of angiosperms. *Syst. Bot.* 8: 96–101.

Rohlf, F., and Sokal, R. 1981. Comparing numerical taxonomic studies. *Syst. Zool.* 30: 459–485.

Schuh, R. T., and J. T. Polhemus. 1980. Analysis of taxonomic congruence among morphological, ecological, and biogeographic data sets for the Leptopodomorpha (Hemiptera). *Syst. Zool.* 29: 1–26.

Schuh, R. T., and J. S. Farris. 1981. Methods for investigating taxonomic congruence and their application to the Leptopodomorpha. *Syst. Zool.* 30: 331–351.

Sober, E. 1975. *Simplicity*, London, Oxford University Press.

———. 1982. Parsimony methods in systematics. *Hennig Society II*. New York, Columbia University Press.

Watrous, L. E., and Q. D. Wheeler. 1981. The out-group comparison method of character analysis. *Syst. Zool.* 30: 1–11.

Wiley, E. O. 1975. Karl R. Popper, systematics, and classification: A reply to Walter Bock and the evolutionary taxonomists. *Syst. Zool.* 24: 233–243.

———. 1981. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. New York, John Wiley and Sons.

XI Race—Social Construction or Biological Reality?

22 Why There Are No Human Races

Kwame Anthony Appiah

My aim in this essay is to defend the claim that American social distinctions cannot be understood in terms of a supposedly biological concept of race.[1] The only human race in the United States, in a slogan, is *the* human race. But (typically for a philosopher, perhaps), I'm going to come at the question in a somewhat roundabout way. And to make my argument I'm going to need to draw on two different and competing philosophical notions of what it is to give an adequate account of the meaning of a word or expression, such as the word "race."

One—we can call this the *ideational* view of meaning—associates the meaning of a term with what you might call an idea. Understanding the idea of race involves grasping how people think about races: what they take to be the central truths about races; under what sorts of circumstances they will apply the idea of race; what consequences for action will flow from that application.

The other picture of meaning—the *referential* view—suggests that what it is to explain what the word "race" means is, in effect, to identify the things to which it applies, the things we refer to when we speak of "races."

These views are not as far apart as they might at first appear. To find out what people are referring to in using the word "race," after all, you might need to know what idea their word "race" expresses: if they had no ideas, no thoughts, about race and if there were no circumstances when they used the word, no consequences to their applying it, then we could hardly suppose that their making the sound "race" meant anything at all. In practice, at least, access to an idea of race is probably needed to find the referent.

And, conversely, once we have identified the referent—found, that is, the races—we can assume that people who understand the word "race" have some beliefs that are at least roughly true of races. For if people are talking about races, it is because they have, or think they have, experience of races: and, generally speaking, some of that experience will be reliable. A little bit of knowledge of what races are like combined with a

K. Anthony Appiah, *Color Consciousness*. © 1996 Princeton University Press. 1998 paperback edition. Reprinted by permission of Princeton University Press.

little information about what people are like—how sensory experience works, for example—will allow us to predict at least some of people's ideas about races.

My aim is not to decide between these two broad traditions of conceiving of meaning. Anyone concerned to understand our concept of race ought, I think, to be interested both in the reality of race and in the way people think about it, in both the referential and the ideational aspects: we can leave it to the philosophers of language to wrangle about which of these ought to have the central place in semantics (or whether, as I suspect, we need both of them).

Perhaps the simplest ideational theory of meaning runs like this: what we learn when we learn a word like "race" is a set of rules for applying the term. Everybody who knows what the word "race" means, which means most competent speakers of English, learns the same rules: so that, while people have different beliefs about races, they share some special beliefs—I'll call them the *criterial* beliefs—that define the concept. These beliefs may not be very high-powered. They might include, for example, the thought that people with very different skin colors are of different races or that your race is determined by the race of your parents. But, on this simplest ideational theory, all of these criterial beliefs have this property: someone who doesn't believe these things, doesn't understand what the English word "race" means.

The simplest theory would also require that if we collected together all these criterial beliefs about race and took them all together, they could be thought of as defining the meaning of the word "race." (This is equivalent to saying that there are things that have to be true of something if it is to be a race—conditions necessary for being a race; and that these necessary conditions are, when taken together, sufficient for being a race.) We can use a device invented by the English philosopher Frank Ramsey in the nineteen-twenties to make this an explicit definition: Something is a race just in case all the criterial beliefs are true of it.[2] Let's call this the "strict criterial theory."

The Ramsey definition makes clear the connection between defining a term and questions of existence: there are races if, but only if, there are things that satisfy all the criteria.

For a number of reasons, which I want to skirt, you won't get many philosophers of language to buy into this strict criterial theory today; but you don't need high-falutin' semantic arguments to be lead to wonder whether we could in fact write a Ramsey-style definition of the word "race." Consider one of the two claims I gave a little while ago. *Your race is determined by the race of your parents.*

Two people marry. The wife has one Ghanaian and one British parent. The father's parents are Norwegian. They have children of various shades, one of whom looks, to all intents and purposes, like an average Norwegian. My friend Georg agrees that the mother's parents are of different races and contends that the Norwegian-looking son is Caucasian, but his darker brothers are not. Does Georg not know what "race" means? Apparently, if people with two parents of the same race are of the same race as their parents. For, if your race is determined by the race of your parents, you must have the same race as your full siblings.

It seems to me simply unconvincing to insist that Georg doesn't know what the word "race" means; at least if knowing what it means is knowing whatever you need to know to count as a competent user of the English word "race". This fails, of course, to establish that we couldn't find a set of beliefs necessary and sufficient for understanding the word "race"; beliefs, that is, that everybody who understands the word "race" must have and such that everybody who has them understands the concept of race. But if even *these* rather uncontroversial-looking claims turn out to be ones that can be denied by someone who understands the word "race," then one might begin to wonder whether *any* claims will turn out to be necessary: and if none are necessary, then certainly the conjunction of the necessary conditions won't be sufficient.

Such doubts about the strict criterial theory—in terms of criteria individually necessary and jointly sufficient—lead us on to the next obvious proposal, one that might seem to be suggested by Wittgenstein's use of the notion of a criterion.[3] Perhaps, what is required to know what "race" means, is that you should believe most of the criterial beliefs (or a good number of them), but not that you should believe any particular ones. The explicit definition that captures the common notion of those who understand the word "race" will then be given by a modified Ramsey-style definition: A race is something that satisfies a good number of the criterial beliefs. I'll call this the "vague criterial theory."

Accepting this theory has certain important consequences. First of all, it isn't going to allow us to draw a sharp line between not knowing what the word "race" means and having unusual views about races. That boundary is vague, because the expression "a good number" is vague.

Second, the theory admits that among the criterial beliefs there are some that are plainly not held by everybody who uses the word "race." These, for example: *Most sub-Saharan Africans are of the Negro race. Most Western Europeans are of the white race. Most Chinese are of the yellow race. Everybody has a race. There are only a few races.*

There are clearly people who count as understanding the term "race" who don't believe each of these things. Somebody who uses the word "race" may have no thoughts at all about Africa or Western Europe or China, need not know even that they exist. I, as you will see, deny that everybody has *a* race, because I think nobody has a race: but there are more moderate folks who think that people of so-called mixed-race are neither of the races of their parents nor of some separate race and deny that everybody has *a* race for that reason.[4] And there have been physical anthropologists who felt that the only useful notion of race classified people into scores of kinds.

If the strict criterial theory had been true, it would have been easy to argue against the existence of races. One would only have had to find the correct definition and then show that nothing in the world actually satisfied it. This looser theory makes it, correspondingly, harder to argue against the existence of races. But the vague criterial theory does suggest a route to understanding the race concept: namely to explore the sorts of things people believe about what they call ''races'' and to see what races would have to be like for these things to be true of them. We can then inquire as to whether current science suggests that there is anything in the world at all like *that*.

Now, suppose there isn't one such thing in the world; then, on this view, there are no races. It will still be important to understand the vague criteria, because these will help us to understand what people who believe in races are thinking. That will be important, even if there are no races: first, because, we often want to understand how other people are thinking, for its own sake; and, second, because people act on their beliefs, whether or not they are true. Even if there are no races, we could use a grasp of the vague criteria for the concept race in predicting what their thoughts and their talk about race[5] will lead them to do; we could use it, too, to predict what thoughts about races various experiences would lead them to have.

Now, I have already declared myself very often on the question whether I think there are any races. I think there aren't. So it is important that I am clear that I also believe that understanding how people think about race remains important for these reasons, even though there aren't any races. To use an analogy I have often used before, we may need to understand talk of ''witchcraft'' to understand how people respond cognitively and how they act in a culture that has a concept of witchcraft, whether or not we think there are, in fact, any witches.

The ideational view might, therefore, lead you to explore contemporary thought and talk about races. But I think that this is likely to produce a confusing picture. This is because current ways of talking about race are the residue, the detritus, so to speak, of earlier ways of thinking about race; so that it turns out to be easiest to understand contemporary talk about ''race'' as the pale reflection of a more full-blooded race-discourse that flourished in the last century. The ideational theory can thus be combined with an historical approach: we can explore the ideational structures of which our present talk is, so to speak, the shadow, and then see contemporary uses of the term as drawing from various different structures, sometimes in ways that are not exactly coherent.

Before we turn to historical questions, however, let me ask what route to understanding the race-concept is suggested by the referential account of meaning.

The answer is most easily understood by thinking about an issue in the history and philosophy of science. From the point of view of current theory some previous theories—early nineteenth century chemistry, say—look as though they classified some things—acids and bases, say—by and large correctly, even if a lot of what they said about those things was pretty badly wrong. From the point of view of current theory, you might argue, an acid is, roughly, a proton-donor. And our recognition of the fact that the classification of acids and bases was in itself an intellectual achievement is recorded in the fact that we are inclined to say that when Sir Humphrey Davy—who, not having any idea of the proton, could hardly be expected to have understood the notion of a proton-donor—used the word ''acid,'' he was nevertheless talking about what we call acids.

The issues here are at the intersection of the philosophy of language and the philosophy of science. And in explaining why it seems proper to think that Sir Humphrey Davy was referring to the things we call proton-donors, even though much of what he believed about acids is not true of proton-donors, philosophers of science have borrowed ideas about reference from recent philosophy of language.

One proposal some have borrowed is what is called the ''causal theory of reference.'' The basic idea is simple enough: if you want to know what object a word refers to, find the thing in the world that gives the best causal explanation of the central features of uses of that word. If you want to know what the name ''New York'' refers to, find the object in the world that is at the root of most of the causal chains that lead to remarks containing the expression ''New York.''

So, in the case of acids, we are urged to believe that the stuffs ''out there'' in the world that really accounted for the central features of Davy's ''acid''-talk really were acids and that that is what accounts for our sense that Davy was not simply talking about something else (or, of course, about nothing at all). Early physiologists (like Descartes) who talked about ''animal spirits'' in the nerve fibers, on the other hand, we now say were referring to nothing at all: there is no currently recognized stuff that can account for what they said about animal spirits; instead there are truths about sodium pumps and lipid bilayers and synapses. There simply is no substance that was usually present when and only when the expression ''animal spirits'' was uttered and that behaves at all as they thought animal spirits behaved.

How can we use these ideas to develop a referential account of the concept of race? Well, we need to explore the sorts of things people have said about what they call ''races'' and see whether there is something in the world that gives a good causal explanation of their talk. If there *is* one thing in the world that best explains that talk, then that will be what the word ''race'' refers too; and that can be true, even if it would surprise most people to know that that was what they were really talking about—just as Sir Humphrey Davy would have been surprised to discover that, when he said ''acids'' he was talking about—referring to—proton-donors.

As a practical matter, at last three things are required for us to allow that a past theorist who spoke of ''Ys'' and was badly mistaken was nevertheless talking about *something*, call it ''X'':

first—*the existence condition—we* must acknowledge the existence of X; and,
second—*the adequacy condition—some* of what was thought to be true of what ''Y'' denoted must be at least approximately true of X; and
third—*the uniqueness condition*—X must be the best candidate for the job of ''Ys'' referent, so that no other thing that satisfies the existence condition satisfies the adequacy condition equally well.

On the causal theory, what it is for X to be the best candidate for the job of ''Ys'' referent in the speech of a community, is for X to be the thing that best causally explains their talk about ''Ys.'' So what we need to do, on this view, is to explore the history of the way the word ''race'' has been used and see if we can identify through that history some objective phenomenon that people were responding to when they said what they said about ''races.''

The difference between ideational and referential theories of meaning, then, is, roughly, that the referential theory requires we should do a historical version of what the ideational theory permits us to do. On the referential theory, exploring the history of the term is central to understanding what it means. Semantic considerations thus steer us towards historical enquiry. (Checking whether a past term meets the existence, adequacy and uniqueness conditions will also require us to draw on current science.)

The history I am going to sketch is the history of the ideas of intellectuals in the United States and the United Kingdom. You might ask why I don't look at the words of more ordinary people: race is statistically most important in ordinary lives. A good question, I say. (This is what you say when you think you have a good answer.) The reason is itself embedded in the history: as we shall see, throughout the nineteenth century the term ''race'' came increasingly to be regarded, even in ordinary usage, as a scientific term. Like many scientific terms, its being in use among specialists did not stop it being used in every day life. Treating it as a scientific term meant not that it was only for use by scientists, but that scientists and scholars were thought to be the experts on how the term worked. That is, with the increasing prestige of science, people became used to using words whose exact meanings they did not need to know, because their exact meanings were left to the relevant scientific experts.

In short, there developed a practice of *semantic deference*: people used words like ''electricity'' outside the context of natural philosophy or physical science, assuming that the physicists could say more precisely than they could what it meant. This semantic deference thus institutes a new form of what Hilary Putnam has called ''linguistic division of labor''; just as older specialties, like theology or law, had for a long time underwritten concepts—the Trinity, landlord—whose precise definition ordinary people didn't know.

The result is that even ordinary users of the term "race," who operated with what I have called vague criteria in applying it, thought of themselves as using a term whose value as a tool for speaking the truth was underwritten by the experts. Ordinary users, when queried about whether their term "race" really referred to anything, would have urged you to go to the experts: the medical doctors and anatomists, and later, the anthropologists and philologists and physiologists, all of whom together developed the scientific idea of race.

This makes the term "race" unlike many other terms in our language: "solid," for example. "Solid" is a term that we apply using everyday criteria: if I tell you that materials scientists say that a hunk of glass is not a solid but a liquid, you may well feel that they are using the term in a special technical sense, resisting semantic deference. Some people might want to defend the word "race" against scientific attacks on its legitimacy, by denying, in effect, that semantic deference is appropriate here. Of this strategy, I will make just this observation: if you're going to go that route, you should probably offer some criteria—vague or strict—for applying the term. This is because, as we shall see, the arguments against the use of "race" as a scientific term suggest that most ordinary ways of thinking about races are incoherent.

The understandings of "race" I am exploring are American; it seems appropriate enough, then, to begin with a thinker who helped shape the American republic: namely, Thomas Jefferson.

So let's look at Query XIV of the *Notes on the State of Virginia*, published in the seventeen-eighties. The emancipation of black slaves is inevitable Jefferson has argued; and it is right. But blacks, once emancipated, will have to be sent elsewhere. Jefferson anticipates that we may wonder why, especially given "the expence of supplying, by importation of white settlers, the vacancies they will leave."

> Deep rooted prejudices entertained by the whites; ten thousand recollections, by the blacks, of the injuries they have sustained; new provocations; the real distinctions which nature has made; and many other circumstances, will divide us into parties, and produce convulsions which will probably never end but in the extermination of the one or the other race.—To these objections, which are political, may be added others, which are physical and moral. The first difference which strikes us is that of colour. Whether the black of the negro resides in the reticular membrane between the skin and scarf-skin, or in the scarf-skin itself; whether it proceeds from the colour of the blood, the colour of the bile, or from that of some other secretion, the difference is fixed in nature, and is as real as if its seat and cause were better known to us. And is this difference of no importance? Is it not the foundation of a greater or less share of beauty in the two races? Are not the fine mixtures of red and white, the expressions of every passion by greater or less suffusions of colour in the one, preferable to that eternal monotony, which reigns in the countenances, that immoveable veil of black which covers all the emotions of the other race? Add to these, flowing hair, a more elegant

> symmetry of form, their own judgment in favour of the whites, declared by their preference for them, as uniformly as is the preference of the Oranootan for the black woman over those of his own species. The circumstance of superior beauty, is thought worthy attention in the propagation of our horses, dogs, and other domestic animals; why not in that of man?[6]

Apart from this difference of color with its attendant aesthetic consequences, Jefferson observes that there are other relevant differences: blacks have less hair on their face and bodies; "they secrete less by the kidnies, and more by the glands of the skin, which gives them a very strong and disagreeable odour"; "[t]hey seem to require less sleep."[7]

> Comparing them by their faculties of memory, reason, and imagination, it appears to me, that in memory they are equal to the whites; in reason much inferior, as I think one could scarcely be found capable of tracing and comprehending the investigations of Euclid; and that in imagination they are dull, tasteless, and anomalous.... [Among African-Americans] [s]ome have been liberally educated, and all have lived in countries where the arts and sciences are cultivated to a considerable degree, and have had before their eyes samples of the best works from abroad.... never yet could I find that a black had uttered a thought above the level of plain narration; never see even an elementary trait of painting or sculpture. In music they are more generally gifted than the whites with accurate ears for tune and time, and they have been found capable of imagining a small catch.... Misery is often the parent of the most affecting touches in poetry.—Among the blacks is misery enough, God knows, but no poetry.[8]

Though he tells us that "[t]he opinion, that they are inferior in the faculties of reason and imagination, must be hazarded with great diffidence,"[9] he nevertheless concludes:

> I advance it as a suspicion only, that the blacks whether originally a distinct race, or made distinct by time and circumstances, are inferior to the whites in the endowments both of body and mind. It is not against experience to suppose, that different species of the same genus, or varieties of the same species, may possess different qualifications. Will not a lover of natural history then, one who views gradations in all the races of animals with the eye of philosophy, excuse an effort to keep those in the department of man as distinct as nature has formed them. This unfortunate difference of colour, and perhaps of faculty, is a powerful obstacle to the emancipation of these people.[10]

After so conspicuously fair and balanced a discussion, it would have been hard not to share Jefferson's "suspicion." His very caution here adds to rather than detracting from the force of his conclusions; and after so much attention to the "difference... of faculty," it is easy to miss the fact that Jefferson believes that Negroes and whites must be kept apart, even if his "suspicion" is mistaken. For Jefferson the political significance of race begins and ends with color.

Jefferson's claims here about the Negro's faculties went neither unnoticed nor unanswered. And we can find, in his letters as in the *Notes*, evidence that he remained willing to entertain the possibility that his skepticism about the capacities of the Negro was unwarranted. Thanking the Abbé Grégoire for sending him a copy of his *De la littérature des Nègres*[11] Jefferson writes:

> Be assured that no person living wishes more sincerely than I do, to see a complete refutation of the doubts I have myself entertained and expressed on the grade of understanding allotted to them by nature, and to find that in that respect they are on a par with ourselves. My doubts were the results of personal observation [one wonders, a little, about the Orangutan here] on the limited sphere of my own State, where the opportunities for the development of their genius were not favorable, and those of exercising it still less so. I expressed them therefore with great hesitation; but whatever be their degree of talent it is no measure of their rights. Because Sir Isaac Newton was superior to others in understanding, he was not therefore lord of the person or property of others.[12]

I have quoted so much of Jefferson in part, of course, because Jefferson is an important figure in the history of American debates about racial politics; but mostly because in these passages I have cited we see something entirely representative of the best thinking of his day: the running together of biology and politics, science and morals, fact and value, ethics and aesthetics. Jefferson is an intelligent, sensitive, educated American shaped by the Western intellectual currents we call the Enlightenment. Race, for Jefferson and his peers, was *a concept they invoked to explain cultural and social phenomena*, it was also grounded in the physical and the psychological natures of the different races; it was, in other words, what we would call a *biological concept*.

I say that it was what *we* would call a biological concept, because the science of biology (even the word "biology") did not exist when Jefferson was writing the *Notes*.[13] What did exist was Natural History; and Jefferson would have agreed that race was a Natural Historical notion, as much as was the idea of species that Linnaeus had developed and which Buffon had popularized.[14] To think of race as a biological concept is to pull out of the Natural History of humans a focus on the body—its structure and function—and to separate it both from mental life—the province of psychology—and from the broader world of behavior and of social and moral life. If Jefferson's discussion, with its movement from questions of the morphology of the skin, to discussions of sexual desire, to music and poetry, strikes us as a hodge-podge, it is because we live at on the other side of a great intellectual chasm, which opens up with increasing speed through the nineteenth century. For we live now with a new configuration of the sciences; and, more especially, with the differentiation from the broad field of natural history, of anatomy, physiology, psychology, philology (i.e., historical linguistics), sociology, anthropology, and a whole host of even more specialized fields that gradually divided between them the task of describing and understanding human nature.

Jefferson's discussion is representative of a transition in the way the word "race" is used in reflecting on the characters of different kinds of peoples: the outer manifestations of race—the black skin of the Negro, the white skin and round eyes of the European, the oval eyes of the Oriental—have taken their place for him besides other, less physical, criteria, in defining race. The race of a person is expressed in all these ways,

> physical, moral, intellectual: they are referred back, so to speak, to a common cause or ground.

Jefferson conceives of racial difference as both physical and moral, but he is not *committed* to the view that race explains all the rest of the moral and social and political matter that is drawn into the portrait of the Negro in the *Notes*. The letter to Grégoire reveals a man who leaves open—at least in theory—the possibility "that nature has given to our black brethren, talents equal to those of the other colors of men"; and throughout the *Notes* Jefferson writes with real affection and respect about Indians, who "astonish you with strokes of the most sublime oratory; such as prove their reason and sentiment strong, their imagination glowing and elevated." The differences between whites and Indians, for Jefferson, hardly constitute a difference of essential natures.

If we move on another century or so from Jefferson's *Notes*, we enter once more a new intellectual landscape: one in which there is no longer any doubt as to the connection between race and what Jefferson calls "talent": and here, of course, the word "talent"—deriving from the New Testament parable of the talents—refers to inherited—to "native"—capacities.

Let me turn, then, from Jefferson, and move on into the second half of the nineteenth century, to the work of a poet and critic who, like Jefferson, uses the concept of race to explain the moral and the literary, but unlike him, is convinced that biological inheritance helps determine every aspect of racial capacity; namely Matthew Arnold.

Arnold was the greatest English critic of the nineteenth-century. He was also a central Victorian poet, an influential essayist, and lecturer: in short, a very public intellectual, whose influence was extended into the United States, not least by his lecture tour here in 1883 to 1884 (in his early sixties) which lead to the publication, in 1885, of *Discourses in America*.

In 1857 Matthew Arnold was elected to the Professorship of Poetry at Oxford, a position he held for about a decade. Ten years later, he published a series of lectures he had given as Professor of Poetry *On the Study of Celtic Literature*. In these lectures he argues that the ancient literature of the Celts—of Ireland and Wales, in particular—is part of the literary heritage of Britain; even of those Britons in England who by then conceived of themselves as heirs to a Saxon heritage and were inclined, by and large, to hold the Irish Celts, in particular, in less than high regard.

Here is how Arnold makes his case:

> . . . here in our country, in historic times, long after the Celtic embryo had crystallised into the Celt proper, long after the Germanic embryo had crystallised into the German proper, there was an important contact between the two peoples; the Saxons invaded the Britons and settled themselves in the Britons' country. Well, then, here was a contact which one might expect would leave

> its traces; if the Saxons got the upper hand, as we all know they did, and made our country be England and us be English, there must yet, one would think, be some trace of the Saxon having met the Briton; there must be some Celtic vein or other running through us.
>
> . . . though, as I have said, even as a matter of science, the Celt has a claim to be known, and we have an interest in knowing him, yet this interest is wonderfully enhanced if we find him to have actually a part in us. The question is to be tried by external and internal evidence; the language and physical type of our race afford certain data for trying it, and other data are afforded by our literature, genius, and spiritual production generally. Data of this second kind belong to the province of the literary critic; data of this first kind to the province of the philologist and the physiologist.
>
> The province of the philologist and the physiologist is not mine; but this whole question as to the mixture of Celt with Saxon in us has been so little explored, people have been so prone to settle it off-hand according to their prepossessions, that even on the philological and physiological side of it I must say a few words in passing.[15]

The ensuing discussion of what Arnold calls "physiology" is not what we should expect: it turns out that he is simply going to discuss the likelihood of mixture—i.e., breeding—between the races. He cites, for example, the opinion of a certain Monsieur Edwards that "an Englishman who now thinks himself sprung from the Saxons or the Normans, is often in reality the descendant of the Britons."[16] The appeal to philology, on the other hand, might seem to suggest an alternative mechanism for the transmission of racial traits—namely through language; but, in fact, philology is, for Arnold and his contemporaries, largely a guide to racial ancestry, with those whose languages are most closely related being also most closely related by blood. Arnold is clear that language can, in fact, be misleading:

> How little the triumph of the conqueror's laws, manners, and language, proves the extinction of the old race, we may see by looking at France; Gaul was Latinised in language manners, and laws, and yet her people remained essentially Celtic.[17]

But he is also convinced, as I say, that it can be a guide to racial character.

What Arnold lays out in these passages is the essence of what I call *racialism*. He believed—and in this he was typical of educated people in the English-speaking world of his day—that we could divide human beings into a small number of groups, called "races," in such a way that the members of these groups shared certain fundamental, heritable, physical, moral, intellectual and cultural characteristics with each other that they did not share with members of any other race.

There are a few complications to this basic picture, which we should bear in mind. First, there are two major ways in which counter-examples to claims about the members of the race could simply be ruled out. It was acknowledged that there were, to begin with, in all races, as there are in animal species, occasional defective members: in animals, the two-headed pigs and three-legged cats so beloved of tabloid journalism in my homeland of Ghana: in human beings, the mute, the mentally disabled, the blind. These individuals were not to count against the general laws governing the racial type. Similarly, the norm for each race might be different for males and females, so that a racial type might be defined by two norms, rather than one.

A second complication derives from the fact that many of the characteristics of the various races were described as dispositions or tendencies: a single person who was not defective might still differ from the average member of his race because his individual character dominated the natural tendencies he had inherited in his racial essence. Celts might all tend towards the sentimental: but a particular Welshman might, through an exercise of will, conquer his natural racial temper. As a result the failure of an individual to fit the norm for her race would not by itself refute the theory: for it might be that that person had simply conquered her inherited disposition. Many of what I shall call the characteristics of a race were thus not, to use a modern term, phenotypic: they did not necessarily display themselves in the observable behavior of every individual.[18]

These characteristics, then, that each normal woman (and man) of a race was supposed to share with every other woman (and man) together determined what we can call the *essence* of that race; they were characteristics that were necessary and sufficient, taken together, for someone to be a normal member of the race. Arnold's concept of race should, then, provide the materials for what I have called a strict criterial theory of the meaning of the term "race."

Arnold was uncharacteristic of his age in many ways: and one of them is the cosmopolitanism—or, at least, the Europeanism—of his temperament: he quotes frequently from French and German scholars. And on the question of race his views conformed with what was coming to be the common sense of Western European intellectuals.

Arnold's discussion in *On the Study of Celtic Literature* makes it plain that he believes that the racial essence accounts for more than the obvious visible characteristics of individuals and of groups—skin color, hair, shape of face—on the basis of which we decide whether people are, say, Asian- or Afro-Americans. For a racialist, then, to say someone is "Negro" is not just to say that they have inherited a black skin or curly hair: it is to say that their skin color goes along with other important inherited characteristics—including moral and literary endowments. By the end of the nineteenth century most Western scientists (indeed, most educated Westerners) believed that racialism was correct and theorists sought, to explain many characteristics—including, as we see here, the character of literatures—by supposing that they were inherited along with (or were in fact part of) a person's racial essence.

Arnold represents, then, a theory couched in terms of the new vocabulary of "race," whose authority derives, in part, from its association with the increasing prestige of the natural sciences. (In the Celtic literature lectures, Arnold uses the word "data" several times.) And the most important theoretical development in the growth of a biological conception of race had already occurred by the time Arnold published *Culture and Anarchy* in 1869. For on November 24, 1859, Charles Darwin had published a work whose full title reads: *The Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*.

The word "race" had been used in this way to refer to kinds of animals and plants, as well as to kinds of people, for some time; but there is no doubt that even for a mid-nineteenth-century ear this title promises something of relevance to the study of human difference. Indeed, the very fact that a single scientific theory promised to account for the variety of kinds of animals, in general, made its application to humans a natural step in the continuing process of placing the study of human anatomy in the context of a comparative zoology.

Darwin suggested, with characteristic caution, in *The Origin of Species*, that his theory might throw light on "the origin of man and his history"; the implication being that human beings developed, like other modern organisms, out of earlier forms. Taken to its "logical conclusion" this view suggested the oneness not only of all human beings—related by common descent—but, at least potentially, the common ancestry, and thus unity, of all life.

Darwin's theory can be thought of as consisting of two components: one is the claim that kinds of organisms develop by "descent with modification."[19] This claim was immediately widely accepted and applied to understanding the classification of organisms, representing, as it did, a continuation of arguments made five decades earlier years by Lamarck.

But Darwin's more distinctive claim was that the mechanism of modification was natural selection: the selective survival of characteristics that gave individuals advantages in the "struggle for life." Darwin here drew on the parallelism with artifical selection of animals that was carried on by horse and cattle breeders and by pigeon-fanciers. Just as they worked only with the natural variation among animals, selecting those with characteristics they favored and breeding from them, so, in Darwin's theory, nature "selected" organisms for breeding, not (as the rather colorful talk of the "struggle for life" suggested) by destroying some and allowing others to survive, but by affecting differentially rates of reproductive success.

This claim was not so easily accepted. To begin with it was not clear that there was sufficient variation within most kinds of organisms on which selection could work; and, indeed, though Darwin and Darwinians did stress the variability of natural populations, they had no account of the origin of the variations on which selection could act. More than this, most selective forces did not look as though they applied sufficient selection pressure to lead to any very substantial effects: it was only much later, with the development of population genetics, that it was possible to show that relatively small differences in survival rates could produce cumulatively large effects, given sufficient time.

And, finally, Darwin had an inadequate and undeveloped theory of inheritance: the modern account, in terms of the gene, had no real impact until after Mendel's work was rediscovered in 1900. The theory of evolution by natural selection required that organisms should inherit the characteristics of their ancestors: otherwise the surviving offspring of an organism with a trait that gave it an advantage on the struggle for life offered no guarantee that its children would carry the same trait. Indeed, since Darwin believed in a sort of blending theory of inheritance, in which what accounted for a particular observable characteristic was the blended mixture of the factors that determined that characteristic in ones parents, he could not really explain why a factor that was rare in a population could survive at all, since it would be constantly "diluted" by more common forms.

There were other problems: if you want to treat all creatures as derived from a single ancient population, there must be some source of new variations: otherwise every characteristic in any modern organism must have existed in the earliest population. It is thus only with the development of Mendelism, with its account of inheritance in terms of genes, and its recognition of the possibility of new variety arising by mutation, that the theory of natural selection was placed on a sound footing.

This second part of Darwin's theory—the view of natural selection—was thus rightly greeted with less immediate enthusiasm than the general idea of descent with modification.

Descent with modification was all that was required, however, to allow biology to give a much more straightforward account of how organisms should be classified. Darwin thought of species as essentially classificatory conveniences;[20] he was interested in how populations changed their character and separated from each other not in drawing boundaries between them. But his theory allowed that the accumulation of differences by selection could gradually produce kinds—varieties or species—that were measurably different; and thus suggested a mode of classification in which kinds that were more closely related by evolution should be classified together.

Thus, the general acceptance of descent with modification and the increasing acceptance of Darwin's theory of natural selection gave scientific support to the idea that human kinds—races—like animal and plant species could be both evolutionarily related and biologically distinct. Furthermore, even though human races were not mutually infertile, the theory of evolution suggested a way of thinking of varieties as being in the process of speciation: races might not be species, but they were, so to speak, moving in that direction.

Darwin, as I have said, thought of the species as essentially a classificatory convenience: he was, in philosophical jargon, a nominalist about species, holding that the boundaries between species were not clearly marked ''in nature''; and if species were not marked in nature then varieties or subspecies (which is what, on his view, human races were), being even less distinct from each other than species, were presumably classificatory conveniences also.

To believe this was already to move away from the sort of racial essences that we find in Arnold. For Arnold, the interest of the characteristics of a race was exactly that you could suppose that its members all shared certain properties; so that having identified a person's race membership from their appearance one could then make inferences about their moral or literary dispositions. It makes sense that Darwin, whose whole analysis depends on the recognition of variation within populations, was more interested in the ways individuals differed from each other within their varieties than in the ways they were similar.

Once we have the modern genetic picture we can see that each person is the product of enormous numbers of genetic characteristics, interacting with each other and an environment, and that there is nothing in the theory of evolution to guarantee that a group that shares one characteristic, will share all or even most others. Characteristics on different chromosomes are, as the Mendelians said, independently assorted. The theory of evolution will also predict that as you move through a geographical range along a gradient of selection pressure, the frequency of certain characteristics—those that affect skin color, for example—may change fairly continuously, so that populations may blend into each other; and characteristics may drift from one neighboring population into another over time by intermarriage (or, to speak less euphemistically, inter-breeding). Indeed, it turns out that, in humans, however you define the major races, the biological variability within them is almost as great as the biological variation within the species as a whole: put another way, while there are some characteristics that we are very good at recognizing—skin color, hair, skull shape—that are very unevenly geographically distributed, the groups produced by these assignments do not cluster much for other characteristics.

Even limiting oneself to the range of morphological criteria available to comparative anatomists it is hard to classify people objectively into a small set of populations; and whichever way you do it, it will turn out that, for biological purposes, your classification will contain almost as much human genetic variation as there is in the whole species.[21]

''Race,'' then, as a biological concept, picks out, at best, among humans, classes of people who share certain easily observable physical characteristics, most notably skin color and a few visible features of the face and head.

The materials for an evolutionary explanation for skin color variation are easily laid out. The original human population had dark skins which give you a selective advantage in the tropics, because they protect you somewhat from skin cancer. Lighter skins developed in colder climes, no doubt in part because skin cancer is less of a problem where you are permanently clothed, because of the cold, and the sun's rays pass more obliquely through the atmosphere. There may have been actual selection for white skins—melanin blocks the sun's rays, which make vitamins in the skin; so the less sun you see, the less melanin is good for you—or it may just be that the mutations that make for white skin developed and survived because there was no longer selection pressure against them.[22] And we may as well mention a third possibility here, one which Darwin noticed as well, which is that skin color was maintained by sexual selection: because, for some reason or other, human beings of one sex or other (or both) developed a preference for mates with lighter skins.

Why does biological variation in skin color not correlate more with other characteristics? Partly, because the other characteristics have been selected (as has, say, sickle-cell disease, in parts of West Africa and the Eastern Mediterranean) under pressures not highly correlated with the presence of harmful amounts of sunlight. Perhaps, too, because there are mechanisms that have evolved to maintain the stability of the genotype, reflecting, among other things, the fact that certain combinations of genes are adaptive only when they are present together.[23] As a result, even after long periods—of the order of hundreds of thousands of years—of geographical separation, human populations do not drift apart significantly with respect to most of their biological properties. And finally, because there has been continuous exchange of genes between the major geographical areas of human settlement over the hundreds of thousands of years since the first humans set off out of Africa.

The United States bears witness to the continuing significance of this phenomenon. It is true that Americans still tend, overwhelmingly, to marry people of their own, as we say, "racial identity." But very large numbers (perhaps as many as two-thirds) of African-Americans have some European forebears; up to two-fifths may have American Indian "blood"; and at least 5 percent of white Americans are thought to have African roots. It is estimated that 20 to 30 percent of the genes of the average African-American come from European and American Indian ancestors.[24] The result is that, even if the four roughly separated populations of the four continents from which the ancestors of most Americans came had each been much less genetically variable than was in fact the case, there would still be large numbers of people whose skin-color predicted very few other biological properties.

We have followed enough of the history of the race concept and said enough about current biological conceptions to answer, on both ideational and referential view, the question whether there are any races.

On the ideational view, the answer is easy. From Jefferson to Arnold, the idea of race has been used, in its application to humans, in such a way as to require that there be significant correlations between the biological and the moral, literary, or psychological characters of human beings; and that these be explained by the intrinsic nature (the "talents" and "faculties" in Jefferson; the "genius," in Arnold) of the members of the race.[25]

That has turned out not be true; the recent fuss generated by *The Bell Curve* about the correlation of race and IQ in the United States notwithstanding. Even if you believed Murray and Herrnstein's estimates of the heritability of IQ within groups in the United States—and you shouldn't—they offer almost no evidence relevant to refuting the claim that the differences between American groups are entirely caused by the environment; say, in particular, by the ways that blacks are treated in a racist society.[26]

Once you have the modern theory of inheritance, you can see why there is less correlation than everyone expected between skin-color and things we care about: people are the product not of essences but of genes interacting with each other and with environments and there is little systematic correlation between the genes that fix color and the like and the genes that shape courage or literary genius. So, to repeat, on the ideational view we can say that nothing in the world meets the criteria for being a Jeffersonian or an Arnoldian race.

The biological notion of race was meant to account only for a narrower range of characteristics, namely, the biological ones, by which I mean the ones important for biological theory. There are certainly many ways of classifying people for biological purposes: but there is no single way of doing so that is important for most biological purposes which corresponds, for example, to the majority populations of each continent or sub-continent. It follows that on an ideational view, there are no biological races either: not, in this case because nothing fits the loose criteria, but because too many things do.[27]

On the referential view we are required to find something in the world that best explains the history of usage of the term. Two candidates suggest themselves for the biological uses of "race": one is the concept of a population that I have been using for a while now. It can be defined as "the community of potentially interbreeding individuals at a given locality."[28] There are interesting discussions in the literature in population genetics as to how one should think about where to draw the boundaries of such communities: sometimes there is geographic isolation, which makes interbreeding in the normal course of things much less likely. But the population concept is generally used in such a way that we speak sometimes of a population defined by one geographical region and also, at other times, of a wider population, defined by a wider range, of which the first population is a part; and at yet other times of a populations that are overlapping.

I have no problem with people who want to use the word "race" in population genetics.[29] What Darwin was talking about—evolution, speciation, adaptation—can best be understood in terms of talk of populations. And the fact is that in many plants and animals there are, in fact, local populations that are reproductively isolated from each other, different in clustered and biologically interesting ways, and still capable of interbreeding if brought artificially together; and biologists both before and after Darwin could have called these "races." It's just that this doesn't happen in human beings. In this sense, there are biological races in some creatures, but not in us.

A more ecumenical proposal in this spirit would be to say that the word "race" refers to populations, more generally. The trouble is that, in this sense, while there are human populations that are and have been for some time relatively reproductively isolated, it is not all plausible to claim that any social sub-group in the United States is such a population. In this sense, then, there are human races, because there are human populations, in the geneticists' sense, but no large social group in America is a race. (The Amish, on the other hand, might come out as a race on this view, since they are a relatively reproductively isolated local population.)

A second candidate for the biological referent would simply be groups defined by skin color, hair and gross morphology, corresponding to the dominant pattern for these characteristics in the major sub-continental regions: Europe, Africa, East and South Asia, Australasia, the Americas, and, perhaps, the Pacific Islands. This grouping would encompass many human beings quite adequately and some not at all: but it is hard to see of what biological *interest* it would be, since we can study the skin and gross morphology separately, and there is, at any rate, a good deal of variation within all these areas, in skin, hair-color and the morphology of the skull. Certainly, this referent would not provide us with a concept that was central to biological thinking about human beings. And once more, in the United States, large numbers of people would not fit into any of these categories, because they are the products of mixtures (sometimes long ago) between people who do roughly fit this pattern, even though the social distinctions we call "racial" in the United States do, by contrast, cover almost everybody.[30] And, so, if we used this biological notion, it would have very little established correlation with any characteristics currently thought to be important for moral or social life.

The bottom line is this: you can't get much of a race-concept, ideationally speaking, from any of these traditions; you can get various possible candidates from the referential notion of meaning, but none of them will be much good for explaining social or psychological life, and none of them corresponds to the social groups we call "races" in America.

Notes

1. This essay is substantially abbreviated and slightly altered from the essay "Race, Culture, Identity: Misunderstood Connections," which appears in *The Tanner Lectures on Human Values*, Vol. 17 (Salt Lake City: University of Utah Press, 1996), pp. 51–136. (Another slightly different version of the longer arguments appears in *Color Conscious: The Political Morality of Race* [Princeton, NJ: Princeton University Press, 1996] with Amy Gutmann.)

2. See "Theories," in Frank Ramsey, *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, D. H. Mellor (ed.) (London: Routledge and Kegan Paul, 1978), pp. 101–125.

3. See P. F. Strawson, "Wittgenstein's conception of a criterion," in *Wittgenstein and the Problem of Other Minds*, Harold Morick (ed.) (Brighton, Sussex: Harvester Press, 1981.)

4. See Naomi Zack, *Race and Mixed Race* (Philadelphia: Temple University Press, 1993).

5. Strictly speaking, if there aren't any races, there's no talk or thought about races. So this is a shorthand for "talk they would assent to (or thoughts they would express) using the word 'race' and its cognates."

6. *Notes of the State of Virginia* (1781–82) in Thomas Jefferson, *Writings* (New York: Library of America, 1984), p. 264.

7. Jefferson, *Notes*, op. cit., p. 265.

8. Jefferson, *Notes*, op. cit., p. 206.

9. Jefferson, *Notes*, op. cit., p. 269.

10. Jefferson, *Notes*, op. cit., p. 270.

11. H. Grégoire, *De la littérature des Nègres: ou, recherches sur leurs facultés intellectuelles, leurs qualités morales et leur littérature: suivies des notices sur la vie et les ouvrages des Nègres qui se sont distingués dans les sciences, les lettres et les arts* (Paris: Maradan, 1808). Grégoire's full title displays the relevance of this work to Jefferson's theme in the *Notes*.

12. February 25, 1806, to Henri Grégoire, *Letters* in *Writings*, op. cit., p. 1202.

13. "The term 'biology' first appeared in a footnote in an obscure German medical publication of 1800. Two years later it again appeared, apparently independently, and was given ample publicity in treatises by a German naturalist (Gottfried Treviranus) and a French botanist turned zoologist (Jean-Baptiste de Lamarck)." William Coleman, *Biology in the Nineteenth Century: Problems of Form, Function and Transformation*, Cambridge History of Science Series (Cambridge: Cambridge University Press, 1971), p. 1.

14. Carolus Linnaeus, *Systema Naturae*, in which people are classified as *Homo sapiens*, appears in 1735.

15. Arnold, *Celtic Literature*, op. cit., pp. 66–67.

16. Arnold, *Celtic Literature*, op. cit., p. 72. Arnold never explicitly discusses sex, of course; and so we are left with the possibility of interpreting this as meaning that there are Englishmen who are of wholly British (i.e., Celtic) descent or thinking that there are some of partially British (i.e., Celtic) descent. Given, however, that some of the former have "passed" many centuries ago, the existence of the latter can be assumed.

17. Arnold, *Celtic Literature*, op. cit., p. 69.

18. Nevertheless, it is a point about the logic of dispositional terms that it is hard (though not impossible) to make sense of applying them to the members of a group if no one in the group ever

displays the disposition: see Anthony Appiah *Assertion and Conditionals* (Cambridge and New York: Cambridge University Press, 1985), Chapter 2, Section 4.

19. My account here is based on William Coleman, *Biology in the Nineteenth Century*, op. cit.

20. See George W. Stocking, *Race, Culture and Evolution* (New York: Free Press, 1968): "Darwin's own position on the question of human races was equally congenial to polygenist thinking. Although he thought it a matter of indifference whether human races were called species or subspecies, he granted that a naturalist confronted for the first time with specimens of Negro and European man would doubtless call them 'good and true species.'" p. 46.

21. "On average there's 0.2 percent difference in genetic material between any two randomly chosen people on Earth. Of that diversity, 85 percent will be found within any local group of people—say, between you and your neighbor. More than half (9 percent) of the remaining 15 percent will be represented by differences between ethnic and linguistic groups within a given race (for example, between Italians and French). Only 6 percent represents differences between races (for example, between Europeans and Asians). And remember that's 6 percent of 0.2 percent. In other words, race accounts for only a minuscule 0.012 percent difference in our genetic material." Paul Hoffman, "The Science of Race," *Discover*, November 1994, p. 4.

22. See Bernard R. Ortiz de Montellano, "Melanin, Afrocentricity and Pseudoscience," *Yearbook of Physical Anthropology*, Vol. 36, 1993, pp. 33–57.

23. Ernst Mayr, *Populations, Species and Evolution* (Cambridge: Harvard University Press, 1970), p. 300.

24. James Shreeve, "Terms of Estrangement," *Discover*, November 1994, p. 58. All these claims should be interpreted bearing in mind the fact that a "recent study found that in the early 1970s, 34 percent of the people participating in a census survey in two consecutive years changed racial groups from one year to the next," loc. cit.

25. That is, *not* produced by the fact that people who have certain physical appearances are treated in ways that produce differences.

26. Since this point is elementary it is perhaps worth explaining. Heritability measures the ratio of variance in a characteristic in an environment that is due to genes to the total variance. The heritability of height in the United States in India and in the human population in general is high. There is, too, a significant difference in average height between Indians (in India) and Americans (in America). But this inter-populational difference is almost entirely due to differences in nutrition. High heritability is quite consistent with most of the difference between populations being environmental.

Herrnstein and Murray, authors of *The Bell Curve* (New York: The Free Press, 1994), are aware of this fact and so seek to offer some rather unconvincing arguments for the suspicion that interracial average differences are in fact significantly genetic in origin. For arguments that they are *not* see Chapter 6 of Thomas Sowell's *Race and Culture: A World View* (New York: Basic Books, 1994).

27. This is essentially the point of Jared Diamond's essay "Race Without Color" in *Discover*, November 1994, pp. 82–89.

28. Mayr *Populations, Species and Evolution*, op. cit., p. 82.

29. I think, however, that this usage carries two risks: first, it gives an ill-deserved legitimacy to ideas that are mistaken, because those who listen in to these conversations may not be aware of the fact that the usage here does not correspond at all to the groups that have mostly been called races in Europe and America; second, because speaking this way, you can actually find yourself relying, illicitly, on those other modes of classification. Still, if you can avoid these two dangers, there's no problem.

30. Where a boundary is vague, profound consequences for which side of the boundary you fall on are bound to seem arbitrary. (Consider abortion, where it seems right to think that the burden of proof against the fetus diminishes with age, not that it suddenly disappears at the start, say, of the second trimester.) So those who take racial identity to have deep moral or political significance would be argumentatively better off if they could draw sharp lines between races.

23 A New Perspective on the Race Debate

Robin O. Andreasen

In the ongoing debate concerning the nature of human racial categories, there is a trend to reject the biological reality of race in favour of the view that races are social constructs. At work here is the assumption that biological reality and social constructivism are incompatible. I oppose the trend and the assumption by arguing that cladism, in conjunction with current work in human evolution, provides a new way to define race biologically. Defining race in this way makes sense when compared to the developments in other areas of systematic biology, where shared history has largely replaced morphological similarity as the foundation of a natural biological classification. Surprisingly, it turns out that cladistic races and social constructivism are compatible. I discuss a number of lessons about the way human biological races have been conceptualized.

> I was born in a century when the walls of race were clear and straight; when the world consisted of mutually exclusive races; and even though the edges might be blurred, there was no question of [the] exact definition and understanding of the meaning of the word. . . . [Of late], the concept of race has . . . changed and presented so much . . . contradiction that as I face Africa I ask myself what it is between us that constitutes a tie which I can feel better than I can explain?
>
> (W. E. B. Du Bois [1940], p. 116)

1 Introduction

Ashley Montagu led a lifelong campaign to rid science of the term "race." In 1964, he made the prediction "Race is the phlogiston of our time" (Montagu [1964] p. xii). Phlogiston, a substance believed to be given off during combustion, was once thought to be real. However, when Lavoisier determined the true nature of combustion, phlogiston turned out to be a mere chimera. Montagu believed that the concept of race should suffer a similar fate.

Most race theorists would say that Montagu's prediction has come true.[1] Although the person on the street may still believe that races are biologically real, science has

The British Journal for the Philosophy of Science 49, no. 2 (June 1998): 199–225. Published by Oxford University Press. Reprinted by permission of the Biometrika Trustees.

proven otherwise. Biological races are supposed to be "subspecies"—formal subdivisions of a species—yet according to most systematic biologists, the subspecies category is arbitrary and theoretically uninteresting. In addition, detailed work in human genetics purportedly reveals that, regardless of whether there are *non-human* subspecies, there are no *human* subspecies. A random sample of genes taken from different human populations classified by location reveals that these populations are too genetically similar to each other to justify dividing humans into races.

Although most race theorists think that races are biologically unreal, I disagree. Systematists and taxonomists have used two main approaches—typological and geographical—in their attempts to define subspecies. Neither of *these* work to define race. However, there is a third possibility that has largely been overlooked in the race literature—namely, the cladistic approach to subspecies. I will defend this account, arguing that races are ancestor–descendant sequences of breeding populations, or groups of such sequences, that share a common origin. Moreover, I will show that current work in human evolution supports this account and offers a new way of thinking about the biological reality of race. Races once existed, but they are on their way out. With the advent of the modern world came the intermixing of previously isolated populations and the gradual dissolution of racial distinctness. It isn't that science must recognize that race, like phlogiston, never existed; rather, human activity is causing race to lose its biological reality.

Once I have defended this new approach, I will describe three lessons that the cladistic concept provides about the current state of the race debate. Most race theorists treat biological races as static categories. From the non-existence of current races, they argue that biological races are, and always have been, illusions that we have projected on the world. The account that I offer, however, shows that biological races are dynamic categories. Second, because biological concepts of race are used repeatedly to justify belief in racial superiority, many have come to associate the biological reality of race with racism. Although this reaction is perfectly understandable, I will argue that these concepts need not go hand in hand, and that the cladistic concept, specifically, lends no support to claims about racial superiority. Third, those who reject the biological concept of race often think that race is a *social construct*—it is a product of our social practices. Although there are many ways of being a constructivist about race, most constructivists assume that their view is incompatible with the idea that races are biologically real. I will argue, however, that these conceptions can be complementary; they should not always be viewed as competitors.

2 Two Arguments Against the Biological Concept of Race

Today's race theorists cite two, purely biological, arguments to support their claim that race is not a biologically meaningful (or objective) category. In the first argument, they appeal to the practice of systematists. Most systematic biologists refuse to divide nonhuman species into subspecies on the grounds that the subspecies concept is theoretically meaningless. They add that if there is no justification for naming subspecies in biology, then there is no justification for dividing humans into biological races. The second argument allows that the subspecies concept might be useful in some contexts; here, theorists argue that there is a special reason why it should not be applied to humans.

Versions of both of these arguments have been presented before (Montagu [1941, 1959]; Barnicot [1964]; Ehrlich and Holm [1964]; Hiernaux [1964]; Livingstone [1964]), but for a long time they did not command general assent because the biological reality of race was thought to be "self-evident." However, the tide has turned; today, the dominant view in biology is that there are no biological races. As a result, the following arguments are now part of the mainstream in race theory.

2.1 The "No Subspecies" Argument

Race, as a biological concept, is synonymous with subspecies. Human races are subspecies of *Homo sapiens* and, like the term "race," "subspecies" has had a tumultuous history. Two systematic definitions have been offered—one typological, the other geographical—but neither is acceptable. Additionally, many biologists feel that there is little reason to search for a better definition. Race theorists conclude that we should learn a lesson from systematic biology. If biologists rarely use the subspecies concept to describe variation in nonhuman contexts, we should not use it in the case of humans.

2.1a Problems with the Typological Subspecies Concept In the spirit of Aristotle, subspecies were first defined as *types*—as natural kinds defined in terms of an *essential property* possessed by all and only the members of the same subspecies.[2] An attribute is "essential" to an object if it is a *necessary*, *intrinsic*, and *explanatory* property that an object must have in order to be the kind of thing it is.[3] For instance, it is a necessary truth that all samples of water have H_2O as their molecular structure.[4] This property is "intrinsic" because it depends for its instantiation only on the entity in which it is instantiated. It is also "non-accidental": nothing can be water and lack this structure and anything possessing this structure must be water. Finally, this property is "explanatory." It allows one to make inferences about other properties that things of this type possess. Being made of H_2O explains many phenomenal properties of water such as its being clear and tasteless.[5] Thus, the typological subspecies concept is the idea that subspecies ought to be objectively classified on the basis of a uniform association of characteristics, transmitted together due to the existence of an essential property.

Since there is extensive variability in nature, one rarely finds a single property possessed by all and only members of the same subspecies. Additionally, even if such a trait were to be found, it would probably not satisfy the modal and explanatory requirements demanded of an essential property. One might think that this is the straightforward reason why the typological subspecies concept is inadequate, but the essentialist has a way out. An Aristotelian essentialist can develop a ''natural state model'' in order to explain, actually explain away, nature's variability (Sober [1980]). According to Aristotle, every object has a ''natural state'' and a number of ''unnatural states.'' An unnatural state is a deviation from the natural state, which occurs when an object is subject to an interfering force. By appeal to this distinction, then, typologists can view variability as a deviation from certain natural tendencies. More specifically, defenders of the typological subspecies concept can recognize unlimited variety within, and continuity between, subspecies—so long as one can expect to find discrete natural tendencies underlying this variation. In their search for natural tendencies, typologists gather numerical data on the phenotypic characteristics of organisms within a population. They then compute the mean and the standard deviation in order to construct a bell curve. The peak of the curve (the average) might be taken to represent the ideal type, and the variance around that mean might represent deviation from type due to interfering forces.

The problem with the typological subspecies concept, and with typological thinking in biology, is that evolutionary theory no longer uses a natural-state model to understand biological diversity (Sober [1980]). When Darwin introduced his theory of evolution by natural selection, he introduced a new way of thinking, which Mayr [1959] calls ''population thinking.'' Populationists define taxonomic categories in terms of the phenotypic differences existing between *populations as a whole*. In contrast, typological thinking involves defining taxonomic categories in terms of the properties possessed by *individual organisms*.

Population thinking eliminates the demand for natural tendencies by providing a new way to account for variability in nature. In Sober's words ([1980], p. 176, emphasis in the original):

> Both typologists and populationists seek to transcend the blooming buzzing confusion of individual variation. Like all scientists, they do this by trying to identify properties of systems which remain constant in spite of the system's changes. For the typologist, the search for invariances takes the form of a search for natural tendencies. The typologist formulates a causal hypothesis about the forces at work on each individual within a population. The invariance underlying this diversity is the possession of a particular natural tendency *by each individual organism*. The populationist, on the other hand, tries to identify invariances by ascending to a different level of organization. For [the populationist], the invariant property across generations within a lineage is the amount of variability, and this is a property of populations.

Unlike typologists, populationists do not try to explain away nature's variability. Instead they treat it as real and work to explain the variation in one generation by appeal to variation in a previous generation and to the laws of heredity. This allows us to see why the typological subspecies concept won't work. Ever since Darwin, population models have played a central role in evolutionary theorizing; the ideas of type and of deviation from type at the species level, and at other taxonomic levels as well, do not feature in evolutionary laws and theories. The typological subspecies concept has no place in contemporary evolutionary biology.

2.1b Problems with the Geographical Subspecies Concept As one might expect, the typological subspecies concept was replaced with a population concept, which I will call the *geographical subspecies concept.* "Geographical subspecies" are morphologically distinct geographic representatives of a species. More formally, a geographical subspecies is an aggregate of phenotypically and genetically similar intraspecific populations, inhabiting a geographic subdivision of the range of that species, and differing significantly from other conspecific populations (Mayr and Ashlock [1991], p. 43). According to this definition, it is not necessary for different subspecies to differ absolutely; it suffices that there be statistically significant differences in the mean values of the characters used to define subspecies membership. Furthermore, membership must be defined with reference to many characteristics—enough to ensure that the addition of new characteristics will not alter subspecies groupings. If these conditions are not met, subspecies classifications will be arbitrary and theoretically uninteresting.

In contrast with those who support the typological subspecies concept, friends of the geographical concept deny that subspecies are natural kinds; instead they adopt a conventionalist stance. Subspecies are taken to be *conventional categories* that get individuated according to practical human interests. This does not mean, however, that geographical subspecies are arbitrary. On the contrary, for this concept to be biologically useful, some designations will have to be better than others. The meaningful designations will be the ones that allow biologically interesting generalizations to be formulated.

Over the years, there have been a number of criticisms of the geographical concept (Wilson and Brown [1953]; Gillham [1956]; Hagmeier [1958]; Mayr [1963, 1982]; Johnston and Selander [1964, 1966]; Storer [1982]; Zusi [1982]); most of them point to the fact that intraspecific variation is often not discrete. It is generally gradual across geographic regions (clinal) and it often ranges in many different directions (discordant). For example, a bird species might gradually increase in size from the southern to the northern regions of its habitat. At the same time, a different pattern of variation, say from light to dark, might range across the east–west regions of the habitat. When intraspecific variation displays such a pattern, there is no principled way to designate geographical subspecies. Designations would be mere subjective partitions of continuous variability. They would be neither stable nor repeatable, nor theoretically interesting.

Although I think that biologists have come to the right conclusion—namely, that geographical subspecies are arbitrary and unreal—they have come to this conclusion for the wrong reason. The above argument demands that there be a precise boundary between different subspecies, but this is unreasonable. It is like demanding that there be a precise line of demarcation between baldness and having a full head of hair, or between being rich and being poor. Just because there are line-drawing problems in these cases does not mean that these properties (wealth and baldness) should be rejected as arbitrary. Similarly, geographical subspecies might be real, even if the boundaries between them are vague.

The real problem with the geographical concept is that it is a *phenetic* concept. Pheneticists define taxonomic groupings based upon overall similarity. For example, a phenetic subspecies is a set of phenotypically and genetically similar conspecific organisms that are distinct from other such sets. Perhaps the worst problem with pheneticism is that there is no reason to believe that the resultant taxa reflect any important biological phenomena. Pheneticism defines taxa using similarity *alone*, but there is no reason to believe that overall similarity represents an objective (or theoretically interesting) feature of reality. A second problem with this concept is that there is rarely a unique way to describe the similarities and differences between pairs of taxa. For example, a pheneticist might attempt to classify a set of organisms into subspecies by pointing to a number of characteristics possessed by one group, but not by the others. However, for each of these characteristics, it is possible to describe others that entail quite different similarity groupings. Finally, the concept of similarity itself can be spelled out in different ways, which further augments pheneticism's embarrassment of riches (Ridley [1986]).

Systematists once thought it possible to provide a biologically meaningful definition of "subspecies." Today however, many biologists question this assumption (Wilson and Brown [1953]; Mayr [1963, 1982]; Johnston and Selander [1964, 1966]; Storer [1982]; Zusi [1982]). Some argue that, due to the dynamic nature of intraspecific variation, it will be difficult to come up with a biologically meaningful subspecies concept. Designating subspecies entails identifying *distinct units* and giving them formal names. Yet the boundaries between "subspecies" are rarely fixed or definite. Others argue that the subspecies concept is superfluous; biologists can use multivariate analysis to study the clinal and discordant variation within a single species. As a result, it is generally accepted that the subspecies category is (at best) a unit of taxonomic convenience. Race theorists conclude that the implications for human race ought to the obvious. If the subspecies concept is inadequate for defining non-human subspecies, we should not use it to divide humans into biological races.

2.2 The "No Human Subspecies" Argument

The second argument against the biological concept of race is a special case of the first. Race theorists who present this argument are agnostic about the overall value of the subspecies category, allowing that it might result in biologically meaningful designations in some non-human contexts. What they argue is that when the geographical concept gets applied to humans, the result is a number of biologically insignificant "racial" groupings. Thus, even if there are non-human subspecies, there are no *human* subspecies.

Over the past fifty years, geneticists have been gathering copious data on the genetics of contemporary human populations in order to measure the genetic differences among individuals, populations, and "races."[6] Studies indicate that the variation within major "races" is slightly greater than the variation between local populations. Moreover, the variation between individuals within a population is substantially greater than the variation between populations or "races." Approximately 85% of human genetic variation is between individuals within the same local populations. About 8% is between the local populations found within major "racial" groups, and the remaining 7% is between "races" (Lewontin [1972]; Nei and Roychoudhury [1972; 1974]; Cavalli-Sforza [1974]; Lewontin, Rose, and Kamin [1984]; Cavalli-Sforza, Menozzi, and Piazza [1994]). Using these statistics, race theorists argue that there can be no justification for recognizing human geographical races. Populations are clustered so closely together that any partitioning into races would be merely subjective.

Like the general argument against the biological reality of subspecies, this argument does not deny the existence of human variation. It merely claims that racial classification is not the best way to understand such variation. At first sight, this argument provides a strong reason to reject the biological concept of race—especially when it is coupled with the previous argument. Nevertheless, I will show that appearances are deceiving; these arguments are inconclusive.

3 Why Not Cladism?

Biologists have often been too quick to infer the general failure of the subspecies concept, and most race theorists have been too quick to reject the biological reality of human race. In this section, I will argue that there is a third option for defining subspecies that has been largely overlooked in the race literature. Using the principles of cladistic classification, we can think of subspecies as (groups of) ancestor-descendant sequences of breeding populations that share a common origin. I will call this the *cladistic approach* to subspecies. Additionally, by appeal to current work in human evolution, I will show that this account can be applied successfully to humans. This will provide a new way to define race that is untouched by the above arguments.

The philosophical debate concerning the status of scientific categories provides a background for my argument. Given the human propensity to divide the world into different sorts of things—gold, humans, bachelors—it is natural to ask which of these categories are artificial kinds, invented by us, and which are natural kinds, corresponding to real divisions in nature. Two answers to this question have been prominent in contemporary philosophy. Essentialists claim that natural kinds are defined in terms of essential properties possessed by all and only members of a kind (Kripke [1972]; Putnam [1975]). Conventionalists, on the other hand, argue that all categories are individuated according to practical human interests (Dewey [1938]; Lewis [1946]; De Sousa [1984]; Sidelle [1989]). Some categories may be more interesting than others, but all categories have the same metaphysical status. However, when it comes to certain categories—namely, evolutionary ones—the possibility arises that neither position is adequate. Both of the above philosophies of classification hold that classifications must be based on the idea of similarity; typological classification requires a shared essential property and the geographical concept requires overall similarity. There is a third possibility, however, that is present in the biological literature—namely, that a natural classification is one that reflects the genealogical relationships among organisms (Darwin [1859]; Hennig [1966]; Ghiselin [1974]; Hull [1978]; Sober [1988]). Darwin was among the first to make this suggestion, but Hennig is the one who gave it rigour.[7]

Hennig is the founder of the systematic school called *cladism* which individuates taxa in terms of common ancestry. By organizing sets of well-defined species into a branching structure, a *phylogenetic tree*, one can depict the relationships between ancestors and their descendants and classify organisms into higher taxa (anything above the species level).[8] (Later I will discuss how this taxonomic philosophy can be extended to lower taxonomic categories such as subspecies.) For example, figure 23.1 represents an ancestral species, A, giving rise to two daughter species, B and C, which then eventually give rise to the terminal taxa H-L as depicted.

In this diagram, the nodes represent the species whose evolutionary relationships are depicted and the branches represent speciation events. Provided that each object in the tree has a unique immediate ancestor—that is, provided that there is branching without reticulation—we can define a taxonomic unit as a *monophyletic unit*, a group composed of an ancestor and all of its descendants. Sober [1993] applies what he calls "the cut method" to explain the concept of monophyly. If you draw a cut across any branch, the nodes immediately above that cut comprise a monophyletic group. For example, in figure 23.1, E is a monophyletic group, so is DHIJ, and so are many other groupings.

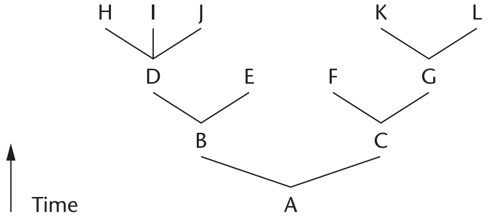


Figure 23.1

Two facts about monophyly are worth noting. The first is that the complement of a monophyletic group is not itself a monophyletic group. By applying the cut method to figure 23.1, it is possible to see that DHIJ is monophyletic, but that the rest of the tree (ABCEFGKL) is not. Furthermore, monophyly is a property of a *bifurcating tree* (a point that will have considerable importance later on). Although the concept of monophyly *can* be applied to a reticulate structure, such as the one depicted in figure 23.2, it results in a partial overlap between monophyletic groups. Again the cut method is useful for seeing why this is so.

Applying the cut method to this reticulate structure produces two groups (BDEH and CFGH), both of which contain species H. Most systematists choose not to develop classifications in such cases. The reason is that when reticulation is extensive, partial overlap will also be extensive, resulting in a nonhierarchical classification scheme.

Cladistic classifications have both a conventional and an objective aspect. The way that monophyletic groups get assigned to a taxonomic level is conventional. According to Henning the smallest terminal taxa—for example, H, I, J, E, F, K, and L in figure 23.1—represent current species; the next largest monophyletic units—DHIJ and GKL—represent genera, and so on up the taxonomic hierarchy.[9] Yet there is nothing to keep us from arranging things differently. For example, one might decide that DHIJ and GKL comprise genera, or that they make up families. There is no fact of the matter at issue here. This conventional aspect, however, should not obscure the fact that the branching structure and the nested hierarchy of monophyletic groups exist objectively. They exist objectively because they reflect the patterns and processes of evolution, which are themselves objective. It is this that makes cladistic classifications objective. We may not always know when we have the right phylogeny, but when we do, the resulting classification will reflect something that exists independently of our classifying activities.

We are now in a position to see why we ought to consider a cladistic view of subspecies. Defenders of the typological approach embrace essentialism and defenders of the geographical approach adopt a conventionalist stance, yet discussions about the biological reality of subspecies have taken place without seriously considering the cladistic approach. Since this third possibility has been immensely important when it

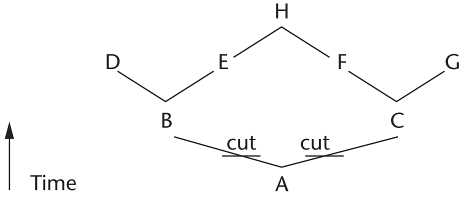


Figure 23.2

comes to defining other categories in the taxonomic hierarchy and, as I am about to demonstrate, it can be adapted for defining subspecies, it is premature to conclude that subspecies are biologically meaningless.

3.1 Cladistic Subspecies

A cladistic view of subspecies would require constructing a phylogenetic tree out of the *breeding populations* in a single species. A ''breeding population'' is a set of local populations linked to one another by reproductive ties that are, for the most part, reproductively isolated from other such populations. For example, a gaggle of geese living in Vilas Park constitutes a local population. When there is interbreeding between this population and other local populations due to migration, these local populations constitute a breeding population. In our tree, then, the nodes would represent breeding populations and the branches would represent the birth of new breeding populations. A breeding population is born when a local subpopulation becomes separated from its parent population and is, for the most part, reproductively isolated from the parent population. Suppose that a few rabbits get separated from the breeding population to which they belong due to a river's changing course. On the assumption that the rabbits cannot cross the river, this constitutes the birth of a new breeding population. Referring again to figure 23.1, we can define subspecies in the following manner. The terminal taxa (H, I, J, E, F, K, and L) represent current breeding populations, the more inclusive monophyletic units (DHIJ, GKL, BEDHIJ, and CFGKL) represent a nested hierarchy of subspecies, and the whole tree represents the species.

There are two reasons why biologists have, by and large, ignored the possibility of defining subspecies cladistically. The first is mere force of habit—cladistic classification is traditionally used for defining membership of higher taxa; to apply it to subspecies would deviate from Hennig's original intention. Second, many authors think that tree reconstruction is too difficult for lower taxonomic levels (Bremer and Wanntorp [1979a, b]; Arnold [1981]). They argue that cladism is appropriately applied whenever groups are diverging. Two groups are likely to diverge when, over long periods of time, there is limited gene flow between them. But in the case of subspecies, divergence rarely occurs because there is often considerable interbreeding between local populations. I think that this point is overstated. Many commonly recognized sexual species have subpopulations between which there is little or no genetic exchange (Ehrlich and Raven [1969]). Furthermore, low levels of interbreeding can be allowed; interbreeding is only a problem when it is extensive enough to cause reticulation. Third, systematists and human evolutionists have come up with methods for estimating degrees of interbreeding, and hence can work around some of the problems that it poses (Nei and Roychoudhury [1993]; Templeton [1993]; Cavalli-Sforza, Menozzi, and Piazza [1994]). Finally, current work in human evolution illustrates that it is possible to reconstruct a phylogenetic tree of human breeding populations—and this means that it is possible to apply cladistic classification below the species level. It is to this possibility that I now turn.

4 Races as Clades

Reconstructing a human phylogenetic tree has been a goal of human evolutionists for some time, but until recently this was only a remote possibility. Early attempts at tree reconstruction were, for the most part, unsuccessful because they were based solely on palaeontological and archaeological data (mainly fossilized bones and artefacts). Due to the incompleteness of the fossil record and the limitations of early dating techniques, these data were met with scepticism. Today, however, the prospects of reconstructing a human phylogeny are quite promising. Not only have dating techniques improved, but new fossil evidence is being discovered quite rapidly. More importantly, however, recent developments in human genetics provide a new and independent source of data. This new source, in conjunction with improved palaeontological and archaeological data, is allowing human evolutionists to approach their goal.

A chief tool for reconstructing human phylogenies is a quantitative measure called *genetic distance*. This is a measure of the difference in gene frequencies between two breeding populations. In this context, it is used for estimating degrees of relatedness between human populations. Roughly, the smaller the genetic distance between two populations, the closer their ancestral relation. The reasoning here is that all humans share a common origin; we all evolved from a single ancestral population at some point in our distant past. Thus as local subpopulations migrated out of their original location and formed new breeding populations, mutations occurred and differences accumulated. Genetic distance, therefore, serves as a means by which to calibrate human evolutionary history.

We should be aware, however, that the assumed correlation between time and genetic distance is only rough; a number of factors can disturb it, resulting in an imperfectly reconstructed tree. When two previously distinct populations live in similar environments, natural selection may cause them to converge (i.e. to develop similarities). The effect of convergence on tree reconstruction is an over-estimation of their degree of relatedness. Secondly, rapid genetic drift in small populations will cause them to evolve more rapidly than larger populations. In such cases, smaller populations appear to be older (more divergent) than they in fact are. Finally, interbreeding between two previously isolated populations can occur, resulting in an overestimation of their degree of relatedness.

These difficulties, however, are not insurmountable. Geneticists can minimize potential biases in a number of ways. They address the problem of convergence by using selectively neutral DNA segments, such as "junk DNA" (DNA that serves no apparent function) or mitochondrial DNA. Secondly, rates of evolution are likely to be uniform when drift is a major cause of change and when populations are roughly the same size on average. Independent evidence confirms the former assumption and careful selection of populations makes the latter probable (Cavalli-Sforza [1991], p. 105). Furthermore, geneticists can greatly reduce the problems posed by hybridization by studying *aboriginal populations*—breeding populations that occupied their present location before the great migratory waves that began with the voyages of discovery in the mid- to late fifteenth century. A fourth source of confirmation comes from agreement with historical, archaeological, and palaeontological data. Finally, and perhaps most importantly, biases can be ''washed out'' by averaging over many genes in calculating genetic distance. The greater the number of measured differences, the more confidence one can have that a tree is unbiased. Fortunately, thousands of genes are known.

For the past fifty years, geneticists have been gathering extensive data on the genetics of living populations (Nei and Roychoudhury [1982]; Cavalli-Sforza *et al.* [1988]; Cavalli-Sforza [1991]; Vigilant *et al.* [1991]; Wilson and Cann [1992]; Cavalli-Sforza, Menozzi, and Piazza [1994]). Their purpose is to infer major patterns of human evolution. For example, Cavalli-Sforza and colleagues calculated the genetic distances between 120 different gene states for forty-two aboriginal populations. Using these data, they inferred the phylogenetic tree depicted in figure 23.3.

Cavalli-Sforza and colleagues tested the accuracy of their phylogeny by comparing measured genetic distances with a number of widely accepted dates suggested by the geological record. The genetic distances between Africans and non-Africans exceeded all other measures. Moreover, this distance was approximately twice that between Australians and South East Asians, and the latter was more than twice that between Europeans and North East Asians (Cavalli-Sforza [1991], p. 106). The corresponding times of

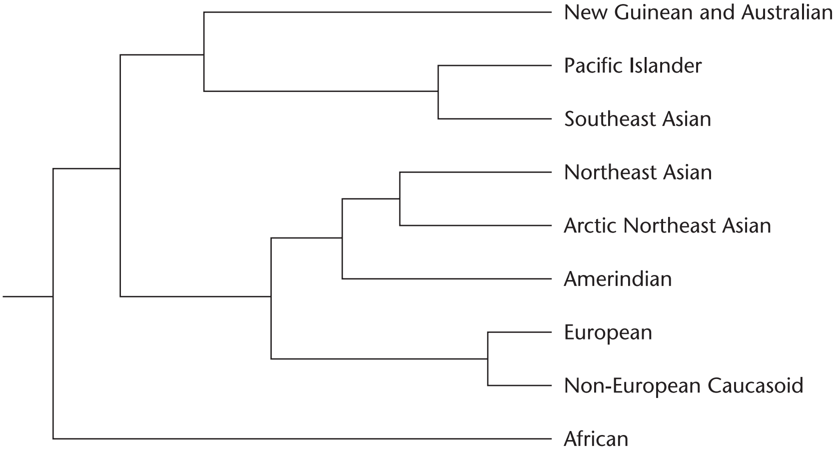


Figure 23.3

separation suggested by paleoanthropology are in similar ratios. Archaeological and palaeontological data indicate that anatomically modern humans evolved in Africa ~200,000 years ago. Thus, the first major split—which separates Africans from all other groups—represents a racially undifferentiated stock of *Homo sapiens* migrating out of Africa ~100,000 years ago. The second split separates Pacific and South East Asians from all other non-Africans. The breeding populations of native Australia and Papua New Guinea reached their location ~50,000 years ago. Within this supercluster, there was a later separation between the Pacific Islanders and the South East Asians. Prior to this split, however, there was a division in the third major branch; the North Eurasians separated from the Caucasians. An approximate date for this event is 40,000 years ago. Finally, a series of divisions occurred in the North Eurasian supercluster.

An interesting result comes from applying cladistic classification to Cavalli-Sforza's tree. People standardly divide humans into three (or more) major races—Africans, Caucasians, Asians. The cladistic concept of race, however, results in racial categories that cross-classify these standard groupings. More specifically, the folk category "Asian" is not a cladistic race. We can see why by looking at the nested hierarchy of monophyletic groups in figure 23.3. Caucasian and African are monophyletic groups, but Asian is not. South East and North East Asians are in two distinct major branches. North East Asians are more closely related to Amerindians and to Caucasians than they are to South East Asians. Similarly, South East Asians are more closely related to Australians than to North East Asians. This conclusion is interesting because it illustrates that the existence of biological races does not depend upon our folk taxonomy being right.

The above results, although fascinating, are somewhat controversial. For example, Cavalli-Sforza (Cavalli-Sforza, Menozzi, and Piazza [1994]) denies that his phylogeny can be used to define race. He relies on the conventional aspect of cladistic classification to support his claim: human phylogenies provide a nested hierarchy of monophyletic groups, and there is no biological reason to apply the term "race" at one level rather than others. However, there is no need to choose a unique level of monophyly that defines racial categories—we can simply define "race" as a nested hierarchy of monophyletic groups. In addition, Cavalli-Sforza's point applies to races no more than it applies to other monophyletic groups in a cladistic hierarchy. For these reasons, we should not accept this argument against the biological reality of races.

Secondly, a number of theorists have criticized Cavalli-Sforza for using phenetic methods—specifically, Nei's genetic distance—to make phylogenetic inferences (O'Grady *et al.* [1989]; Bateman *et al.* [1990a, b]; Bayard [1990]); they argue that cladistic methods provide a better way to measure propinquity of descent. I will address this issue, as well as a number of others, at the end of this section. At present, however, my concern is to address a different matter. This criticism appears to contradict my proposal that we use *cladistic* classification to define race, since it asserts that Cavalli-Sforza uses phenetic rather than cladistic methods of phylogenetic inference. To understand why this contradiction is only apparent, it is important not to confuse the problem of *phylogenetic inference* (which concerns how one is to infer what the phylogenetic tree is for a given set of taxa) with the problem of *classification* (which concerns how one is to organize sets of organisms into taxa (Felsenstein [1984])).[10] There is no contradiction between using phenetic methods for reconstructing trees and using cladistic methods for constructing a classification scheme (Sober [1993]).

Finally, in addition to the controversy surrounding Cavalli-Sforza's methods of phylogenetic inference, there is some controversy over the specifics of his tree. For example, although Cavalli-Sforza's results agree with one line of fossil evidence, there is some disagreement with other palaeoanthropological data (Stringer [1990]). Additionally, other research groups have inferred phylogenies that show a somewhat different pattern of migration and subsequent divergence (Nei and Roychoudhury [1982, 1993]; Vigilant *et al.* [1991]; Wilson and Cann [1992]). For example, Wilson's group obtained a phylogeny in which Asian is monophyletic, but African is not. Nei and his colleagues, on the other hand, found that all three major races are monophyletic. Finally, Cavalli-Sforza's tree supports a hypothesis (the "out-of-Africa" hypothesis) that sees racially undifferentiated modern humans as evolving in Africa about 200,000 years ago and subsequently spreading around the world. Under this hypothesis, racial differentiation occurs after the initial migration out of Africa as a result of incomplete but effective reproductive isolation among groups. Some human evolutionists, however, reject this hypothesis (Wolpoff, Wu, and Thorne [1984]; Wolpoff *et al.* [1988]; Wolpoff [1989a, b]; Thorne and Wolpoff [1992]). Instead, they defend what is called the "multiregional evolution" hypothesis. This theory holds that, as a result of genetic continuity over time and gene flow among contemporaneous populations, modern humans evolved not only in Africa but also in Europe and Asia from their already racially differentiated *Homo erectus* forebears.

In spite of the issues that remain to be resolved, we should not lose sight of the great importance of this research. Cavalli-Sforza's work illustrates that human evolutionists are approaching their goal; they are on their way towards reconstructing a human family tree that accurately reflects the patterns and processes of human evolution. This carries with it important implications for the race debate. It means that it is possible to give a biologically objective definition of race. Races are monophyletic groups; they are ancestor-descendant sequences of breeding populations, or groups of such sequences, that share a common origin. Even if the empirical details change, this conceptual point—as well as others that I am about to make—will remain in place.

5 Lessons

I have just argued that cladistic classification, in conjunction with current work in human evolution, vindicates the biological reality of race. My aim in this section is to discuss the impact that this has on the current state of the race debate. Specifically, I will discuss three lessons that the cladistic view provides about the way race theorists have traditionally understood the biological concept of race.

5.1 Biological Races Are Dynamic

Current race theorists tend to think of biological races as static categories. The presupposition is that biological races have either always existed or they have never existed. Although this type of thinking is rampant in the race literature, it is best illustrated by Montagu's prediction. Montagu thought that biological races don't exist and that they never have. Racial categories are merely convenient fictions invented by humans in order to render intelligible their observations of human differences.

The cladistic concept of race, however, shows that biological races are dynamic categories; races once existed, but due to recent historical events, they are on their way out. As noted in Section 3, cladistic classification requires that evolution take the form of a branching process. Subspecific evolution takes this form when a species splits into several breeding populations that experience different evolutionary forces under a significant degree of reproductive isolation. Current work in human evolution illustrates that this condition was met in the past. Specifically, the genetic data used to reconstruct phylogenetic trees indicates that Old World human populations had low levels of genetic contact for a substantial portion of time (Nei and Roychoudhury [1993]; Templeton [1993]; Cavalli-Sforza, Menozzi, and Piazza [1994]). Further support comes from the fossil record. Had there been substantial interbreeding among ancestral populations, the fossil remains of our ancestors would show significantly more continuity than they in fact show (Stringer and Andrews [1988]; Stringer [1990]; Aiello [1993]; Shreeve [1995]). From this evidence, and from Cavalli-Sforza's tree, we can conclude that races once existed.

Yet these newly reconstructed phylogenetic trees do not indicate the existence of current races. They describe racial ancestry—this is why human evolutionists focus on aboriginal populations (such as the Australian aborigines who, even today, remain reproductively isolated to a very high degree) and attempt to control for any outbreeding that has occurred in more modern times. The importance of this point can be seen by returning to the ''no human subspecies'' argument. Unlike the data used to reconstruct phylogenetic trees, the data used in this argument come from modern (i.e. non-aboriginal) populations. For example, when measuring the frequencies of various gene states, Lewontin used black and white Londoners and Nei and Roychoudhury used African and Caucasian Americans as representatives of the African and Caucasian ''races'' (Nei and Roychoudhury [1972]; Lewontin, Rose, and Kamin [1984]). Additionally, there were no attempts to control for outbreeding when these studies were done. These data, in conjunction with the fact that there has been reduced reproductive isolation in modern times, indicate that races are fading out of existence. Ever since the voyages of discovery, colonization and immigration have been blurring racial distinctness.

Thus, if we focus on the synchronic question—is there any justification for dividing current populations into races—the answer may very well be "no." The boundaries between races are becoming blurry due to a lack of reproductive isolation. This is, or at least should be, the upshot of the "no human subspecies" argument. But we should not conclude from this, as Lewontin and others have done, that race is merely an illusion that we have projected on the world. The cladistic concept focuses on a diachronic question about the biological reality of race; it seeks to define race via evolutionary history. If we focus on this type of question, a better conclusion to draw is that *ancestral races* existed, but biological races are anastomosing.

5.2 The Relation Between Biological Races and Racism

Anyone who attempts to argue that races are biologically objective should do so with great trepidation. Appeals to the biological reality of race have been used repeatedly to justify the belief that some races are biologically superior to others; this, in turn, gets used to justify oppressive social practices. As a result, people have come to associate any talk of the biological objectivity of race with racism. One particular concern is that some might think that Cavalli-Sforza's diagram—with Africans splitting off first and Caucasians last—is itself racist. Does the diagram entail that Africans are the most "primitive" of races and that Caucasians are the most "advanced"? The answer is "no." Cavalli-Sforza's tree tells us nothing about the relative "values" of different races; this is so for two reasons.

Claims about biological objectivity entail claims about biological difference, but they do not justify conclusions about racial superiority. As Hume taught us, one cannot infer normative conclusions from purely descriptive premises. It follows that inferences from biological difference to biological superiority are fallacious. Statements about biological differences are descriptive; they are statements of empirical fact. Assertions of racial superiority are normative claims that are born out of social and political motives. They are a result of imposing a value system upon the fact of biological variation, and this value system has no intrinsic relationship to biological diversity itself. In Cavalli-Sforza's words, "[p]olitical convenience, and a variety of motives totally unconnected with science are the basis of racism" (Cavalli-Sforza, Menozzi, and Piazza [1994], p. 19).

In addition, even if one could sometimes draw conclusions about the relative values of different races from facts about biological differences, Cavalli-Sforza's tree does not support such an inference. As I argued above, this tree represents facts about racial ancestry. Specifically, it illustrates that in our distant past, there was little genetic contact between human breeding populations. As a result, we can conclude that ancestral races existed. However, as I also argued above, the current situation is much less clear cut.

With the advent of the modern world came an explosion of migration resulting in a great increase of interbreeding among previously isolated populations. Consequently, the boundaries between current human populations are becoming increasingly fuzzy. Given this, it makes no sense to use this concept to argue for biological superiority. If it is unclear that races exist today or in our recent past, and I have argued that it is, surely this fact provides no basis for drawing racist conclusions.

5.3 The Relation Between the Cladistic Account and Constructivist Accounts

Those who reject the biological reality of race often think that race is a *social construct*—it is a product of our social practices. Although it is often assumed that "social constructivism" has a single well-understood meaning, this is not the case. Constructivists often use the metaphor in a variety of ways, and rarely distinguish between its many uses. For example, although most forms of constructivism oppose realism about natural kinds, this opposition can be either *local* or *global*.[11] "Local constructivists" accept that some kinds might be natural; they merely deny that the category in question is a natural kind. "Global constructivists," however, deny that any kind is natural. At this point, one can draw a distinction between *causal* and *conceptual* forms of constructivism. "Conceptual constructivism" is the idea that some (or all) categories conceptually depend for their existence and features on the way that human inquirers think about things. This thesis is most often formulated as a global thesis that is intended to apply to *all* categories (see, for example, Goodman [1978]; Latour and Woolgar [1979]; and Woolgar [1988]), yet it also can be stated as a local thesis about some limited set of categories (see, for example, Appiah [1986, 1992, 1996] and Goldberg [1993]).[12] In contrast with conceptual constructivism, "causal constructivism" is usually formulated as a local realist thesis; it holds that humans can create socially real kinds as a result of their causal interactions with the world. For example, some constructivists argue that the act of classifying people can cause these classifications to become real and causally meaningful (McIntosh [1968]; Foucault [1978]; Hacking [1986, 1988, 1991, 1992]). Human kinds become real when the act of classifying people influences their self-understandings and behaviours, causing them to act in conformity with a label. To see the difference between these two constructivisms, consider the following example. A conceptual constructivist might hold that the existence of individual dinosaurs, or of dinosaur taxa, depends on how we think about the world; the meaning of the term "dinosaur" inevitably involves considerations that have to do with human capacities. In contrast, a causal constructivist would never say that dinosaurs depend for their existence and features on us—at least not if dinosaurs existed before human beings started to interact with the world.

Using this brief taxonomy of constructivisms, we can better understand what it means to say that race is a social construct. Race constructivism is most often (and most sensibly) formulated as a *local* thesis; race constructivists deny that race is a biologically real category, while allowing the biological reality of other categories (such as species). Moreover, although most race theorists agree that races are not biologically real, they disagree over the metaphysical implications of this. Appiah [1986, 1992, 1996] and Goldberg [1993], for example, argue that races are neither biologically real nor socially real. For them, races are *conceptual constructs*—they are merely a product of the way that we think about human differences. Many race theorists, however, find this view implausible. They argue that race is a central element in many people's identities; it also plays a prominent role in how people identify and treat others. For these theorists races are *causal constructs*; as a result of the way that people treat others in the name of race, races have become socially real categories (on a par with categories such as marital status, class, and religion). Defenders of this view hold that race ought to be defined in terms of socially normed biological and cultural factors (Du Bois [1940]; Omi and Winant [1994]; Outlaw [1995]).

The third lesson that I want to draw from the cladistic concept of human race is that it poses no threat to the constructivist project; in fact, there is a sense in which the two conceptions are complementary. The reader is probably wondering how this could be: if the cladistic approach provides a biologically objective definition of race and constructivists deny the biological reality of race, how can the two coexist? The answer is that the cladistic account falls outside the constructivist's domain of enquiry. Questions about biological classification can be about ordinary language classifications, or they can be about scientific classifications. For example, the question "is there a biologically objective way to define race" could be asking whether biology vindicates our common-sense notions of race. Alternatively, it could be asking whether there are *any* biologically objective ways to divide humans into races. The second question is more general than the first; it accepts the possibility that our conventional racial categories may be unjustified while allowing that biological races might still exist. As it turns out, constructivists are interested in the first question only. I, however, am using the cladistic account to address the second question. Moreover, since the cladistic account supports the biological reality of human race without vindicating popular conceptions, it is not in competition with constructivist accounts.

Evidence that constructivists are interested in the first question and not the second comes from two sources. First, there is the practice of constructivists; as a matter of fact constructivists have focused their studies on popular racial categories. Second, there is the constructivist project. Constructivists are concerned with the sociology of race and race relations. In order to make sense of the problems of interracial conflict, they seek to understand the role that the race concept has played in modern Western thought and experience, and this understanding is gained by critically examining popular conceptions of race. Specifically, constructivists are interested in the commonly held belief that there are at least three biological races (Caucasians, Africans, and Asians) that differ significantly in their morphological, behavioural, and intellectual characteristics.

Their concern with this conception stems from their belief that it forms the core of many common-sense notions about race. For example, it is (unfortunately) often only a short step from beliefs about morphological, behavioural, and intellectual traits to beliefs about biologically based racial superiority.

When constructivists deny the biological reality of race, they are denying that biology vindicates our common sense notions; they are not disagreeing with the conclusions generated from the cladistic account. We learn two main lessons from applying cladistic classification to Cavalli-Sforza's human family tree: firstly, biological races once existed, and the conditions necessary for maintaining racial distinctness no longer exist; and secondly, what subdivisions there are in the human species are being rapidly diminished by extensive outbreeding. Neither of these lessons helps constructivists achieve their goals. Constructivists are interested in the impact of popular biological conceptions on human behaviour. Yet scientists did not seek scientific backing for popular conceptions of race until around the seventeenth century (Banton and Harwood [1975]). By this time, biological races were already on the road to obsolescence. Moreover, it is likely that if current races *do* exist, they cross-classify folk racial categories. For example, most people believe that African-Americans are more closely related to the Bushmen than they are to Caucasian-Americans. The cladistic account, however, would lead to the opposite conclusion. The reason is that the Bushmen are aboriginal populations that have remained largely reproductively isolated (even in more modern times). African-Americans and Caucasian-Americans, on the other hand, have not been reproductively isolated. Hence, the latter two groups would probably be in the *reticulated* part of the tree, while the Bushmen might get their own distinct branch. This constitutes further evidence that popular notions of race are indeed social constructions without any basis in biological fact.

By clarifying the ambiguity in questions about the existence of biological races, it is possible to appreciate the importance of both conceptions of race. The cladistic conception helps systematists understand the patterns and processes of human evolution. Constructivist conceptions, on the other hand, aid our understanding of social and political implications of current uses of the term. Both perspectives should be recognized as legitimate.

6 Conclusion

I have argued that race theorists have been too quick to reject the biological reality of race. Cladistic classification, in conjunction with current work in human evolution, shows that biological races once existed, but they may no longer exist. It is possible to accept a biological account of race without accepting odious claims about genetically based racial superiority. Furthermore, the historical character of the cladistic concept means that it is not in conflict with most constructivist views.

Acknowledgements

I would like to thank Tracy Aiello, Milo Aukerman, Dan Hausman, Dave Lorvick, Laura Sizer, Doug Smith, Chris Stephens, Vanessa Tanaka, and especially Elliott Sober for their valuable criticisms and suggestions. Thanks also to the anonymous referees of this journal for helpful comments on an earlier draft.

Notes

1. I am using the term ''race theorist'' to mean someone who makes theoretical claims about race.

2. See Mayr [1963] for a brief discussion of the history of the subspecies concept in biology.

3. This is David Hull's [1978] version of Aristotelian essentialism.

4. Zemach ([1976], p. 120), who is sceptical about the existence of essential properties, challenges this example by arguing that ''there is no chemical constitution common to all bodies of water.'' I will not address this criticism since my aim here is merely to provide an intuitive example.

5. Although this characterization of essentialism is fairly vague—for example, the concept of explanation needs more development—it should be good enough for my purposes.

6. For the purposes of these studies, ''racial'' categories were assumed to correspond with major skin-colour groupings.

7. In addition to essentialism, pheneticism, and cladism, there is a fourth school of classification—*evolutionary taxonomy*—which uses both genealogy and adaptive similarity for individuating taxa. I mention this school only to set it aside; it fails to offer a nonarbitrary criterion for when adaptive similarity matters more than propinquity of descent, and vice versa (Sober [1993]).

8. It is important to recognize the difference between a phylogenetic tree and a cladogram. Cladograms represent only one feature of phylogenetic development, namely, branching sequence. Phylogenetic trees, on the other hand, represent a number of features such as branching sequence and the relation between ancestors and descendants.

9. It has been argued that species cannot be required to be monophyletic. If species are monophyletic, then ancestral species cannot exist. On this view, an ancestral species *belongs* to monophyletic groups, but they cannot *be* monophyletic groups. Thus, although the monophyly criterion is useful for classifying superspecific taxa (and, as I will argue, subspecific taxa), it is not an appropriate criterion for designating species (Sober [1993]).

10. I have already discussed the distinction between cladistic and phenetic methods of classification—pheneticism uses overall similarity to define its taxa and cladism uses monophyly. Here is the difference between pheneticism and cladism when it comes to phylogenetic inference. *Pheneticists* use overall similarity to measure degrees of relatedness. When two breeding populations are found to be more similar to each other than either is to the third, a pheneticist would conclude that the first two are more closely related to each other than either is to a third. *Cladists*,

on the other hand, distinguish between two types of similarity—those that are derived through descent from a common ancestor (synapomorphies), and those representing retained ancestral characters (symplesiomorphies)—and argue that only the former count as evidence when assessing degrees of relatedness.

11. Constructivism, in addition to being a thesis about kinds, can be a thesis about objects, properties, or reality. I am focusing on the social construction of kinds because of my interest in race as a social construct.

12. See Scheffler [1980], Wolterstorff [1987], Giere [1988], Stove [1991], and Devitt [1991] for persuasive criticisms of global conceptual constructivism.

References

Aiello, L. [1993]: "The Fossil Evidence for Modern Human Origins in Africa: A Revised View," *American Anthropologist*, 95, pp. 73–96.

Appiah, K. A. [1986]: "The Uncompleted Argument: Du Bois and the Illusion of Race," in H. L. Gates (*ed.*), *"Race," Writing and Difference*, Chicago, University of Chicago Press, pp. 21–37.

Appiah, K. A. [1992]: *In My Father's House: Africa in the Philosophy of Culture*, New York, Oxford University Press.

Appiah, K. A. [1996]: "Race, Culture, Identity: Misunderstood Connections," in K. A. Appiah and A. Gutmann (*eds*), *Color Conscious: The Political Morality of Race*, Princeton, Princeton University Press, pp. 30–105.

Arnold, E. N. [1981]: "Estimating Phylogenies at Low Taxonomic Levels," *Zeitschrift für Zoologische Systematic und Evolutionforschung*, 19, pp. 1–35.

Banton, M. and Harwood, J. [1975]: *The Race Concept*, London, David & Charles.

Barnicot, N. A. [1964]: "Taxonomy and Variation in Modern Man," in A. Montagu (*ed.*), *The Concept of Race*, New York, Collier Books, pp. 180–227.

Bateman, R., Goddard, I., O'Grady, R., Funk, V. A., Mooi, R., Kress, W. J., and Cannell, P. [1990a]: "Speaking of Forked Tongues: The Feasibility of Reconciling Human Phylogeny and the History of Language," *Current Anthropology*, 31, 1, pp. 1–13.

Bateman, R., Goddard, I., O'Grady, R., Funk, V. A., Mooi, R., Kress, W. J., and Cannell, P. [1990b]: "On Human Phylogeny and Linguistic History: Reply to Comments," *Current Anthropology*, 31, 1, pp. 177–83.

Bayard, D. [1990]: "Comments on 'speaking of Forked Tongues,'" *Current Anthropology*, 31, 1, pp. 14–15.

Bremer, K. and Wanntorp, H. [1979a]: "Geographic Populations or Biological Species in Phylogeny Reconstruction?" *Systematic Zoology*, 28, 2, pp. 220–4.

Bremer, K. and Wanntorp, H. [1979b]: "Hierarchy and Reticulation in Systematics," *Systematic Zoology*, 28, 4, pp. 624–7.

Cavalli-Sforza, L. L. [1974]: "Genetics of Human Populations," *Scientific American*, 231, 3, pp. 81–9.

Cavalli-Sforza, L. L. [1991]: "Genes, Peoples and Languages," *Scientific American*, 265, pp. 104–10.

Cavalli-Sforza, L. L., Menozzi, P., and Piazza, A. [1994]: *The History and Geography of Human Genes*, Princeton, Princeton University Press.

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., and Mountain, J. [1988]: "Reconstruction of Human Evolution: Bringing Together Genetic, Archeological, and Linguistic Data," *Proceedings of the National Academy of Science*, 85, pp. 6002–6.

Darwin, C. [1859]: *On the Origin of Species*, Cambridge, MA, Harvard University Press, 1964.

De Sousa, R. [1984]: "The Natural Shiftiness of Natural Kinds," *Canadian Journal of Philosophy*, 14, 4, pp. 561–80.

Devitt, M. [1991]: *Realism and Truth*, 2nd edition, Princeton, Princeton University Press.

Dewey, J. [1938]: *Logic: The Theory of Inquiry*, New York, Holt.

Du Bois, W. E. B. [1940]: *Dusk of Dawn: An Essay Toward an Autobiography of a Race Concept*, New Brunswick, Transaction Publishers.

Ehrlich, P. and Holm, R. [1964]: "A Biological View of Race," in A. Montagu (*ed.*), *The Concept of Race*, New York, Collier Books, pp. 153–79.

Ehrlich, P. and Raven, P. [1969]: "Differentiation of Populations," *Science*, 165, pp. 1228–32.

Felsenstein, J. [1984]: "The Statistical Approach to Inferring Evolutionary Trees and What it Tells Us About Parsimony and Compatibility," in T. Duncan and T. Stuessy (*eds*), *Cladistics: Perspectives on the Reconstruction of Evolutionary History*, New York, Columbia University Press, pp. 169–91.

Foucault, M. [1978]: *The History of Sexuality*, Robert Hurley, trans., New York, Pantheon Books.

Ghiselin, M. [1974]: "A Radical Solution to the Species Problem," *Systematic Zoology*, 23, pp. 536–44.

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

Gillham, N. W. [1956]: "Geographic Variation and the Subspecies Concept in Butterflies," *Systematic Zoology*, 5, pp. 110–20.

Goldberg, D. T. [1993]: *Racist Culture: Philosophy and the Politics of Meaning*, Cambridge, Basil Blackwell.

Goodman, N. [1978]: *Ways of Worldmaking*, Indianapolis, Hackett Publishing Company.

Hacking, I. [1986]: "Making Up People," in T. C. Heller, M. Sosna, and D. E. Wellber (*eds*), *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought*, Stanford, Stanford University Press, pp. 222–36 and 347–8.

Hacking, I. [1988]: "The Sociology of Knowledge about Child Abuse," *Noûs*, XXII, pp. 53–63.

Hacking, I. [1991]: "The Making and Molding of Child Abuse," *Critical Inquiry*, 17, pp. 253–88.

Hacking, I. [1992]: "World-Making by Kind-Making: Child Abuse for Example," in M. Douglas and D. Hull (*eds*), *How Classification Works*, Edinburgh, Edinburgh University Press, pp. 180–238.

Hagmeier, E. M. [1958]: "The Inapplicability of the Subspecies Concept in the North American Marten," *Systematic Zoology*, 7, pp. 1–7.

Hennig, W. [1966]: *Phylogenetic Systematics*, Urbana, University of Illinois Press.

Hiernaux, J. [1964]: "The Concept of Race and the Taxonomy of Mankind," in A. Montagu (*ed.*), *The Concept of Race*, New York, Collier Books, pp. 29–45.

Hull, D. [1978]: "A Matter of Individuality," *Philosophy of Science*, 45, pp. 335–60.

Johnston, R. F. and Selander, R. K. [1964]: "House Sparrows: Rapid Evolution of Races in North America," *Science*, 144, pp. 548–50.

Johnston, R. F. and Selander, R. K. [1966]: "On Supposed 'Subspecies' of North American House Sparrows," *Systematic Zoology*, 3, pp. 357–8.

Kripke, S. [1972]: *Naming and Necessity*, Cambridge, MA, Harvard University Press, 1980.

Latour, B. and Woolgar, S. [1979]: *Laboratory Life: The Social Construction of Scientific Facts*, Beverly Hills, CA, Sage.

Lewis, C. I. [1946]: *An Analysis of Knowledge and Valuation*, La Salle, Open Court.

Lewontin, R. C. [1972]: "The Apportionment of Human Diversity," *Evolutionary Biology*, 6, pp. 381–98.

Lewontin, R. C., Rose, S., and Kamin, L. [1984]: *Not in Our Genes: Biology, Ideology, and Human Nature*, New York, Pantheon.

Livingstone, F. B. [1964]: "On the Nonexistence of Human Races," in A. Montagu (*ed.*), *The Concept of Race*, New York, Collier Books, pp. 46–60.

McIntosh, M. [1968]: "The Homosexual Role," *Social Problems*, 16, 2, pp. 182–93.

Mayr, E. [1959]: "Typological versus Population Thinking," in E. Sober (*ed.*) [1994], *Conceptual Issues in Evolutionary Biology*, 2nd edition, Cambridge, MA, MIT Press, pp. 157–60.

Mayr, E. [1963]: *Animal Species and Evolution*, Cambridge, MA, Harvard University Press.

Mayr, E. [1982]: "Of What Use are Subspecies," *Auk*, 99, pp. 593–5.

Mayr, E. and Ashlock, P. [1991]: *Principles of Systematic Zoology*, New York, McGraw Hill.

Montagu, A. [1941]: "The Concept of Race in the Human Species in the Light of Genetics," in A. Montagu (*ed.*) [1964], *The Concept of Race*, New York, Collier Books, pp. 1–11.

Montagu, A. [1959]: "The Concept of Race," in A. Montagu (*ed.*) [1964], *The Concept of Race*, New York, Collier Books, pp. 12–28.

Montagu, A. [1964]: "Introduction," in A. Montagu (*ed.*) [1964], *The Concept of Race*, New York, Collier Books, pp. xi–xviii.

Nei, M. and Roychoudhury, A. K. [1972]: "Gene Differences Between Caucasian, Negro, and Japanese Populations," *Science*, 177, pp. 434–6.

Nei, M. and Roychoudhury, A. K. [1974]: "Genetic Variation Within and Between the Three Major Races of Man, Caucasoids, Negroids, and Mongoloids," *American Journal of Human Genetics*, 26, pp. 421–43.

Nei, M. and Roychoudhury, A. K. [1982]: "Genetic Relationships and Evolution of Human Races," *Evolutionary Biology*, 14, pp. 1–59.

Nei, M. and Roychoudhury, A. K. [1993]: "Evolutionary Relationships of Human Populations on a Global Scale," *Molecular Biology and Evolution*, 10, 5, pp. 927–43.

O'Grady, R. T., Goddard, I., Bateman, R. M., DiMichelle, W. A., Funk, V. A., Kress, W. J., Mooi, R., and Cannell, P. [1989]: "Genes and Tongues," *Science*, 243, p. 1651.

Omi, M. and Winant, H. [1994]: *Racial Formation in the United States: From the 1960s to the 1990s*, 2nd edition, New York, Routledge.

Outlaw, L. [1995]: "On W. E. B. Du Bois's 'The Conservation of Races,'" in L. A. Bell and D. Blumenfeld (*eds*), *Overcoming Racism and Sexism*, Lanham, MD, Rowman and Littlefield Publishers, pp. 79–102.

Putnam, H. [1975]: "The Meaning of 'Meaning,'" in H. Putnam (*ed.*), *Mind, Language, and Reality*, Cambridge, Cambridge University Press, pp. 215–71.

Ridley, M. [1986]: *Evolution and Classification: The Reformation of Cladism*, London, Longman.

Scheffler, I. [1980]: "The Wonderful Worlds of Goodman," *Synthese*, 45, pp. 201–9.

Shreeve, J. [1995]: "The Neanderthal Peace," *Discover*, Sept., pp. 70–81.

Sidelle, A. [1989]: *Necessity, Essence, and Individuation: A Defense of Conventionalism*, Ithaca, Cornell University Press.

Sober, E. [1980]: "Evolution, Population Thinking, and Essentialism," *Philosophy of Science*, 47, pp. 350–83, reprinted in E. Sober (*ed.*) [1994], *Conceptual Issues in Evolutionary Biology*, 2nd edition, Cambridge, MA, MIT Press, pp. 161–89 (page references to the latter).

Sober, E. [1988]: *Reconstructing the Past: Parsimony, Evolution, and Inference*, Cambridge, MA, MIT Press.

Sober, E. [1993]: *Philosophy of Biology*, Boulder, Westview Press.

Storer, R. W. [1982]: "Subspecies and the Study of Geographic Variation," *Auk*, 99, pp. 599–601.

Stove, D. [1991]: *The Plato Cult and Other Philosophical Follies*, Cambridge, MA, Basil Blackwell.

Stringer, C. B. [1990]: "The Emergence of Modern Humans," *Scientific American*, 263, pp. 98–104.

Stringer, C. B. and Andrews, P. [1988]: "Genetic and Fossil Evidence for the Origin of Modern Humans," *Science*, 239, pp. 1263–8.

Templeton, A. R. [1993]: "The 'Eve' Hypothesis: A Genetic Critique and Reanalysis," *American Anthropologist*, 95, 1, pp. 51–72.

Thorne, A. G. and Wolpoff, M. H. [1992]: "The Multiregional Evolution of Humans," *Scientific American*, 266, pp. 76–83.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A. C. [1991]: "African Populations and the Evolution of Human Mitochondrial DNA," *Science*, 253, pp. 1503–7.

Wilson, A. C. and Cann, R. L. [1992]: "The Recent African Genesis of Humans," *Scientific American*, 266, pp. 68–73.

Wilson, E. O. and Brown, W. L., Jr. [1953]: "The Subspecies Concept and Its Taxonomic Application," *Systematic Zoology*, 2, pp. 97–111.

Wolpoff, M. H. [1989a]: "Multiregional Evolution: The Fossil Alternative to Eden," in P. Mellars and C. Stringer (*eds*), *The Human Revolution*, Edinburgh, Edinburgh University Press, pp. 62–108.

Wolpoff, M. H. [1989b]: "The Place of the Neandertals in Human Evolution," in E. Trinkaus (*ed*.), *The Emergence of Modern Humans*, Cambridge, Cambridge University Press, pp. 97–141.

Wolpoff, M. H., Spuhler, J. N., Smith, F. H., Radovcic, J., Pope, G., Frayer, D., Eckhardt, R., and Clark, G. [1988]: "Modern Human Origins," *Science*, 241, pp. 772–3.

Wolpoff, M. H., Wu, X., and Thorne, A. G. [1984]: "Modern *Homo Sapiens* Origins: A General Theory of Hominid Evolution Involving the Fossil Evidence from East Asia," in F. H. Smith and F. Spencer (*eds*), *The Origins of Modern Humans*, New York, Alan R. Liss., pp. 411–83.

Wolterstorff, N. [1987]: "Are Concept-Users World-Makers?" in J. E. Tomberlin (*ed*.), *Philosophical Perspectives, 1: Metaphysics*, Atascadero, CA, Ridgeview Publishing Company, pp. 233–67.

Woolgar, S. [1988]: *Science: The Very Idea*, Chichester, Sussex, Ellis Horwood.

Zemach, E. M. [1976]: "Putnam's Theory on the Reference of Substance Terms," *Journal of Philosophy*, 73, 5, pp. 116–27.

Zusi, R. L. [1982]: "Infraspecific Geographic Variation and the Subspecies Concept," *Auk*, 99, pp. 606–8.

XII Cultural Evolution

24 Does Culture Evolve?

Joseph Fracchia and Richard C. Lewontin

The drive to describe cultural history as an evolutionary process has two sources. One from within social theory is part of the impetus to convert social studies into "social sciences" providing them with the status accorded to the natural sciences. The other comes from within biology and biological anthropology in the belief that the theory of evolution must be universal in its application to all functions of all living organisms. The social-scientific theory of cultural evolution is pre-Darwinian, employing a developmental model of unfolding characterized by intrinsic directionality, by definable stages that succeed each other, and by some criterion of progress. It is arbitrary in its definitions of progress, and has had the political problem that a diachronic claim of cultural progress implies a synchronic differential valuation of present-day cultures. The biological scheme creates an isomorphism between the Darwinian mechanism of evolution and cultural history, postulating rules of cultural "mutation," cultural inheritance and some mechanism of natural selection among cultural alternatives. It uses simplistic ad hoc notions of individual acculturation and of the differential survival and reproduction of cultural elements. It is unclear what useful work is done by substituting the metaphor of evolution for history.

I Culture, the Two Cultures, and History

In his well-known "Two Cultures" essay C. P. Snow reported a gap between the literary and natural-scientific cultures. Acknowledging that "a good deal of the scientific feeling" is shared by some of his "American sociological friends," Snow was well aware that there was a degree of artificiality in limiting the number of cultures to the "very dangerous" one of two. Yet, he based his binarist decision largely on the cohesion of the natural-scientific and literary communities that made of them cultures "not only in an intellectual but also in an anthropological sense."[1] The intellectual division of labor and the development of disciplinary languages certainly seem to substantiate his reference to two incommensurate cultures. Anyone who has sat on a university committee reviewing grant proposals from, and consisting of citizens of, each of the

From *History and Theory*, 1999, 8: 52–78. Reprinted by permission of Blackwell Publishing.

cultures must have observed the pattern of who accuses whom of using jargon and be convinced that at least the academic version of Snow's gap, that between the humanities and the natural sciences, has widened into a seemingly unbridgeable abyss. It has become commonplace that the two cultures have nothing in common.

Perhaps, however, too much has been made of this abyss. Members of the literary culture, and of the humanities in general, may be appalled at the thought of scientists mucking around on cultural terrain and subjecting it to "scientific analysis." But natural scientists seem more irritated than intimidated by the apparent independence of human culture from scientific study. And social scientists expressing their discontent about being dangled over the abyss helped prompt Snow to take "A Second Look" and to acknowledge the "coming" of a "third" social-scientific culture with the potential to "soften" the communication difficulties between the other two.[2] Cultural anthropologists, moreover, at least those with a "scientific" rather than a "relativist" bent, could point to a long tradition in their discipline of attempting to bridge the abyss by subjecting culture and its "evolution" to scientific study.

The idea that culture evolves antedated the Darwinian theory of organic evolution and, indeed, Herbert Spencer argued in support of Darwin that, after all, everything else evolves.[3] Of course, the validation of the theory of organic evolution has in no way depended on such argument by generalization. It is Darwinism that became the theory of evolution, and, standing Spencer on his head, one inspiration for theories of cultural evolution since 1859. There has been a long and bloody Hundred Years War among cultural anthropologists over whether human culture can be said to evolve, a war in which the contending parties alternate in their periods of hegemony over the contested territory. That struggle has, in part, been a philosophical consequence of a diversity in the understanding of what distinguishes an evolutionary from a "merely" historical process. In greater part, however, it can only be understood as a confrontation between the drive to scientize the study of culture and the political consequences that seemed to flow from an evolutionary understanding of cultural history.

Until the last decade of the nineteenth century, partly under the influence of Darwinism, but also as an extension of pre-Darwinian progressivist views that characterized a triumphant industrial capitalism, anthropological theory was built on an ideology of evolutionary progress. Lewis Henry Morgan's construal of the history of culture as the progress from savagery through barbarism to civilization was the model. In the 1890s Boas successfully challenged the racism and imperialism that seemed the inevitable consequences of Morgan's progressivist views and set an anti-evolutionist tone that characterized cultural anthropology until after the Second World War. Beginning with the celebration in 1959 of the hundredth anniversary of the publication of the *Origin of Species*, there was a demand from within anthropology to reintroduce an evolutionary perspective into cultural history from which it had been purged by the Boasites, a demand that was later given collateral support by the development within biology of sociobiological theories of human nature. But again the implication that there were ''higher'' and ''lower'' stages of human culture, an implication that seemed built into any evolutionary theory, could not survive its political consequences, and so by 1980 cultural anthropology once again returned to its Boasian model of cultural change, cultural differentiation, and cultural history, but without cultural evolution.

In his Preface to the manifesto of cultural evolution *redivivus*, *Evolution and Culture*, Leslie White bitterly attacked the Boas tradition, conflating it with general creationist anti-evolutionism:

> The repudiation of evolutionism in the United States is not easily explained. Many nonanthropological scientists find it incredible that a man who has been hailed as ''the world's greatest anthropologist''. . . , namely Franz Boas, a man who was a member of the National Academy of Sciences and President of the American Association for the Advancement of Science, should have devoted himself assiduously and with vigor for decades to this antiscientific and reactionary pursuit.[4]

But why does White insist, illogically and counterfactually, that a denial of cultural evolution is anti-evolutionism *tout court*? There is a hint in the word ''antiscientific,'' but all is explicitly revealed two pages later: ''The return to evolutionism was, of course, inevitable *if . . . science was to embrace cultural anthropology*. The concept of evolution has proved itself to be too fundamental and fruitful to be ignored indefinitely by *anything calling itself a science*'' (emphasis added).[5] Thus, the demand for a theory of cultural evolution is really a demand that cultural anthropology be included in the grand twentieth-century movement to scientize all aspects of the study of society, to become validated as a part of ''social science.'' The issue was particularly pressing for cultural anthropologists because they were engaged in an institutional struggle for support of their research and academic prestige with members of their own academic departments who practiced the undoubtedly scientific activity of physical anthropology.

But the demand for a theory of cultural evolution also arose from among the natural sciences, particularly among evolutionary biologists for whom the ability to explain all properties of all living organisms, using a common evolutionary mechanism, is the ultimate test of the validity of their science. Ever scornful of what they acronymiously dubbed the SSSM (the ''standard social science model'' based on Durkheim's axiom), evolutionary biologists doubted not that the scientific analysis and understanding of the place and evolution of culture in the life history of *Homo sapiens* was properly the province of students of human evolution. The advent of culture was, after all, a biological adaptation and it must therefore be explicable by biological science. Yet a combination of two inhibiting factors kept the forays of evolutionary biologists into the cultural realm to a minimum at least from the end of World War II into the mid-1970s. These were: the close link between biologically based pseudoscientific social and cultural theories and genocide; and the lack of a properly comprehensive theory.

This latter problem, as most recent cultural evolutionists agree, was finally solved with the concluding chapter of E. O. Wilson's *Sociobiology* (1975) which provided the impetus for the latest round of attempts to subject human history to evolutionary explanation. There, Wilson sketched the certainty that, as he put it a few years later in *On Human Nature*, the appropriate instrument for closing the "famous gap between the two cultures" is "general sociobiology, which is simply the extension of population biology and evolutionary theory to social organization."[6]

While rather adamant about their scientific right to explain not just the evolution of human cultural capacities, but also cultural evolution, biologists are also rather uneasy about their self-imposed obligation to do so. For they wager the *raison d'être* of science on establishing the validity of the principle of reductionism: in order for science to remain tenable, it must have universal explanatory power; and this means "nesting" the human sciences in the great hierarchy of sciences. If evolutionary biology cannot explain human culture, then perhaps its explanations of other phenomena ought to be reexamined. Intrigued by the challenge, Wilson noted that reduction is "feared and resented"[7] by too many in the human sciences and, in a bold Napoleonic metaphor, he sniffed "a not unpleasant whiff of grapeshot" in the thought that the applicability of sociobiology to human beings is a battle on which hangs the fate of "conventional evolutionary theory."[8] Thrilled by the challenge and inspired by the apparent potential of the sociobiological synthesis, an increasing number of scientists attempted to build on Wilson's blueprint in order to bridge the abyss and lay claim to the territory on the other side.

Some members of the social sciences, those who preferred to be recognized as *bona fide* scientists and not just as members of a "third" culture, were meanwhile growing uneasy over the proliferation of opposing theories and models that had apparently brought the production of social-scientific knowledge to a standstill. Such social scientists began to question their own SSSM and turned increasingly to the new and seemingly infallible sociobiological synthesis for the models and explanatory mechanisms that would put their own disciplines on proper scientific footing. Alexander Rosenberg, for example, bemoans the inability of the social sciences to live up to John Stuart Mill's hope for them, namely, to be based on explanatory laws. In a telling formulation he claims that

> the social sciences would be of only passing interest, only entertaining diversions, like an interesting novel or an exciting film, unless they too stood the chance of leading to the kind of technological achievements characteristic of natural science. For a social science conceived as anything less practical in ultimate application would simply not count as knowledge, on my view. And if it does not count as *knowledge*, disputes about its methods and concepts are no more important than learned literary criticism or film reviews are to our uninformed enjoyment of the books and movies we like.[9]

Rosenberg expects this to be rectified as soon as the social sciences are treated as life sciences; and he optimistically predicts that the study of human behavior, once set on a biological footing, "will admit of as much formally quantified and mathematical description as the most mathematical economist could hope for." Against all claims for their uniqueness he insists that the traditional social sciences have been "superseded" by, and will only become truly scientific when subsumed under, sociobiology.[10]

More recently, anthropologist John Tooby and psychologist Leda Cosmides have also chastised the social sciences for their "self-conscious stance of intellectual autarky"; their "disconnection from the rest of science has left a hole in the fabric of our organized knowledge where the human sciences should be." The lack of progress in the social sciences has been caused by their "failure to explore or accept their logical connections to the rest of the body of science—that is, to causally locate their objects of study inside the larger network of scientific knowledge."[11]

This desideratum is the cornerstone of the journal *Politics and the Life Sciences* whose editors and contributors insist that the social sciences must be nested within the life sciences. The hopes for a synthesis implicit in the journal's name were expressed by Richard Shelly Hartigan in a flattering review of Richard D. Alexander's *The Biology of Moral Systems* (1987). Predicting marital bliss, Hartigan confidently asserts that "the lengthy divorce of the natural from the human sciences is about to end with reunion. Though the nuptials may be delayed awhile, the parties are at least getting to know each other again more intimately."[12] The reunion consists of articles devoted to the "Darwinian" explanation of such topics as social alienation, the nuclear arms race, the legal process, social stratification, oral argument in the supreme court, the relation between human intelligence and national power, and even feminism.[13]

These examples could be multiplied, but as this brief overview indicates, the biggest engineering project attempting to bridge the gap at least between the cultures of the natural and the human sciences over the last few decades has been initiated by natural scientists, anxious perhaps about having wagered their *raison d'être* on the success of their imperialist venture; and it has quickly drawn the participation of those social scientists optimistic about overcoming their inferiority complex and gaining respectability by grounding their own disciplines in the natural sciences. The bridge itself is the concept of "cultural evolution" whose scientific girders are the categories and explanatory laws either directly borrowed or derived from a narrowly selectionist approach to the study of biological evolution.

At the outset we must make clear what the issue of cultural evolution is *not* about. First, there is no question that culture as a phenomenon has evolved from the absence of culture as a consequence of biological change. Whether or not other primates have culture on some definition, the insectivores, from which the primates evolved, do not, so at some stage in biological evolution culture appeared as a novelty. Second, no one challenges the evident fact that human cultures have changed since the first appearance of *Homo sapiens*, but not even the most biologistic theory proposes that major changes within the phenomenon of culture—say the invention of an alphabet or of settled agriculture—was a consequence of genetic evolution of the human central nervous system. Human culture has had a history, but to say that culture is a consequence of a historical process is not the same as saying that it evolves. What constitutes an evolutionary process as opposed to a ''merely'' historical one? What explanatory work is done by claiming that culture has evolved?

Leslie White's *cri de coeur* accusing the Boasians of aligning themselves with anti-evolutionist creationism confounds two quite different issues. The mid-nineteenth-century struggle against evolution, mirrored in modern Christian creationism, was not over whether the succession of life forms from earlier times to the present has some law-like properties that give some shape to that history. Rather it was, and remains, a denial that organismic forms have had a history at all, that there has been significant change in species and that present-day life forms arose from others quite unlike them. But on one denies that culture has had a history, that industrial production arose from societies that were at a previous time pastoralist and agricultural. Not even the most literal of fundamentalists thinks that God created the motor car on the sixth day. Ironically it is a form of traditional Christianity that simultaneously denies an intelligible history to organic life as a whole, while asserting a directionality to human history, the ascent toward final redemption from the depths of the Fall.

White's identification of the struggle over cultural evolution with the struggle over organic evolution, if it is more than a deliberate piece of propaganda in a battle for academic legitimacy, is really a struggle over the nature of historical processes. At base, it is meant to be a rejection of the proposition that human cultural history is just one damn thing after another, claiming that, on the contrary, there is an underlying nomothetic process. But in asserting the claim that culture evolves White claimed more than what was necessary. History may indeed be law-like in some sense, but does that make a historical process evolutionary? There may be law-like constraints on historical change like Ibn Khaldun's rule that ''Bedouins can gain control only over flat territory,'' but we do not therefore characterize the *Muqaddimah* as providing an ''evolutionary'' theory of history, any more than Hegel's third kind of history, the Philosophical, is claimed to be a theory of evolution.[14]

It might be asserted that for theories to qualify as evolutionary they must consist of more than mere constraints and prohibitions; rather they must be characterized by generative laws or mechanisms whose operations produce the actual histories. But the *Muqaddimah* offers laws of the origin, transformation, differentiation, and eventual extinction of political formations: ''Dynasties of wide power and large royal authority have their origin in religion based either on prophethood or truthful propaganda''; ''The authority of the dynasty at first expands to its limit and then is narrowed down in successive stages, until the dynasty dissolves and disappears"; "With regard to the amount of prosperity and business activity in them, cities and towns differ in accordance with the different size of their population."[15] These are not simply empirical generalizations. Each is derived as the necessary consequence of basic properties of human motivation, just as the war of all against all is derived by Hobbes from the basic assumptions that human beings are, by nature, self-expanding in their demands and that the resources for their expansion are limited. The ease with which the concept of the "evolution of culture" has been employed in anthropology and human evolutionary biology finds no parallel in the discourse of contemporary historians. When François Furet and Mona Ozouf write, in their Preface to *A Critical Dictionary of the French Revolution*, that "ignoring the evolution of historiography means overlooking an important aspect of the event itself," they mean only that historiography has changed, that is, that it has had a history.[16]

It might be that "evolution" and "history" are meant to be separated by questions of scale and grain. Modes of production, familial and other group relationships, forms of political organization, levels of technology are seen as general properties of human social existence. They are also "culture" and they are said to "evolve" while spatio-temporally individualized sequences like the events in France from the Estates General to Thermidor are only instantiations of classes of cultural phenomena, schemata that are repeated in different places and at different times. So Leslie White makes the distinction between the particularity of micro (historical) events and the generality of macro (evolutionary events): "I should like to call the temporal particularizing process, in which events are considered significant in terms of their uniqueness and particularity, 'history' and call the temporal generalizing process which deals with the phenomena as classes rather than particular events, 'evolution.'"[17] But if this is what is meant to discriminate evolution from mere history, then the cultural evolutionist departs radically from theories of evolution of the physical world. For Darwinism, not only organic life as a whole, but each species and each population in each species evolves. The standard model of organic evolution begins with the evolutionary forces that cause local populations to change over relatively short times, and derives the evolution of individual species in time from changes in populations that comprise them. Moreover, in its usual reductionist form, evolutionary theory explains the evolution of life as a whole as a mechanical consequence of the rise and fall of individual species. So why, if human culture evolves, has not Bedouin culture evolved, or the Middle East, or the state called Saudi Arabia?

The attempt to differentiate "cultural evolution" from "history" brings us to the edge of a different kind of abyss—one that is broader and older, though obscured by, the more visible one between the human and natural sciences. This abyss cuts across established disciplinary boundaries, and separates nomological and historical modes of explanation. Civil wars always inflict the deepest wounds. And the battles *within* the human sciences (between historians emphasizing contingency and particularity and social scientists insisting on general laws and models) and *within* the natural sciences (between biologists who insist on the contingency, the historicity, of evolution and those who view evolution as a lawful process of selection and adaptation) are by virtue of the proximity of the antagonists frequent, intense, and have perhaps the longest lasting effects.

Snow's depiction of the abyss along disciplinary lines makes those battles appear as perhaps bitter, but nevertheless only intradisciplinary squabbles, as merely different perspectives on common problems. Yet, the cross-disciplinary affinities of "historians" versus "scientists" are nowhere more evident than in the issue that both claim as their own: that which appears to one group as "cultural evolution," to the other as "human histories." The ease, for example, with which confirmed selectionists among evolutionary biologists and those social scientists similarly concerned with explanatory laws have found common cause in the concept of cultural evolution indicates that on fundamental ontological and epistemological issues there is no abyss between them. That ease finds its counterpart in the ease with which the two authors of this essay, a historian and a geneticist, agree on a historical approach to cultural change. The differences between these two perspectives are incommensurable, not because of disciplinary boundaries, but because they involve different conceptions about the nature of "scientific" inquiry, different ontological and epistemological assumptions, and accordingly different modes of explanation.

Darwinian theorists of cultural evolution universally agree that selection is *the* explanatory law, the key to explaining all "evolutionary" or "historical" developments at any sociocultural and historical coordinates. In this way human history is reduced to a unitary process, its complex dynamics to a rather singular logic, and the particularity of historical time is reduced to "empty abstract time" (Walter Benjamin).[18]

We begin with different assumptions about historical objects and, accordingly, about historical time. We view historical phenomena as particulars embedded in particular sociocultural forms, each with its own systemic properties and discrete logic of production and reproduction, its own dynamics of stasis and change. Each sociocultural form therefore has, to borrow an appropriate phrase from Louis Althusser, its own time and history. Because every historical phenomenon has its own particular locus in a particular sociocultural constellation with its own concrete and particular time and history, there is no one transhistorical law or generality that can explain the dynamics of all historical change. Our contention, therefore, is that cultural evolutionary theories have not been (nor will be) able to meet even their own claims to explain the past and predict the future. And this is because of the problematic assumptions about the nature of culture and the problematic conflation of historical and evolutionary processes.

II The Forms of Evolutionary Theory

Models of the evolution of phenomena are traditionally models of the temporal change in the nature of ensembles of elements. The individual elements in the ensemble can be physical objects like organisms or stars or properties like size or chemical composition or syntactic structure. So when we speak of the "evolution of human beings" we mean a change in the composition of the ensemble of physical individuals that we identify individually as human, but we can as well consider the "evolution of European painting" as a change in the ensemble of materials, techniques, subjects, and design principles that characterize the production of that art. Whether it is physical objects or attributes or artifacts, it is not any individual element, but the composition of the ensemble that is at the center of interest.

Evolutionary theories as they have been constructed for the physical world and as they have been taken over into human social phenomena can be classified according to two properties. First, they may be either transformational or variational. In a transformational theory, the ensemble of elements changes in time because each of the elements in the ensemble undergoes roughly the same secular change during its individual history. That is, the evolution of the ensemble is a result of the developmental pattern of each individual. The transformational model characterized all evolutionary theories until Darwin, and has remained the model for the evolution of the physical universe since Kant and Laplace produced the Nebular Hypothesis for the origin of the Solar System. The collection of stars in the cosmos has been evolving because every star is individually undergoing an aging process from its birth at the Big Bang, through a sequence of nuclear reactions until it exhausts its nuclear fuel and then collapses into a dead mass. It is this model that is embodied in the very word "evolution," an unfolding or unrolling of a history that is already immanent in the object. It is a model of evolution that takes as its cause the development (*desarollo*, *Entwicklung*), the unrolling or unfolding of the predetermined fate of each element in the ensemble.

The alternative, invented by Darwin to explain organic evolution, is a variational evolutionary scheme. In variational evolution, the history of the ensemble is not a consequence of the uniform unfolding of individual life histories. Rather, variational evolution through time is a consequence of variation among members of the ensemble at any instant of time. Different individuals have different properties and the ensemble is characterized by the collection of these properties and their statistical distribution. The evolution of the ensemble occurs because the different individual elements are eliminated from the ensemble or increase their numbers in the population at different rates. Thus, the statistical distribution of properties changes as some types become more common and others die out. Individual elements may indeed change during their lifetime, but if they do, these changes are in directions unrelated to the dynamic of the collection as a whole and on a time scale much shorter than the evolutionary history of the group. So, the developmental changes that characterize the aging of every living organism are not mirrored in the evolution of the species. Every human being may become grayer and more wrinkled with age, but the species as whole has not become so in 5 million years of evolution from its common ancestor with other primates. Organic evolution is then a consequence of a twofold process: the production of some variation in properties among individual elements followed by the differential survival and propagation of elements of different types. Moreover, the production of the variation is causally independent of its eventual fate in the population. That is what is meant by the claim that organic evolution is based on "random" variation. It is not that the changes in individual properties are uncaused, or the consequence of some force outside of normal physical events. Rather it is that the forces of change internal to organisms, leading to the production of variant individuals, are causally random with respect to the external forces that influence the maintenance and spread to those variants in the population. Many are called, but few are chosen.

The invention of the variational scheme for organic evolution, with its rigorous separation of internal developmental forces from external culling forces, is the major epistemological break achieved by Darwin. All other evolutionary schemes that had been postulated until the appearance of the *Origin* in 1859, whether of the evolution of the cosmos, of organisms, of language, or of ideas, were transformational. The Darwinian variational scheme, with its denial of the causal role of individual developmental histories was, in fact, a negation of evolution as it had previously been understood. The retention of the term "evolution" by Darwinists, while stripping it utterly of its former structural implication, has led to a considerable confusion and ambiguity in subsequent arguments about cultural evolution, for there has been no agreement among cultural evolutionists about just what sort of evolution they mean.

The choice of a transformational, developmental theory of evolution implies properties of the process that are not integral to, although they may be present in, a variational theory: directionality and staging. In an unfolding process the possibility of each successive transformation is dependent on the completion of a previous step of transformation to provide the initial state for the next change. It is not necessary that the complete unfolding the predictable from the very origin of the system because successive steps may be contingent. There may be more than one local unfolding possible from a given state, and these alternatives may be chosen, contingent on various external circumstances. Transformational theories, nevertheless, usually assume a very restricted contingency, putting very strong constraints on which states may succeed each other, and in what order. Indeed the standard theory of embryonic development which provides a metaphorical basis for developmental theories of evolution assumes that there is one and only one possible succession of states. Thus, there is one direction, or at most a few alternative possible directions of change immanent in the nature of the objects. Directionality does not in itself imply that change is monotone or that there is a repeated cycling among states along some simple axis, yet again and again transformational theories take the form of a "Law of Increase of . . . ," complexity, efficiency, control over resources or energy, of Progress itself. The task of filling in the blanks we leave to later pages. A variational theory, in contrast, does not have directionality built into it because the variation on which the sorting process operates is not intrinsically directional, and changes in the statistical distribution of types in the ensemble are assumed to be the consequence of external circumstances that are causally independent of the variation. Nevertheless, one-way directionality has penetrated Darwinism by means of a claim about natural selection. If the differential numerical representation of different types in a species occurs not by chance events of life and death, but because the properties of some organisms confer on them greater ability to survive and reproduce in the environment in which they find themselves, might there not be some properties that would confer a general advantage over most or all environments? Such properties, then, ought to increase across the broad sweep of organisms and over the long duration of evolutionary history, putting aside any particularities of history. So, for example, it has been claimed that complexity has increased during organic evolution, since complex organisms are supposed somehow to be able to survive better the vagaries of an uncertain world. Unfortunately no agreement can be reached on how to measure complexity independent of the explanatory work it is supposed to do. It is, in fact, characteristic of directionality theories that organisms are first arrayed along an axis from lower to higher and then a search is instituted for some property that can be argued to show a similar ordering.

From directionality it is only a short step to a theory of stages. Transformational developmental theories are usually described as a movement from one stage to the next in the sequence, from savagery to barbarism to civilization, from artisanal production to competitive industrial capitalism to monopoly capital. Development begins by some triggering, starting the process from its germ, but there are thought to be a succession of ordered stages through which each entity must pass, the successful passage through one stage being the condition for moving on to the next. Variation among individual entities then arises because there is some variation in the speed of these transitions, but primarily because of arrested development, the failure to pass on to the next stage. Freudian and Piagetian theories are of this nature. It should be no surprise to anthropologists that transformational evolutionary theories of culture identify present-day hunters and gatherers as being in an arrested stage of cultural evolution.

The second property that distinguishes among evolutionary schemes is the mortality of the individual objects in the ensemble. Members of the ensemble may be either immortal, or at least have potential lifetimes that are of the same order as the ensemble as a whole, or they may be mortal or at least have lifetimes significantly shorter than the duration of the entire collection whose evolution is to be explained. The lifetime of the material universe is the same as the lifetime of the longest lived of individual stars. Individual organisms, on the other hand, invariably have their entrances and their exits, but the species may persist. The classification of an evolutionary system as either mortal or immortal is independent of whether it is transformational or variational and the construction of an evolutionary theory for a domain of phenomena—culture, for example—will require model assumptions about both of these properties. Two of the schemata are illustrated by phenomena to which the concept of evolution is commonly applied. Stellar evolution is a transformational evolution of a system composed of immortal objects; organic evolution is variational and its objects, individual organisms, are mortal. Although we do not ordinarily think of it in such terms, an example of an evolutionary process that is variational, but whose objects are immortal, is any separation of a mixture of physical materials by sieving, as for example in panning for gold. The lighter particles are washed away, leaving the flakes of gold behind so that the concentration of gold becomes greater and greater as the process continues, yet the same bits of gold are present at the end of the process as at the beginning. Pre-Darwinian theories of organic evolution were transformational, the entire species evolving as a consequence of slow directional changes in individuals who were, nevertheless, mortal.

The mortality of the individual objects in an evolutionary process raises a fundamental problem, namely, how the changes in the composition of the ensemble that occur within the lifetime of short-lived elements are to be accumulated over the long-term evolution of the group. Whether the evolution is variational or transformational there must be some mechanism by which a new generation of successors retains some vestige of the changes that occurred in a previous time. In the classical vulgar example of Lamarckian transformational evolution, if the ancestors of giraffes slightly elongated their necks to reach up into trees, all the effort would have been wasted, for after their deaths their offspring would need to repeat the process *ab initio*. Nor does the variational scheme of Darwin solve the problem. Were slightly longer-necked variant giraffes to survive better or to leave more offspring than their short-necked companions, and so enrich the proportion of the longer variant in the species, no cumulative change would occur over generations unless the bias introduced by the sieving process in one generation were somehow felt in the composition of the next. That is, it demands some mechanism of inheritance of properties, in the broadest sense. Beyond the observation that offspring had some general resemblance to their parents, neither Darwin nor Lamarck had the benefit of a coherent theory of inheritance, so they had to content themselves with a variety of ad hoc notions about the passage of characteristics, all of which had in common that the properties of individual organisms were somehow directly influenced by the properties of their biological parents at the time of conception. Theorists of cultural evolution, conscious of the need for a theory of inheritance, yet deprived of any compelling evidence for particular law-like mechanisms for the transgenerational passage of cultural change, are in a much more difficult position, although they do not seem to have realized it, because they do not even know whether an actor-to-actor, not to speak of a parent-to-offspring, model of the passage of culture has any general applicability.

III.A The Paradigms of Cultural Evolutionary Theory: Transformational Theories of Cultural Evolution

A remarkable feature of the history of attempts to create a theory of cultural evolution is the disjuncture between the powerful impetus given to those attempts by the triumph of Darwinism, and the form that those essays have taken until recently. Darwin's substitution of the variational scheme of evolution for a transformational one eliminated the need for the postulation of intrinsic directional forces driving the process of change and consequently avoided the need for a theory of progress. If directionality and its special variant, progress, are claimed to be features of a variational evolutionary scheme, they must be imported by means of a force not inherent in the variational process itself. If there is directionality, it must come from outside of organisms, as a claim, for example, about the nature of environments and their histories. Differential reproduction and survival of randomly generated variants contains no intrinsic direction. Developmentalist, transformational theories of evolution, in contrast, are directional by necessity because the motive mechanism is some form of unfolding of an already immanent program.

Beginning with Edward Burnett Tylor's *Primitive Culture* (1871) and Lewis Henry Morgan's *Ancient Society* (1877), cultural evolutionary theory, called forth by the historical phenomenon of Darwinism, ignored the structure of Darwinian explanation, and remained transformational for nearly 100 years. Nearly all of the theories of cultural evolution have had more in common with Herbert Spencer's *Progress: Its Law and Cause* (1857) than with Darwin's *Origin*. First, they have been dominated by notions of progress and direction. This accent on direction and progress has even been used to characterize organic evolution itself. In the most important manifesto of cultural evolutionism since its revival after the Second World War, *Evolution and Culture*, Marshall Sahlins provides a diagram of the evolution, reproduced here, not of culture, but of all animal life. Superimposed on the upward trend along the axis of "Levels of General Progress," identified by Sahlins as "general evolution," are minor diversifications within a level of progress, symptomatic of "specific evolution" (mere history, perhaps).[19] While diagrams like this were icons of nineteenth-century evolutionism, notions of general progress in biology have been expunged from current descriptions of organic evolution. In the modern practice of reconstructing phylogenetic relationships, the antonym of "primitive" is not "advanced," but "derived."

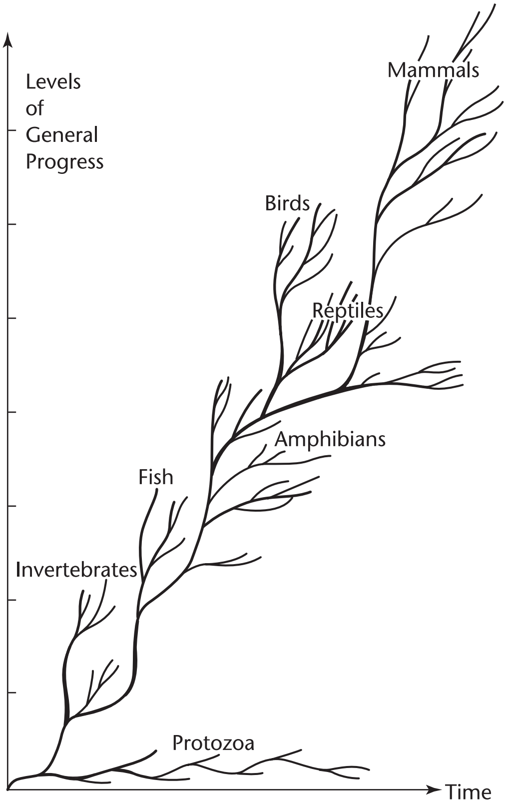


Figure 24.1
Diversity and progress among major lineages of animal life (schematized).

Second, given a commitment to directionality and progress, it then becomes necessary to decide what criteria should be used to determine progress aside from later as against earlier. In theories of organic evolution, recurrent attempts to use the notion of progress have foundered on this issue. It is clear from the fossil record that there has been no increase in the duration of species since the earliest record of multicellular organisms. Nor would anyone be so foolish as to predict that vertebrates will outlast the bacteria, should a major catastrophe overtake all of life on earth. Increasing complexity has been a favorite of progressivist theorists, both for organic evolution and for cultural and political structures, but there is no agreement among physical scientists no how complexity is to be measured and there is the recurrent danger that it will be conveniently defined, post hoc, to put *Homo sapiens* at the top. Sahlins dismisses that shibboleth of bourgeois economic theory, efficiency, as a measure on the grounds that ''an organism can be more efficient than another and yet remain less highly developed."[20] By more highly developed he means having more parts and subparts, more specialization of parts and more effective integration and, subserving these, the transformation of more total energy. Exactly how that cashes out in the great progress from fishes to reptiles in the diagram is not made clear. It is clear, however, what work is done in the domain of culture. Industrial capitalism certainly turns over more calories per capita than does the economy of the Yanamamo of the Orinocan rain forest, and almost any description of a European polity of 1999 will show it to have more parts and subparts with greater specialization than a fief in thirteenth-century Europe, although the question of the relative integration of feudal and bourgeois society as a whole can be debated. Nor can this characterization of an increasing level of cultural progress be attacked on the grounds that some earlier cultures, say Athenian democracy, as most would agree, were more progressive than Carolingian feudalism. The combination of general and specific evolution allows for local exceptions, especially if cultures in different parts of the world are undergoing independent evolutionary trajectories because accidents of geography prevent any effective contact between them or because catastrophic historical events have left a culture without a sufficient population to sustain it. It is only the long sweep of human cultural history that is meant to be progressive. The problem with such a theory is that it is hard to imagine any observation that could not be rationalized. The mere numerosity of the human species makes it impossible to return to feudal agricultural production, although a global nuclear war with a 95% mortality rate might do the trick. Would that be an example of specific or general cultural evolution?

Third, transformational evolution demands a mechanism, or at the very least, a set of empirical law-like regularities that are characteristic of all times and places, even if these cannot be generated from lower level mechanical principles. Transformational theories of cultural evolution, to the extent that they attempt to generate putative trends from some lower level principles at all, usually do so from middle level laws of the same ontological status as Ibn Khaldun's generative rules, rather than deriving them explicitly from properties of human beings and their consequent interactions in assemblages, as Hobbes did. *Evolution and Culture* provides a "Law of Cultural Dominance" that assures that more advanced cultures will spread and replace the less advanced when they come in contact, and a "Law of Evolutionary Potential" that asserts that the more specialized and adapted to local circumstances a culture is, the less likely it is to progress to a higher stage. Beyond appealing to the reasonable notion that cultures that control more energy are likely to take over those that control less, provided they do not destroy themselves in the meantime, and the rather more ideological prejudice that progress comes from struggle, no lower level mechanisms are adduced that generate these laws.

Although transformational theories do not have carefully articulated lower level mechanisms providing the mediation for the law-like higher level properties that are claimed, there is general agreement on elements that would go into such a theory of mediation. Human beings have certain properties:

1. They have great physical power to alter their surrounding circumstances;
2. They have self-reflexive consciousness so they can assess and react to their own psychic states;
3. They can imagine and plan what does not yet exist, so they can invent novelties;
4. They have a recursive linguistic function that allows them to communicate complex hypothetical structures and causal assertions;
5. They are always born and develop psychically in group contexts.

These properties are sufficient to allow groups of human beings to generate a variety of artifacts, activities, and group relations, to decide how well these satisfy their physical and psychic desires, to consciously plan and alter their activities and beliefs, and to pass information about these activities and beliefs between individuals and across generational boundaries, and they generate the possibility of coercing or convincing other groups to adopt particular patterns of activity.

The problem with this list of properties of human beings and the powers that derive from them is that they contain no assertions about the nature of the transformation of individual properties into group properties and structures, or the way in which individuals are transformed by the group, or the manner in which group properties have their own dynamic relationships. That is, there is no social theory or psychosocial theory. Of course, a completely atomistic and reductionist evolutionary theory would not require such a social theory, but no transformational theory of cultural evolution denies the relevance of social and psychosocial causes. There is simply no agreement on what these are or how they would generate the ‘‘laws’’ of directionality and progress. It has remained for variational theories of cultural evolution to play the reductionist game.

III.B The Paradigms of Cultural Evolutionary Theory: Variational Theories of Cultural Evolution

Variational models for cultural evolution have appeared in the last twenty years as a concomitant of the invention of sociobiology and its transformation into evolutionary psychology. It was the intention of sociobiology to give an orthodox Darwinian explanation of the origin of major features of human culture like religion, warfare, family structure, and so on, as manifestations of the higher reproductive rate of individuals with certain behavioral properties, but not to explain changes that have occurred in the forms of those phenomena during the process of human history. Indeed, the chief evidence offered for the origin of these features through biological, genetic evolution was precisely that they were universal. All human cultures have religion, all engage in warfare, and E. O. Wilson claimed that male domination in society would persist indefinitely.[21] The ambition to extend classical Darwinism to the explanation of all aspects of species life, including species social behavior, resulted in an immense popularity of adaptive evolutionary thinking in fields like economics, political science, and psychology that were in search of more ''scientific'' explanatory schemes. One result of this intellectual fashion was, ironically, the creation of formal Darwinian models of differentiation and temporal change of social institutions, but without the biological genetic content of organic evolution. It is important to stress that Darwinian theories of the evolution of human cultural diversity in time and space are emphatically not theories that this diversity is based in genetic differences and that genetic evolution is at the base of the change from agricultural to industrial societies, or the development of the centralized state. Instead, a variety of theories of cultural evolution have been created that are *isomorphic* with the skeletal structure of Darwinian evolutionary theory, substituting for its various concrete biological elements analogical features from culture.

The skeletal structure of the Darwinian variational scheme for organic evolution consists of three assertions:

1. Individual organisms within populations vary from one another in their characteristics. This variation arises from causes within organisms that are orthogonal to their effects on the life of the organism (The Principle of Random Variation).
2. Offspring resemble their parents (and other relatives) on the average more than they resemble unrelated organisms (The Principle of Heredity).
3. Some organisms leave more offspring than others (The Principle of Differential Reproduction). The differential reproduction may be a direct causal consequence of the characteristics of the organism (natural selection), or it may be a statistical variation that arises from purely random differential survival. This latter possibility is often ignored in vulgar expositions of Darwinian evolution, and all changes are ascribed to natural selection, but it is now certain that a great deal of evolution, especially molecular evolution, is a consequence of stochastic variations in reproduction.

If there is no variation among organisms, then even if different individuals leave different numbers of offspring, nothing will change. If there were no heredity of characteristics, then even if different organisms left different numbers of offspring, there would be no effect on the characteristics of the next generation. Finally, if different organisms all left exactly the same number of offspring no change would be expected in the composition of the population. In order to produce a scheme of cultural evolution that is isomorphic with the Darwinian variational structure there must be analogs of its elements.

The production of those analogs has occupied a great many people in a variety of disciplines over the last few decades. With so many competing models produced, it is hardly surprising that there is a great deal of spirited debate among the authors of the large and expanding literature on cultural evolution.[22] But however full of sound and fury, this debate is essentially an intramural affair. For beneath all the differences in details, there is a paradigmatic unity among Darwinian theories of cultural evolution based on the assumption that cultural evolution can and must be explained in terms isomorphic with the three principles of Darwin's variational scheme. Before they can proceed with that explanation, however, cultural evolutionists undertake a clean-up project, accomplished through sleights of conceptual hand, that clears away anything between the "biological" and the "cultural" that might have a constitutive effect in the production and "evolution" of cultural forms. This entails first of all the disappearance of the social or, at least, depriving the social of causal efficacy, and then the neutralizing of culture.

The easiest way to make society disappear is simply to dissolve it by definitional fiat into a mere population. E. O. Wilson, for example, writes: "When societies are viewed strictly as populations, the relationship between culture and heredity can be defined more precisely."[23] Robert Boyd and Peter Richerson state rather categorically that "cultural evolution, like genetic evolution in a sexual species, is always a group or population phenomenon"; and in a later work: "because cultural change is a population process, it can be studied using Darwinian methods."[24] A more nuanced way of dissolving society into a collection of atomistic individuals is to create a choice between two extreme alternatives. Melvin Konner, for example, correctly rejects the society-as-organism metaphor by contrasting the cell that is devoted "entirely to the survival and reproduction of the organism" with "the purposes of the individual human [that] are wedded to the survival and reproduction of the society only transiently and skeptically." But he overdraws the consequences of this obvious insight and concludes that evolution "has designed the individual with a full complement of independence and a canny ability to subvert, or at least try to subvert, the purposes of society to its own. Every time a human being gets fed up with his or her society or church or club or even family, and voluntarily changes affiliation, we have another factual disproof of the central metaphor of social and political science."[25] Here he assumes that the repudiation of the obviously false metaphor of society as organism is a justification for an equally obviously false atomistic individualism that renders society a mere population.

However accomplished, the dissolution of societies into populations or, as in more nuanced approaches, the reduction of differential social power to the status of a subordinate variable,[26] precludes the possibility that social systems might have properties unique to them as organized systems, that is, that social relations might be characterized by structures of unequal power that affect individual social behavior and the fitness of cultural traits. This dissolution means, in turn, that social hierarchy and inequality are explained as just the consequence of the differential cultural fitness of individuals or of the cultural traits they bear, rather than, say, as a consequence of antagonistic and exploitative social relations.[27]

Having taken the crucial preliminary step of dissolving society, the next step is, perhaps surprisingly, to neutralize culture as well. In order to qualify as an instance of a variational theory of evolution, culture must be proven to consist of isolatable, individual entities, and to be only the sum of its parts. It is thus necessary to refute any and all claims that cultures have unique and discrete properties and a system-specific logic that require them to be analyzed each on its own terms. This is sometimes done by definitional fiat aimed at another superorganismic straw man. E. O. Wilson, for example, insists that ''cultures are not superorganisms that evolve by their own dynamics.'' Culture, concurs Jerome Barkow, ''is not a 'thing,' not a concrete, tangible object. It isn't a cause of anything. To describe behaviour as 'cultural' tells us only that the action and its meaning are shared and not a matter of individual idiosyncrasy.''[28]

The definitional fiats that posited population-like models of culture received at least two slight challenges. Discontent with an excessively atomistic view of culture, Bernardo Bernardi, for example, constructs a constellation of ''anthropemes'' consisting of ''ethnemes,'' themselves subdivided into ''idioethnemes'' and ''socioethnemes''; and Martin Stuart-Fox divides memes into mentemes.[29] Though these attempts appear to reject the notion of isolated, individual memes and to aim at systematic complexity, they fall short. Tellingly, in suggesting the division of the meme into mentemes, Stuart-Fox quite consciously attempted to construct a categorial analogy with modern linguistic terminology. But he did not follow up this overture and consider Saussure's fundamental insight on which modern linguistics is based, namely that meaning is system-specific, that each term (sign) acquires its historically-specific meaning by virtue of its place within a discrete set of differential relations. By neglecting this insight, attempts such as Stuart-Fox's and Bernardi's focus only on the aggregate rather than the systemic. Only additive in method, they treat memes as aggregates of smaller entities, as cultural molecules composed of cultural atoms—which effects only a slight displacement of their ontological individualism, reproducing it at the level of compounds.

Coevolutionists have also made overtures to the systemic character of culture by removing it from a tight genetic leash and insisting that culture evolves relatively autonomously on its own cultural track. But regardless of the number of evolutionary tracks advocated, all theories of cultural evolution pay only lip service to the complexity of culture: because they persist in treating culture as merely the sum total of individual cultural units at a given stage in the selection process, as a kind of ''state of the 'memes' '' at a given point in time, they deny culture any system-specific characteristics; and this, in turn, allows all cultures to be explained according to the same (transhistorical and therefore ahistorical) selectionist logic.

With society and culture reduced to mere aggregates and deprived of any systemic and system-specific characteristics, the ground is prepared for the construction of a scheme of cultural evolution that is isomorphic with the Darwinian variational structure. This, as mentioned above, requires the construction of cultural analogs to the three fundamental principles of the Darwinian variational scheme.

First, a decision has to be made about the Principle of Random Variation, about the identity of the objects that have variation, heredity, and differential reproduction. Are these objects individual human beings who are the bearers of different cultural characteristics and who pass on those characteristics to other human beings by various means of social and psychological communication, and who have differential numbers of cultural "offspring"? This is the approach generally favored by those focusing on behavior and defining cultural in behavioralist terms. Or are they the characteristics themselves with properties of heredity and differential reproduction? This is the more common approach in recent years, especially among the "coevolutionists" who have taken an "ideational" view of culture using so-called "trait-based" models of the evolutionary process. An example of the former is Cavalli-Sforza's and Feldman's theory of cultural transmission, while Dawkins's "memes" are an example of the latter.[30]

Either way, a fundamental problem results from the assumption that these cultural units, say the idea of monotheism, or the periphrastic "do," somehow spread or disappear in human populations, namely: no theory of cultural evolution has provided the elementary properties of these abstract units. Presumably they are mortal and so need rules of heredity. But, for a variational theory, it must be possible the count up the number of times each variant is represented. What is the equivalent for memes of the number of gene copies in a population? Perhaps it is the number of individual human beings who embody them, but then the death of a human carrier means the loss of a meme copy and so memes do, after all, have the problem of heredity. A major problem of creating a variational theory of cultural evolution is that the task of building a detailed isomorphism has not been taken seriously enough.

Once the individual units are settled upon, little time is spent determining the sources of variation in those units, the "cultural analogs of the forces of natural selection, mutation, and drift that drive genetic evolution."[31] Following a quick definitional determination of the sources of variation—randomness and drift, selection, and perhaps the addition of a uniquely cultural source such as intentionality—the next step is to find the cultural analogs to the Principle of Heredity.

Most cultural evolutionists simply accept as given that culture is a system of heredity or at least of unidirectional transmission. Boyd and Richerson state axiomatically that "Darwinian methods are applicable to culture because culture, *like genes*, is information that is transmitted from one individual to another" (emphasis added). In a later essay they turn inheritance into the defining characteristic of cultural evolutionary theory: "The idea that unifies the Darwinian approach is that culture constitutes a system of inheritance"; and after a brief discussion that moves from inheritance through the "population-level properties" of culture that makes it "similar . . . to gene pools," they conclude that "because cultural change is a population process, it can be studied using Darwinian methods."[32]

To be sure, however, Boyd and Richerson spoke a bit too inclusively. While some cultural evolutionists use "inheritance" and "transmission" interchangeably, others are uneasy about the genetic and parental overtones of "inheritance" and prefer "transmission." But both terms refer to a process of descent that occurs in the same unidirectional manner between an active donor and a passive recipient. The semantic advantage of "transmission" is that it drops the genetic connotational baggage of "inheritance" while preserving the portrayal of cultural change as a unidirectional process of descent with modification and selection.

Whether conceptualized as "heredity" or "transmission," however, the problematic issue is that both terms require the establishment of some laws of the heredity of units or their characteristics if human individuals are the units. We then require the details of the passage of culture to new individuals, by analogy with the Mendelian mechanism of the passage of genetic information from parent to offspring by way of DNA. In making this analogy, however, the biological model implies constraints that have not been apparent to cultural evolutionists. We say that parents "transmit" their genes (or at least copies of their genes) to their offspring, so models of cultural evolution begin with models of the "transmission" of cultural traits from one set of actors to others by analogy with the transmission of genes. Parents may transmit traits to their children, or teachers to their pupils, or siblings and other peers to each other by a variety of simple rules. The outcomes of evolutionary models of this kind turn out to be extremely sensitive to the postulated rules of transmission, and since there is no firm basis on which to choose the rules, almost anything is possible. But there is a deeper problem. Is culture "transmitted" at all? An alternative model, one that accords better with the actual experience of acculturation, is that culture is not "transmitted" but "acquired." Acculturation occurs through a process of constant immersion of each person in a sea of cultural phenomena, smells, tastes, postures, the appearance of buildings, the rise and fall of spoken utterances. But if the passage of culture cannot be contained in a simple model of transmission, but requires a complex mode of acquisition from family, social class, institutions, communications media, the work place, the streets, then all hope of a coherent theory of cultural evolution seems to disappear. Of course, it was simpler in the Neolithic, but there was still the family, the band, the legends, the artifacts, the natural environment.

Some dissenters present serious challenges to the inheritance/transmission model even though they remain faithful to its explanatory principle. Martin Daly questions the value of the inheritance model because he finds no cultural analog to the gene, because cultural traits "are not immutable" like genetic traits, because cultural "transmission need not be replicative," because the recipients are not "simply vessels to be filled,'' and because ''social influence'' makes the processes of cultural change less regular than is implied by the term ''transmission.''[33] Though Daly and others raise perfectly legitimate and very important questions about inheritance and transmission analogies, they deprive their insights of real force by still maintaining that cultural change is a process that can and must be explained in terms isomorphic with ''the evolutionary model of man.''[34]

This assumption brings us to the third analogical element in theories of cultural evolution, the Principle of Differential Reproduction. Whether they define the units as cultural atoms or cultural molecules, whether they speak of cultural change as inheritance, or of transmission to passive recipients or to active acquisitors, they all insist that cultural change is a process of *descent* with modification; and as such it has all the attributes of a variational evolutionary process eligible for Darwinian, that is, selectionist explanation. To all cultural evolutionists may be extended that which Martin Stuart-Fox said of himself, namely that they ''*take for granted* (a) the scientific status of the synthetic theory of evolution and (b) that this theory provides *the most likely model* on which to base a theory of cultural evolution'' (emphasis added).[35]

However, the forces that cause the differential passage of culture across generations and between groups seem not to be encompassed by the reductionist model in which individual actors have more cultural offspring by virtue of their persuasiveness or power or the appeal of their ideas, or in which memes somehow outcompete others through their superior utility or psychic resonance. Atomistic models based on the characteristics of individual humans or individual memes can be made, but they appear as formal structures with no possibility of testing their claim to reality. How are we to explain the disappearance of German and French as the languages of international scientific discourse, and their universal replacement by English without terms like ''Nazi persecution of Jews,'' ''industrial output,'' ''military power in the Cold War,'' or ''gross national product.'' That is, no variational theory of cultural change can be adequate if it attempts to create a formal isomorphism with Darwinist individualism.

Historical, political, social, and economic phenomena, in short, must be dismantled in order to be molded into the raw material for selectionist theories of cultural evolution. This is effected through the dissolution of social systems with structural asymmetries of power into individuals; and through the reduction of cultural systems to eclectic aggregates of differentially reproduced memes. This dual process strips historical phenomena of their sociocultural particularity. Once transformed in this way, they may be subjected to nomological explanation as individual instances of the exogenous, because transhistorical, law of selection. Even the recognition given by William Durham and others to the systemic character of culture and to the possibility that social asymmetries of power might affect cultural transmission and fitness are drained of content by the fundamental assumptions of the cultural evolutionist paradigm: the definition of culture as an aggregate of individual, heritable units and the selectionist explanation of its evolution. And in these assumptions lies the self-validating circularity of cultural evolutionary theories: selectionist explanation requires individual, heritable units of culture; and reduction of culture to an aggregate of such units renders it susceptible to selectionist explanation—whose scientific status had been taken for granted from the very beginning.

As its etymology suggests, any "theory" is a way of looking at the world, and what one sees is that which is visible through one's particular set of theoretical lenses. Cultural evolutionary theories, however, base (and wager) their claim to break through all theoretical biases and to attain scientific status on their verifiability, their ability to postdict past and predict future cultural evolution. If, with the emergence of the hegemony of the physical sciences, the cornerstone of a scientific theory has been the elimination of the historical, its touchstone has been its predictive capacity—a matter that cultural evolutionists address with increasing confidence.

We have already encountered Alexander Rosenberg's optimism about the use of mathematical models in the new sociobiologically based social sciences and his confidence in their predictive capacities.[36] The same optimism is prevalent among the contributors to *Politics and the Life Sciences* who are convinced that the predictive powers of the new evolutionary political science will render it capable of informing policy decisions. Certain that Darwinian models of cultural evolution can produce "a useful retrodiction of ethnography," Lumsden and Wilson were somewhat circumspect, anticipating only predictions of "short term changes in the forms of ethnographic distributions." Nevertheless, they remained—and Wilson has become ever more—optimistic that "the history of our own era can be explained more deeply and more rigorously with the aid of biological theory," and that this approach might enable us to look "down the world-tube of possible future histories."[37] Similarly, Boyd and Richerson quickly overcame their initial caution to assert that "Darwinian models can make useful predictions."[38]

Though they wager the validity of their theories on their predictive capacities, theorists of cultural evolution rig the explanatory game in a variety of ways. One is by covering all bets. This can be done by playing with probabilistic explanations. In the gambling hall, probabilities only provide the odds, but probabilistic predictions of cultural evolution are guaranteed winners, since they encompass all possibilities. Because, for example, of our evolved capacity to reason we could be soberly advancing down the road towards wisdom, courage, and compassion; or because of our innate capacity for aggression we could be headin' for nuclear armageddon—or anything in between. Or it can be done by constructing a historical analog to random drift in theories of biological evolution—the catch-all explanation of that which cannot be subsumed under selection.

A second way to rig the game is with postdictive readjustment. The cultural evolutionist, like the economist, is "an expert who will know tomorrow why the things he predicted yesterday didn't happen today."[39] The gambler's losses might be recouped in a later game, but cannot be undone. But in cultural evolutionary explanation and prediction, the game may be replayed indefinitely until the model is successfully readjusted. Combined with probabilistic explanations, postdictive readjustment renders the model invulnerable by disarming its weaknesses.

The irony here is that the constant recourse to postdictive readjustments brings the science of cultural evolution into the neighborhood of "just plain history"—almost. The difference is that the faith in the scientific status of the law of selection erects a third safeguard for theories of cultural evolution. This belief precludes as "not scientific" any non-evolutionary, that is, historical, explanation of cultural change. But because cultural evolutionary theories are based on a unitary, transhistorical principle, they produce explanations that are too broad to be either falsifiable or explanatory.

Historians, cultural evolutionists argue, are too close to the fray, and their time scales too short—which leads them into all kinds of unimportant detours and false starts that appear to the historical eye as enterprises of great pith and moment. To gain proper perspective, therefore, cultural evolutionists draw back, occasionally indulging in imaginary space travel, in order to attain a sufficiently distant viewpoint from which to view the human species as one among many and to avoid the "anthropocentrism" that would exempt culture (a biological adaptation) from biological explanation. But distance can also be deceiving.

From their distant viewpoint cultural evolutionists willingly see only the broad patterns of cultural evolution, and ignore the inconvenient and contingent details of history that do not fit into those patterns. This conscious oversight produces theories of cultural evolution that are explicitly or implicitly progressivist: since culture is a successful and cumulative adaptation that breaks free of natural selection, the more culture, the better for human welfare and survival. This linear logic points to the contemporary West with the most advanced level of science and technology (the ultimate cultural adaptations insuring human welfare and survival) as the current pinnacle of cultural evolution. But the road to modern Western civilization has taken a series of abrupt and thoroughly unpredictable turns. What general theory of cultural evolution could postdict the collapse of the Roman Empire and the "Dark Ages"? Or the emergence on a distant frontier of the Eurasian landmass of a new geo-cultural entity, a "continent" called Europe? Or that in a very brief historical time span this new culture would overtake much more advanced Asian cultures and establish itself as the most powerful and dominant in the world, with one of its tiny "populations," the English, having acquired an empire on which the sun never set? But the result of all those unpredictable turns, the late modern West, which should be the pinnacle of cultural evolution, has been the epitome of barbarism (which only a small group of *fin de siècle* artists and intellectuals, members of the "literary culture," suspected).

From their distant viewpoint, cultural evolutionists may ignore acts of barbarism in Western history like the genocide of Native Americans or the Nazi Holocaust as just specks of dust on the plain of history, momentary aberrations irrelevant to the question of cultural evolution. Alternatively, they may subject both to the same explanatory principle as just two examples of human aggression explained through some selectionist variation or combination of inclusive fitness, innate aggression, the stress of overpopulation, and/or the need for *Lebensraum*. But to explain the character, causes, and consequences of these two forms of genocide according to the same transhistorical principle would lead to a gross misunderstanding of each and would tell us little about their historically and politically significant differences. Such an approach, for example, is far too broad either to postdict the success of Nazism or to predict the ongoing consequences of the Nazi period, of the historical memory that continues to affect significantly the history not only of Germany and Europe, but also of the Middle East. Whether they forcibly subsume disparate historical phenomena under a transhistorical explanatory principle or write off as mere contingencies historically significant events that cannot be so subsumed, cultural evolutionary theories cannot answer the many crucial questions pertaining to the particularity, the uniqueness, of all historical phenomena. In failing to live up to their own claims to be able to explain history, including that of our own era, "more deeply and more rigorously," cultural evolutionary theories also fail to live up to their further claim to explain history more "usefully"—to explain Nazism, for example, with sufficient precision to prevent its reoccurrence and to develop appropriate policies to deal with its consequences.

It is therefore no use to fall back on yet another safeguard, the claim that the field is still young, the models are still being built, and one day.... The problem is more serious than "not yet enough time." Cultural evolutionary theories are carefully constructed, logically consistent, and very neat. Their neatness, however, is achieved either by dismissing as inessential to cultural evolution the contingencies that are so essential to historical change or by subsuming them to a single transhistorical principle of explanation. But this formulaic treatment is fully inappropriate to the labyrinthine pathways, the contingent complexity, the many nuances, and general messiness of history. And it results in linear explanations that approach closely enough to history to allow the distant observer to mistake proximity for causality. These analytical lines, however, are actually false tangents—briefly nearing, but never touching, the contours of history.

We conclude, finally, by returning to the question of whether any useful work is done by considering cultural evolution as distinct from the history of human societies. Transformational theories of cultural evolution have the virtue that they at least provide a framework of generality with which to give human long-term history the semblance of intelligibility. But the search for intelligibility should not be confused with the search for actual process. There is no end of ways to make history seem orderly.

Variational isomorphisms with Darwinian evolution suffer from the inverse problem. Rather than being so flexible as to accommodate any historical sequence, they are too rigid in structure to be even plausible. They attempt to mimic, for no reason beyond the desire to appear scientific, a theory from another domain, a theory whose structure is anchored in the concrete particularities of the phenomena that gave rise to it.

Notes

1. C. P. Snow, *The Two Cultures and a Second Look* (Cambridge, Eng., 1964), 8–9.

2. *Ibid.*, 70.

3. Herbert Spencer, *The Principles of Biology* [1867] (New York and London, 1914), 432–433.

4. Leslie White, Preface to *Evolution and Culture*, ed. Marshall Sahlins and Elman Service (Ann Arbor, 1960), v.

5. *Ibid.*, vii.

6. E. O. Wilson, *On Human Nature* (Cambridge Mass., 1978), x.

7. *Ibid.*, 13.

8. *Ibid.*, 34.

9. Alexander Rosenberg, *Sociobiology and the Preemption of Social Science* (Baltimore, 1980), 22–23.

10. *Ibid.*, 4, 158.

11. Leda Cosmides and John Tooby, "The Psychological Foundations of Culture," in *The Adapted Mind*, ed. Jerome H. Barkow, Leda Cosmides, and John Tooby (Oxford, 1992), 22–23.

12. Richard Shelly Hartigan, "A Review of *The Biology of Moral Systems*" by Richard D. Alexander, in *Politics and the Life Sciences* 7, no. 1 (1988), 96.

13. See for example the following essays, all from *Politics and the Life Sciences*: Elliot White, "Self-selection and Social Life: The Neuropolitics of Alienation—The Trapped and the Overwhelmed" (vol. 7, no. 1, 1989); John H. Beckstrom, "Evolutonary Jurisprudence: Prospects and Limitations on the Use of Modern Darwinism Throughout the Legal Process" (vol. 9, no. 2, 1991); Lee Ellis, "A Biosocial Theory of Social Stratification Derived from the Concepts of Pro/Anti-sociality and r/K Selection" (vol. 10, no. 1, 1991); Hames N. Schubert *et al.*, "Observing Supreme Court Oral Argument: A Biosocial Approach" (vol. 11, no. 1, 1992); Larry Arnhart, "Feminism, Primatology, and Ethical Naturalism" (vol. 11, no. 2, 1992).

14. Ibn Khaldun, *The Muqaddimah* (Princeton, 1958), chapter 2: 24.

15. *Ibid.*, chapters 3: 3, 3: 46, 3: 11.

16. *A Critical Dictionary of the French Revolution*, ed. François Furet and Mona Ozouf (Cambridge, Mass., 1989), xvi.

17. Leslie White in *Social and Cultural Evolution*, volume III: *Issues in Evolution*, ed. S. Tax and C. Callender (Chicago, 1960), Panel 5.

18. In the introduction to a collection of essays on *History and Evolution*, ed. Matthew Nitecki and Doris Nitecki (Albany N.Y., 1992) Matthew Nitecki states categorically: "The common element of evolutionary biology and history is the concept of *change* over time" (6). Despite great differences in their definitions of the relation between biology and history, all authors included in the volume (and probably all cultural evolutionists) share Nitecki's definition of history as change over time. This definition allows them to elevate history to scientific status and to subject history to evolutionary explanation, as do, for example, Boyd and Richerson in their categorical claim that "Darwinian theory is both scientific and historical" (in Nitecki, 179–180). While there are many problems with defining history simply as change over time, we will only make two comments here: this definition almost inevitably results in treating people living in societies in which change is not the norm as people without history; and once change is defined as a transhistorical constant, it is very likely, though not logically necessary, that the next step will be to seek a transhistorical explanatory law—which for cultural evolutionists is that of selection.

19. *Evolution and Culture* is the work of four authors, each of whom contributed a chapter to the volume: Thomas Harding, David Kaplin, Marshall Sahlins, and Elman Service. The most influential of these has been Sahlins's "Evolution: Specific and General." Sahlins's approach to culture has, of course, evolved considerably since 1960.

20. M. Sahlins, "Evolution: Specific and General," in *Evolution and Cuture*, ed. Sahlins and Service, 20.

21. E. O. Wilson, "Human Decency Is Animal," *New York Times Magazine* (October 12, 1975).

22. On the basis of the two questions he asks of each cultural evolutionary theory, William Durham in *Coevolution: Genes, Culture, and Human Diversity* (Stanford, 1991) is able thoroughly to survey the cultural evolutionist plain and, in so doing, he provides a sense of its paradigmatic unity. His questions are: "Is culture a second inheritance system? What are the best units to use in the study of cultural transmission?" (155). Based on the responses, he establishes in an "approximate[ly] chronological" order (155) a tripartite division of the cultural evolutionist terrain. The earliest theories of cultural evolution tended to be "models without dual inheritance." These conceptualize culture not "as part of the phenotype"; and they explain "phenotypic change in human populations in terms of a single fitness principle, namely, reproductive fitness in one of its guises" (155–156). Examples include: both the stronger and weaker versions of sociobiology, David Barash's "genetic determinism," and the "genes on a leash" model of Wilson's *On Human Nature*. The second type, "models with dual inheritance and trait units," also conceptualizes culture "as part of the phenotype" but views it "as a second, nongenetic inheritance system whose units are defined as culturally heritable aspects of phenotype. These units are recognized as having their own measure of fitness within the cultural system [i.e. 'cultural fitness']. . . . The differential transmission of traits or behaviors within a population constitutes cultural evolution" (156). Examples are: Cavalli-Sforza's and Feldman's cultural transmission model, Richard Alexander's social learning model, Lumsden's and Wilson's gene-culture transmission model, and his own early coevolution model. The third type is "models with dual inheritance and ideational units." These

also treat culture ‘‘as a separate ‘track’ of informational inheritance,’’ but they focus not on phenotypical traits, but the differential transmission of ideas, values, and beliefs in a population’’ (156). Examples include: Albert G. Keller’s social selection model, H. Ronald Pulliam’s and Christopher Dunford’s ‘‘programmed learning model,’’ Boyd’s and Richerson’s ‘‘Darwinian culture theory,’’ and Durham’s own book, *Coevolution*. The ‘‘evolution’’ of theories of cultural evolution has, as Durham indicated, roughly followed the sequence of his types: from behavioral to (Clifford Geertzian) ideational definitions of culture and from single inheritance, culture-on-a-genetic-leash models to dual inheritance models. This evolution resulted from dissatisfaction with single inheritance models for tying culture to a short genetic leash, and with behavioral definitions of culture for their unreliability, the impossibility of knowing precisely which meme motivates a given behavior.

23. Wilson, *On Human Nature*, 78.

24. Robert Boyd and Peter Richerson, *Culture and the Evolutionary Process* (Chicago, 1985), 292; and ‘‘How Microevolutionary Processes Give Rise to History’’ in *History and Evolution*, ed. Nitecki and Nitecki, 181.

25. Melvin Konner, *The Tangled Wing: Biological Constraints on the Human Spirit* (New York, 1982), 414.

26. Unaware of the implications of their reduction of societies to populations, Boyd and Richerson, much to their surprise, found themselves criticized by David Rindos (‘‘The Evolution of the Capacity for Culture: Sociobiology, Structuralism, and Cultural Selectionism,’’ *Current Anthropology* 27 [1986], 315–316) and William Durham (*Coevolution*, 179ff.) for not having adequately addressed the social. In their direct response to Rindos (included in Rindos, 327), Boyd and Richerson claim, correctly, that they spent an entire chapter of their *Culture and the Evolutionary Process* on ‘‘the scale of human social organization,’’ implying, incorrectly, that therewith the matter was resolved. That chapter first develops a taxonomy of biases (direct, indirect, and frequency dependent) and then constructs models to analyze how the frequency of these biases affects the transmission of culture. Though such biases certainly affect social behavior, their origins and persistence are nowhere discussed. Consequently, they end up explaining how social biases affect individual choice by transforming clichés into explanatory principles: ‘‘When in Rome, do as the Romans do’’ becomes the law of ‘‘frequency-dependent bias’’ (286) and ‘‘keeping up with the Joneses’’ the law of ‘‘indirect bias’’ (287). The questions of whether all Romans do as some Romans do or of whether keeping up with the Joneses makes sense in societies not based on commodity production and exchange are crucial questions that disappear in their biases.

Durham makes perhaps the most concerted effort to consider asymmetries of social power and the ‘‘imposition’’ of group values on individual ‘‘choice’’ (*Coevolution*, 198–199). He identifies ‘‘reference groups’’ within a given population, thereby acknowledging ‘‘the simple fact that cultural evolution is an intrinsically political process’’ (211). Because he does not ask the essential questions of why particular ‘‘reference groups’’ exist and what is the distinctly and discrete social logic behind particular asymmetries in group power, Durham can only treat any particular set of reference groups and social asymmetries of power as arbitrary and subordinate variable factors affecting individual choice, rather than as constitutive factors of social and cultural forms and their ‘‘evolution.’’

27. The most perspicacious critic of theories that reduce societies to populations was Karl Marx. The hallmark of political economy and the source of its errors, Marx argued, was that it took as its starting point the population without having determined the components of the populations, its "subgroups" or classes *and* the logic of their internal relations. Such an approach would produce not "a rich totality of many determinations and relations," but "ever thinner abstractions" and "a chaotic conception of the whole" (*Marx-Engels Reader*, ed. Robert Tucker [New York, 1978], 237). Or as he later summarized it more succinctly: "[s]ociety does not consist of individuals, but expresses the sum of interrelations, the relations within which these individuals stand" (247). The analysis of a society reveals much about its population, but the converse is not necessarily true.

28. E. O. Wilson, *On Human Nature*, 78; Jerome Barkow, *Darwin, Sex, and Status: Biological Approaches to Mind and Culture* (Toronto, 1989), 142.

29. Bernardo Bernardi, "The Concept of Culture: A New Presentation" in *The Concepts and Dynamics of Culture*, ed. Bernardo Bernardi (The Hague, 1977); Martin Stuart-Fox, "The Unit of Replication in Socio-cultural Evolution," *Journal of Social and Biological Structures* 9 (1986), 67–90.

30. L. L. Cavalli-Sforza and M. W. Feldman, *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton, 981); Richard Dawkins, *The Selfish Gene* (New York, 1976).

31. See Cavalli-Sforza and Feldman, *Cultural Transmission and Evolution*, 10; Boyd and Richerson, *Culture and the Evolutionary Process*, 8ff, and "How Microevolutionary Processes Give Rise to History," 182.

32. Peter J. Richerson and Robert Boyd, "A Darwinian Theory for the Evolution of Symbolic Cultural Traits," in *The Relevance of Culture*, ed. Morris Freilich (New York, 1989), 121; Boyd and Richerson, "How Microevolutionary Processes Give Rise to History," 181.

33. Martin Daly, "Some Caveats about Cultural Transmission Models," *Human Ecology* 10 (1982), 402–404.

34. *Ibid.*, 406. See also David Hull, "The Naked Meme," in *Learning, Development, and Culture: Essay in Evolutionary Epistemology*, ed. H. C. Plotkin (Chichester, Eng., 1982); H. Kaufman, "The Natural History of Human Organizations," *Administration and Society* 7 (1975); Timothy Goldsmith, *The Biological Roots of Human Nature: Forging Links Between Evolution and Behavior* (New York, 1991).

35. Stuart-Fox, "The Unit of Replication in Socio-Cultural Evolution," 68.

36. Rosenberg, *Sociobiology and the Preemption of Social Science*, 151.

37. Charles Lumsden and E. O. Wilson, *Genes, Mind and Culture: The Coevolutionary Process* (Cambridge, Mass., 1981); 358, 360, 362. See also Wilson, *Consilience* (Cambridge, Mass., 1998).

38. Boyd and Richerson, *Culture and the Evolutionary Process*, 25; "How Microprocesses Give Rise to History," 203.

39. Laurence J. Peter, *Peter's Quotations: Ideas for Our Times* (New York, 1977), 477.

25 Models of Cultural Evolution

Elliott Sober

As least since the time of Darwin, there has been a tradition of borrowing between evolutionary theory and the social sciences. Darwin himself owed a debt to the Scottish economists who showed him how order can be produced without conscious design. Adam Smith thought that socially beneficial characteristics can emerge in a society as if by an ''invisible hand''; though each individual acts only in his or her narrow self-interest, the result, Smith thought, would be a society of order, harmony, and prosperity. The kind of theory Darwin aimed at—in which fitness improves in a population without any conscious guidance—found a suggestive precedent in the social sciences.

The use of game theory by Maynard Smith[1] and others provides a contemporary example in which an idea invented in the social sciences finds application in evolutionary theory. Economists and mathematicians were the first to investigate the payoffs that would accrue to players following different strategies in games of a given structure. Biologists were also to see that game theory does not require that the players be rational or even that they have minds. The behavior of organisms exhibits regularities; this is enough for us to talk of them as pursuing strategies. The payoffs of the behaviors that result from these strategies can be measured in the currency of fitness—i.e., in terms of their consequences for survival and reproduction. This means that the idea of payoffs within games allows us to describe evolution by natural selection. Here again is a case in which a social scientific idea has broader scope than its initial social science applications might have suggested.

At present, there is considerable interest and controversy surrounding borrowings that go in the opposite direction. Rather than apply social science ideas to biological phenomena, sociobiology and related research programs aim to apply evolutionary ideas to problems that have traditionally been thought to be part of the subject matter of the social sciences. Sociobiology is the best known of these enterprises. It has been

In P. Griffiths, ed., *Trees of Life: Essays in the Philosophy of Biology*, Australasian Studies in the History and Philosophy of Science (Kluwer, 1991).

criticized on a variety of fronts. Although I think that these criticisms differ in their force, I don't want to review them here. My interest is in a somewhat lesser-known movement within biology, one that strives to extend evolutionary ideas to social scientific phenomena. I want to discuss the models of cultural evolution put forward by Cavalli-Sforza and Feldman[2] and by Boyd and Richerson.[3] These authors have distanced themselves from the mistakes they see attaching to sociobiology. In particular, they wish to describe how cultural traits can evolve for reasons that have nothing to do with the consequences the traits have for reproductive fitness. In a very real sense, their models describe how it is possible for mind and culture to play an irreducible and autonomous role in cultural change. For this reason, there is at least one standard criticism of sociobiology that does not apply to these models of cultural evolution. They deserve a separate hearing.

In order to clarify how these models differ from some of the ideas put forward in sociobiology, it will be useful to describe some simple ways in which models of natural selection can differ. I focus here on natural selection, even though there is more to evolutionary theory than the theory of natural selection, and in spite of the fact that the two books I am considering sometimes exploit these nonselectionist ideas. Although there are nonselectionist ideas in these two books, the bulk of their models assigns a preeminent role to natural selection and its cultural analogs. So a taxonomy of selection models will help us see how models of cultural evolution are related to arguments put forward in sociobiology.

There are two crucial ingredients in a selection process. Given a set of objects that exhibit variation, what will it take for that ensemble to evolve by natural selection? By evolution, I mean that the frequency of some characteristic in the population changes. The first requirement is that the objects differ with respect to some characteristic that makes a difference in their abilities to survive and reproduce. Second, there must be some way to ensure that offspring resemble their parents. The first of these ingredients is called *differential fitness*; the second is *heritability*.

In standard formulations of the genetical theory of natural selection, different genes or gene complexes in a population encode different phenotypes. The phenotypes confer different capacities to survive and reproduce on the organisms that possess them. As a result, some genes are more successful in finding their way into the next generation than others. In consequence, the frequency of the phenotype in question changes. This is evolution by natural selection with a genetic mode of transmission. Note that traits differ in fitness because some organisms have more babies than others. It may seem odd to say that "having babies"[4] is one way to measure fitness, as if there could be others. My reason for saying this will become clearer later on.

The phenotype treated in such a selection model might be virtually any piece of morphology, physiology, or behavior. Biologists have developed different applications of this Darwinian pattern to characteristics of all three sorts in a variety of species. One way—the most straightforward way—to apply biology to the human sciences is to claim that some psychological or cultural characteristic became common in our species by a selection process of this sort. This is essentially the pattern of explanation that Wilson was using when he talked about aggression, xenophobia, and behavioral differences between the sexes. An ancestral population is postulated in which phenotypic differences have a genetic basis; then a claim is made about the consequences of those phenotypes for survival and reproduction. This is used to explain why the population changed to the configuration we now observe.

The second form that a selection process can take retains the idea that fitness is measured by how many babies an organism produces, but drops the idea that the relevant phenotypes are genetically transmitted. Strictly speaking, evolution by natural selection does not require genes. It simply requires that offspring resemble their parents. For example, if characteristics were transmitted by parents teaching their children, a selection process could occur without the mediation of genes.

A hypothetical example of how this might happen is afforded by that favorite subject of sociobiological speculation—the incest taboo. Suppose that incest avoidance is advantageous because individuals with the trait have more viable offspring than individuals without it. The reason is that outbreeding diminishes the chance that children will have deleterious recessive genes in double dose. If offspring learn whether to be incest avoiders from their parents, the frequency of the trait in the population may evolve. And this may occur without there being any genetic differences between those who avoid incest and those who do not. Indeed, incest avoidance could evolve in this way in a population of genetically identical individuals, provided that the environmental determinant of the behavior runs in families.[5]

In this second kind of selection model, mind and culture displace one but not the other of the ingredients found in models of the first type. In the first sort of model, a genetic mode of transmission works side by side with a concept of fitness defined in terms of reproductive output—what I have called ''having babies.'' In the second, reproductive output is retained as the measure of fitness, but the genetic mode of transmission is replaced by a psychological one. Teaching can provide the requisite heritability just as much as genes.

The third pattern for applying the idea of natural selection abandons both of the ingredients present in the first. Genes are abandoned as the mode of transmission. And fitness is not measured by how many babies an organism has. Individuals acquire their ideas because they are exposed to the ideas of their parents, of their peers, and of their parents' generation. So the transmission patterns may be vertical, horizontal, and oblique. An individual exposed to a mix of ideas drawn from these different sources need not give them all equal credence. Some may be more attractive than others. If so, the frequency of ideas in the population may evolve over time. Notice that there is no need for organisms to differ in terms of their survivorship or degree of reproductive success in this case. Some ideas catch on while others become passé. In this third sort of selection model, ideas spread the way a contagion spreads.

It is evident that this way of modeling cultural change is tied to the genetical theory of natural selection no more than it is tied to epidemiology. Rumors and diseases exhibit a similar dynamic. The spread of a novel characteristic in a population by natural selection, like the spread of an infection or an idea, is a diffusion process.

This third type of selection model has a history that predates sociobiology and the models of cultural evolution that I eventually want to discuss. Consider the economic theory of the firm.[6] Suppose one wishes to explain why businesses of a certain sort in an economy behave as profit maximizers. One hypothesis might be that individual managers are rational and economically well informed; they adjust their behavior so as to cope with market conditions. Call this the learning hypothesis. An alternative hypothesis is that managers are not especially rational or well informed. Rather, firms that are not efficient profit maximizers go bankrupt and thereby disappear from the market. This second hypothesis posits a selection process.

Note that the selection hypothesis involved here is of type III. Individual firms stick to the same market strategies, or convert to new ones, by some process other than genetic transmission. In addition, the biological kind of survival and reproduction (what I have called "having babies") does not play a role. Firms survive differentially, but this does not require any individual organism to die or reproduce.

A different example of type III models, which will be familiar to philosophers of science, is involved in some versions of evolutionary epistemology. Karl Popper suggested that scientific theories compete with each other in a struggle for existence.[7] Better theories spread through the population of inquirers; inferior ones exit from the scene. Popper highlighted the nonbiological definition of fitness used in this view of the scientific process when he said that "our theories die in our stead."[8]

The three possible forms that a selection model can take are summarized in table 25.1. By "learning," I don't want to require anything that is especially cognitive; imitation is a kind of learning. In addition, "having students" should be interpreted broadly, as any sort of successful influence mediated by learning.[9]

Table 25.1

Three types of selection models

	Heritability	Fitness
I	Genes	Having babies
II	Learning	Having babies
III	Learning	Having students

Note: The description of Type III models, in which fitness is measured by "having students," is due to Peter Richerson.

The parallelism between type I and type III models is instructive. In the type I case, individuals produce different numbers of babies in virtue of their phenotypic differences (which are transmitted genetically). In the type II case, individuals produce different numbers of students in virtue of their phenotypic differences (which are transmitted by learning).

Selection models of cultural characteristics that are of either pattern I or pattern II can properly be said to provide a "biological" treatment of the characteristic in question. Models of type III, on the other hand, do not really propose biological explanations at all. A selectional theory of the firm, or a diffusion model that describes the spread in popularity of an idiom in a language, are no more "biological" than their competitors. In type III models, the mode of transmission and the reason for differential survival and replication may have an entirely autonomous cultural basis. Genes and having babies are notable by their absence; the biological concept of natural selection plays the role of a suggestive metaphor, and nothing more.

It is important to recognize that this threefold taxonomy describes the process of, natural selection, not the product that process may yield. For example, once a type I process of natural selection has run its course, it is an open question whether the variation that remains is genetic or nongenetic. Consider the work in sociobiology by Richard Alexander.[10] He believes that human beings behave so as to maximize their inclusive fitness. This means that there is an evolutionary explanation for the fact that people in one culture behave differently from those in another. But Alexander does not think that this is due to there being genetic differences between the two cultures. Rather, his idea is that the human genome has evolved so that a person will select the fittest behavior, given the environment he or she occupies. The fact that people behave differently is due to the fact that they occupy different environments. So, in terms of the current variation that we observe, Alexander is, in fact, a radical environmentalist. This is worth contemplating if you think that sociobiology stands or falls with the thesis of genetic determinism.

Matters change when we consider not the present situation, but the evolutionary past that generated it. The genome that Alexander postulates, which gives current humans their ability to modify behavior in the light of ecological conditions, evolved because it was fitter than the alternatives against which it competed. That is, the process of natural selection that led to the present configuration is one in which genetic differences account for differences in behavior.

So Alexander sees genetic differences as being crucial to the process of evolution, but environmental differences as characterizing the product of that evolution. He is a type I theorist, since these types pertain to the process of natural selection, not its product.

The distinction between process and product is perhaps a bit harder to grasp when we think of the evolution of some behavioral or psychological trait, but it really applies to any evolutionary event. For the fact of the matter is that evolution driven by a type I selection process feeds on (additive) genetic variation, and uses it up. A morphological character can display the same double aspect. The opposable thumb evolved because there was a genetic difference between those with the thumb and those without it. But once that trait has finished evolving, the difference between those with and those without a thumb may owe more to industrial accidents and harmful drugs taken prenatally than to genetic oddities.

This threefold division among selection models is of course consistent with there being models that combine two or more of these sorts of process. My taxonomy describes "pure types," so to speak, whereas it is often interesting to consider models in which various pure types are mixed. This is frequently the case in the examples worked out by Cavalli-Sforza and Feldman and by Boyd and Richerson. I want to describe one example from each of these books. The point is to discern the way in which quite different selection processes interact.

In the nineteenth century, Western societies exhibited an interesting demographic change, one that had three stages. First, oscillations in death rates due to epidemics and famines became both less frequent and less extreme. Second, overall mortality steadily declined. This latter change had a multiplicity of causes; improved nutrition, sanitation, and (if the more recent past is also considered) medical advances played a role. The third part of this demographic transition was a dramatic decline in birth rates. Typically, there was a time lag; birth rates began to decline only after death rates were already on the way down. Cavalli-Sforza and Feldman (p. 181) give the somewhat idealized rendition of this pattern shown in figure 25.1.

Cavalli-Sforza and Feldman consider the question of how fertility could have declined in Europe. From the point of view of a narrowly Darwinian outlook, this change is puz-

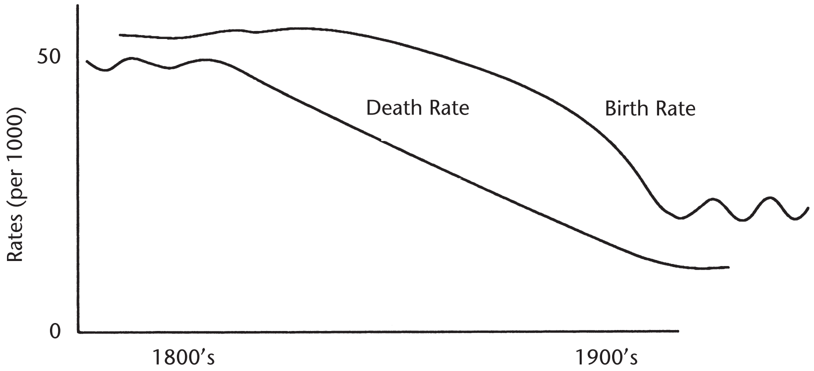


Figure 25.1

Cavalli-Sforza and Feldman's (p. 181) idealized representation of the demographic transition in Europe. Mortality rates decline; then, after a time lag, the birth rate declines also. (Reprinted by permission of Princeton University Press.)

zling. A characteristic that increases the number of viable and fertile offspring will spread under natural selection, at least when that process is conceptualized from the point of view of a type I model. Cavalli-Sforza and Feldman are not tempted to appeal to the theory of optimal clutch size due to Lack, according to which a parent can sometimes augment the number of offspring surviving to adulthood by having fewer babies.[11] Presumably, this Darwinian option is not even worth exploring, because women in nineteenth-century Europe easily could have had more viable fertile offspring than they in fact did. People were not caught in the bind that Lack attributed to his birds.

The trait that increased in the modern demographic transition was one of reduced biological fitness. The trait spread *in spite of* its biological fitness, not *because of* it. In Italy, women changed from having about five children on average to having about two. The trait of having two children, therefore, has a biological fitness of 2/5, when compared with the older trait it displaced.

Cavalli-Sforza and Feldman focus on the problem of explaining how the new custom spread. One possible explanation is that women in all social strata gradually and simultaneously reduced their fertilities. A second possibility is that two dramatically different traits were in play and that the displacement of one by another cascaded from one social class down to the next. The first hypothesis, which posits a gradual spread of innovation, says that fertilities declined from 5 to 4.8 to 4.5 and so on, with this process occurring simultaneously across all classes. The second hypothesis says that having five children competed with having two, and that the novel character was well on its way to displacing the more traditional one among educated people before the same process began among less educated people. This second hypothesis is illustrated in figure 25.2. There is some statistical evidence that the second pattern is more correct, at least in some parts of Europe.

Cavalli-Sforza and Feldman emphasize that this demographic change could not have taken place if traits were passed down solely from mothers to daughters. The Darwinian disadvantage of reduced fertility is so great that purely vertical transmission is not enough to offset it. This point holds true whether fertility is genetically transmitted or learned. A woman with the new trait will pass it along to fewer offspring than a woman with the old pattern, if a daughter is influenced only by her mother.

What is required for the process is some mixture of horizontal and oblique transmission. That is, a woman's reproductive behavior must be influenced by her peers and by her mother's contemporaries. However, it will not do for a woman to adopt the behavior that she finds represented on average in the group that influences her. What is required is that a woman find small family size more attractive than large family size even when very few of her peers possess the novel characteristic. There must be a "transmission bias" in favor of the new trait.

Having a small family was more attractive than having a large one, even though the former trait had a lower Darwinian fitness than the latter. Cavalli-Sforza and Feldman

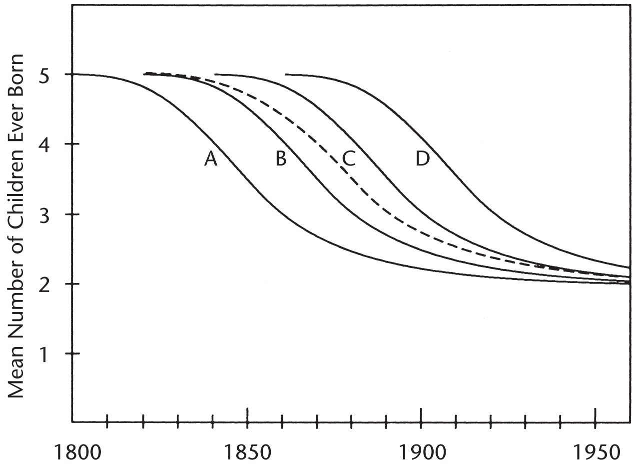


Figure 25.2

Cavalli-Sforza and Feldman's (p. 185) idealized picture of the demographic transition in Italy. A is the most educated class; B, C, and D are progressively less educated. (Reprinted by permission of Princeton University Press.)

show how the greater attractiveness of small family size can be modeled by using ideas drawn from population genetics. However, when these genetic ideas are transposed into a cultural setting, one is talking about cultural fitness, not biological fitness. So the model they end up with for the demographic transition combines two selection processes. When fitness is defined in terms of having babies, having a small family is selected against. When fitness is defined in terms of the attractiveness of an idea ("having students"), there is selection favoring a reduction in family size. Cavalli-Sforza and Feldman show how the cultural process can overwhelm the biological one; given that the trait is sufficiently attractive (and their models have the virtue of giving this idea quantitative meaning), the trait can evolve in spite of its Darwinian disutility.

The example I want to describe from Boyd and Richerson's book is developed in a chapter that begins with a discussion of Japanese kamikaze pilots during World War II. Self-sacrificial behavior—altruism—has been an important problem for recent evolutionary theory. Indeed, Wilson called it "the central problem of sociobiology."[12] Although some apparently altruistic behaviors can be unmasked—shown to be predicated on the selfish expectation of reciprocity, for example—Boyd and Richerson are not inclined to say this about the kamikazes. They died for their country. Nor can one explain their self-sacrifice by saying that it was coerced by leaders; kamikaze pilots volunteered. Nor is it arguable that the pilots volunteered in ignorance of the consequences; suicide missions were common knowledge in the Japanese air force.

So why did kamikaze pilots volunteer? Boyd and Richerson (pp. 204–5) refer to one historian who "argues that the complex of beliefs that gave rise to the kamikaze tactic can be traced back to the Samurai military code of feudal Japan which called for heroic self-sacrifice and put death before dishonour. When the Japanese military modernized in the nineteenth century, the officer corps was drawn from the Samurai class. These men brought their values and transmitted them to subsequent generations of officers who in turn inculcated these values in their men."

Boyd and Richerson (pp. 204–5) say that this historical explanation is "unsatisfactory for two reasons. First, it is incomplete. It tells us why a particular generation of Japanese came to believe in heroic self-sacrifice for the common good; it does not tell us how these beliefs came to predominate in the warrior class of feudal Japan. Second, it is not general enough. The beliefs that led the kamikazes to die for their country are just an especially stark example of a much more general tendency of humans to behave altruistically toward members of various groups of which they are members." They then impose two conditions of adequacy on any proposed explanation: (1) it must show how the "tendency to acquire self-sacrificial beliefs and values could have evolved"; (2) it must show "why altruistic cooperation is directed toward some individuals and not others" (p. 205).

In answer to these requirements, Boyd and Richerson then construct a group selection model that incorporates a certain form of learning. Altruists and selfish individuals exist in each of several groups. Within each group, altruists do less well than selfish people. However, groups of altruists go extinct less often and found more colonies than groups of selfish individuals. These ideas are standard fare in the models of group selection that evolutionary biologists have considered.[13] A type I selection model of the evolution of altruism will require a between-group process favoring altruism that offsets the within-group process that acts to eliminate the trait.

The new wrinkle introduced by the idea of cultural transmission is as follows. Boyd and Richerson postulate that cultural transmission favors common characteristics and works against rare ones. Within a group, individuals are especially biased toward adopting altruism if most individuals are altruists and toward becoming selfish if most people are selfish. What I mean by "especially" biased is illustrated in figure 25.3. In all cases of cultural transmission, the state that a naive individual acquires is influenced by the frequency of traits in the population. Boyd and Richerson impose a more extreme demand. They require that the probability of acquiring a common trait be higher than its population frequency; this is what they call "frequency-dependent biased transmission" (depicted in figure 25.3c).

The process of cultural transmission can work within the time frame of a single biological generation. The effect is to augment the amount of variation there is among groups. Whereas traditional genetic models of group selection allow for a continuum of local frequencies of altruism, the result of this biased transmission rule is to push

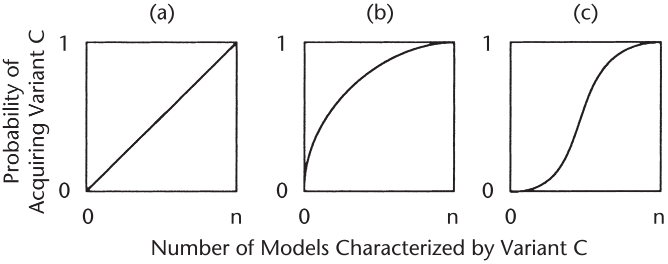


Figure 25.3

Boyd and Richerson's (p. 207) characterization of three patterns of cultural transmission. In all cases, the probability that a naïve individual will acquire a trait depends on the frequency of the trait among the individual's models. (s) represents unbiased transmission, (b) directly biased transmission, and (c) frequency-dependent biased transmission. (Reprinted by permission of University of Chicago Press.)

each local population toward 100 percent altruism or 100 percent selfishness. This has the effect of raising the probability that altruism will evolve and be maintained.

Boyd and Richerson also raise the question of how this biased "conformist" transmission rule could have evolved in the first place. They speculate that if a species is composed of a set of local populations, and if these populations inhabit qualitatively different micro habitats, an individual moving into a new habitat may do best by imitating the traits that are common there. Their proposal is a Darwinian explanation for acting Roman in Rome, so to speak. Once this transmission bias has evolved, it may have various spin-off consequences that have the effect of harming organisms rather than helping them. If you find yourself living with altruists, the transmission bias will lead you to become altruistic yourself, even though you would be better off remaining selfish. Boyd and Richerson admit that there is little or no psychological evidence that people deploy the extreme form of transmission bias that their model postulates.

Just as in the example discussed from Cavalli-Sforza and Feldman, this model of Boyd and Richerson's mixes together the concepts of biological and cultural fitness. Altruism is deleterious to individuals, when fitness is calibrated in terms of the survival and reproduction of organisms. But common characteristics are more contagious than rare ones, when the individuals use a conformist transmission rule. This means that when altruism is common, it is more catching than selfishness. In such cases, the cultural fitness of altruism is greater than the cultural fitness of selfishness, when one considers a group in which altruism is common. The net result is that the special cultural transmission rule can allow a characteristic to evolve that could not evolve without it. Within purely biological models, altruism is eliminated in a large range of parameter values. The prospects for altruism to evolve are enhanced when culture is included in the model. Just as in Cavalli-Sforza and Feldman's discussion of the demographic transition, assumptions about cultural transmission lead to predictions that would not be true if a purely biological and noncultural process were postulated.

The two examples I have described are typical of the models discussed in the two books. The models aim to show how different patterns of cultural transmission make a difference for how a psychological or social characteristic will evolve. Although most of the emphasis is placed on identifying cultural analogs of natural selection, the authors do develop nonselective models of cultural change. For example, population geneticists have described how genes of nearly identical fitness can change frequency in a population by doing a random walk. The models developed for random genetic drift, as it is called, can be used to describe the process by which family names disappear. This helps explain why the descendants of the *Bounty* mutineers have come to share the same surname. A reduction in variation is the expected consequence of random walks, both genetic and cultural.[14]

What are we to make of the research program embodied in these books? Biologists interested in culture are often struck by the absence of viable general theories in the social sciences. All of biology is united by the theory of biological evolution. Perhaps progress in the social sciences is impeded because there is no general theory of cultural evolution. The analogies between cultural and genetic change are palpable. And at least some of the disanalogies can be taken into account when the biological models are transposed. For example, the Weismann doctrine tells us that variation is ''undirected''; mutations do not occur because they would be beneficial. But ideas are not invented at random. Individuals often create new ideas—in science, for example—precisely because they would be useful.[15] Another and related disanalogy concerns the genotype/phenotype distinction. An organism's genotype is a cause of the phenotype it develops; that same genotype also causally contributes to the genotype of the organism's offspring. But there is no further pathway by which a parental phenotype can causally shape the genotype of its offspring. This is one way of describing the idea that there is no ''inheritance of acquired characteristics.'' No such constraints seems to apply to the learning that occurs in cultural transmission.

These disanalogies between genetic and cultural change do not show that it is pointless or impossible to write models of cultural evolution that draw on the mathematical resources provided by evolutionary theory. In a sense, it is precisely because of such differences that there is a point to seeing the consequences of models that take these differences into account. These structural differences between genetic and cultural evolution do not undermine the idea that models of cultural evolution have a point.

Another reservation that has been voiced about models of cultural evolution is that they atomize cultural characteristics. Having two children rather than five, or being a kamikaze pilot, are characteristics that are abstracted from a rich and interconnected network of traits. The worry is that by singling out these traits for treatment, we are losing sight of the context that gives them cultural meaning.

It is worth mentioning that precisely the same question has been raised about various models in genetic evolution itself. If you wish to understand the population frequency of sickle cell anemia, for example, you cannot ignore the fact that the trait is correlated with resistance to malaria. In both cultural and genetic evolution it is a mistake to think that each trait evolved independently of all the others. Of course, the lesson to be drawn from this is not that one should not atomize characteristics, but rather that the atoms one identifies should be understood in terms of their relationship to other atoms.

In fact, this emphasis on context is one of the virtues that Boyd and Richerson think their approach has over the approach taken by sociobiology. According to the models under review, genetic selection has given our species the ability to engage in social learning. Once in place, this cultural transmission system allows characteristics to evolve that could not have evolved without it. In other words, it is only because the traits in question evolve in the context of a cultural transmission system that they are able to evolve at all.

We need to recognize that the descriptors singled out for treatment in science always abstract from complexities. If there is an objection to the descriptors used in models of cultural evolution, it must concern the details of how these models are constructed, not the mere fact that they impose a descriptive framework of some sort or other.[16]

Although the criticisms I have reviewed so far do not seem very powerful, there is a rather simple fact about these models that does suggest that they may be of limited utility in the social sciences. Insofar as these models describe culture, they describe systems of cultural transmission and the evolutionary consequences of such systems. Given that the idea of having two children was more attractive than the idea of having five, and given the horizontal and oblique transmission systems thought to be in place, we can see why the demographic transition took place. But as Cavalli-Sforza and Feldman recognize, their model does not begin to describe why educated women in nineteenth-century Italy came to prefer having smaller families, or why patterns adopted in higher classes cascaded down to lower ones. The model describes the consequences of an idea's being attractive, not the causes of its being attractive.

This distinction between the consequences of fitness differences and the causes of fitness differences also applies to theories of biological evolution.[17] A population geneticist can tell you what the evolutionary consequences for a population will be, if the genes in the population bear various fitness relationships to each other. It is a separate question to say why a given gene in fact is fitter than the alternatives. For example, consider the simplest of one-locus two-allele models for a diploid population.

There are three genotypes possible at the locus in question, which we might label AA, Aa, and aa. If the heterozygote genotype is fitter than the two homozygote forms, the population will evolve to a stable polymorphism. Neither allele will be eliminated by the selection process. This is a simple algebraic fact, one having nothing to do with the biological details of any living population. Models such as this one can be thought of as intellectual resources that biologists interested in some particular population might find reason to use.

When human geneticists apply this model to the sickle cell system, they say that Aa is the fittest genotype because heterozygotes at the locus in question have enhanced resistance to malaria and little or no anemia. The two homozygotes have lesser fitnesses because they are either anemic or lack the malaria resistance. These specific remarks about the locus in the relevant human population describe the sources of fitness differences. Alternatively, a fruitfly geneticist may take the same population genetics model and apply it to a locus in some Drosophila population by saying that the heterozygote has enhanced temperature tolerance over the two homozygotes. The population consequences of heterozygote superiority are the same in the two cases; a stable polymorphism evolves. It is the sources of the fitness differences that distinguish the human application from the application to fruitflies.

This, I think, is the main shortcoming of the models of cultural evolution I am considering. The illumination they offer of culture concerns the consequences of cultural transmission systems. But there is far more to culture than the consequences of the rules that describe who learns what from whom. Social scientists have not wholly ignored the way that patterns of influence are structured in specific cases. A historian of nineteenth-century Italy might attempt to explain why some traits found among educated people were transmitted to lower social strata, while others were not. Again, it is the sources of the transmission system that will interest the social scientist. The social scientist will take it for granted that the consequences of this influence will be that ideas cascade from one class to another.

Models of transmission systems describe the quantitative consequences of systems of cultural influence. Social scientists inevitably make qualitative assumptions about the consequences of these systems. If it could be shown that these qualitative assumptions were wrong in important cases, and that these mistakes actually undermine the plausibility of various historical explanations, that would be a reason or social scientists to take a greater interest in these models of cultural evolution. But if the qualitative assumptions turn out to be correct, it is perhaps understandable that historians should not accord much importance to these investigations.[18]

Population genetics really is a unifying framework within evolutionary theory. Fruitflies and human beings differ in many ways, but if a one-locus system exhibits heterozygote superiority, the population consequences will be the same, regardless of whether we are talking about people or Drosophila. Evolutionary theory is much less unified when we consider what it has to say about the sources of fitness differences. There are many, many models that treat a multiplicity of life-history characteristics and ecological relationships. Evolutionary theory achieves its greatest generality when it ignores sources and focuses on consequences.

The transposition of evolutionary models to the social sciences is a transposition of the most unified and complete part of evolutionary theory, one that leaves behind less unified theoretical ideas. This is not a criticism of the models of cultural evolution that result, but a fact about the price one pays for very general theorizing of this type. Cultural learning is a cultural universal. And patterns of cultural learning conveniently divide into vertical, horizontal, and oblique subcases. When ideas differ in their attractiveness, the system of transmission will determine the rate of change and the end-state that the population achieves. Only because they develop theories *within* this narrow compass do these models *of* cultural evolution have the generality they do.

Many of the examples discussed in the two books I have been considering describe evolution within a culture, not the evolution of the cultural transmission system itself. However, Boyd and Richerson, especially, also concern themselves with the way a system of cultural learning could have evolved by straightforward Darwinian means. Here the authors are not giving a model of how human cultures work, once they exist, but are trying to show how cultural learning became a possibility in the first place. This project obviously is a very important one, but not one that applies to many social scientific research programs. A correct genetic explanation of this important feature of the human phenotype would not provide a unifying framework within which social scientists would then do their work. They would not use this theory at all. It is one thing to explain the demographic transition in nineteenth-century Italy, something else to explain why human beings are able to learn from individuals who are not their biological parents.

In spite of these shortcomings, there is a basic achievement of these models of cultural evolution that deserves emphasis. A persistent theme in debates about sociobiology, about the nature/nurture controversy, and in other contexts as well is the relative "importance" that should be accorded to biology and culture. I place the term "importance" in quotation marks to indicate that it is a vague idea crying out for explication. Nonetheless, it has been a fundamental problem in these controversies to assess the relative "strength" or "power" of biological and cultural influences.

One virtue of these models of cultural evolution is that they place culture and biology into a common framework, so that the relative contributions to an outcome are rendered commensurable. What becomes clear in these models is that in assessing their relative importance of biology and culture, *time is of the essence*. Culture is often a more powerful determiner of change than biological evolution because cultural changes occur faster. When biological fitness is calibrated in terms of having babies, its basic temporal unit is the span of a human generation. Think how many replication events can occur in that temporal interval when the reproducing entities are ideas that jump from head to head. Ideas spread so fast that they can swamp the slower (and hence weaker) impact of biological natural selection.

There is a vague idea about the relation of biology and culture that these models help lay to rest. This is the idea that biology is ''deeper'' than the social sciences, not just in the sense that it has developed further, but in the sense that it investigates more fundamental causes. A social scientist will explain incest avoidance by describing the spread of a custom; the evolutionary biologist goes deeper by showing us why the behavior evolved. The mind-set expressed here is predisposed to think that culture is always a weak influence when it opposes biology. The works described here deserve credit for showing why this common opinion rests on a confusion.

In spite of this achievement, I doubt that these models of cultural evolution provide a general framework within which social scientific investigations may proceed. My main reason for skepticism is that these models concern themselves with the consequences of transmission systems and fitness differences, not with their sources. Social scientists interested in cultural change generally focus on sources and make do with intuitive and qualitative assessments of what the consequences will be. It isn't that the biologists and the social scientists are in conflict; rather, they are talking past each other.

Dobzhansky is famous for having said that ''nothing in biology can be understood except in the light of evolution.'' His idea was not the modest one that evolution is necessary for full understanding; that would be true even if evolution's contribution were minor, though ineliminable. Rather, Dobzhansky had in mind the stronger claim that evolutionary considerations should be assigned pride of place in our understanding of the living world. A transposition of Dobzhansky's slogan to the topic of this chapter would say that ''nothing in the social sciences can be understood except in the light of models of cultural evolution.'' My suspicion is that only the weaker reading of this pronouncement is defensible.

Notes

I worked on this chapter while a William Evans Fellow at the University of Otago during parts of July and August 1990; my thanks to the university and to the members of the Philosophy Department for inviting me and for making my stay such an enjoyable one. This work expands upon a talk I gave in December 1985 at the University of Palma de Mallorca, ''Natural Selection and the Social Sciences.'' I'm grateful to Robert Boyd, Dan Hausman, Peter Richerson, and David S. Wilson for comments on an earlier draft.

1. John Maynard Smith (1982), *Evolution and the Theory of Games*, Cambridge University Press.

2. L. Cavalli-Sforza and M. Feldman (1981), *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton University Press.

3. R. Boyd and P. Richerson (1985), *Culture and the Evolutionary Process*, University of Chicago Press.

4. ''Having babies'' should be interpreted broadly, so as to include ''having grandbabies,'' ''having greatgrandbabies,'' and so forth. In some selection models (e.g., Fisher's sex ratio argument), fitness differences require that one consider expected numbers of descendants beyond the first generation.

5. See R. Colwell and M. King (1983), ''Disentangling Genetic and Cultural Influences on Human Behavior: Problems and Prospects,'' in D. Rajecki (ed.), *Comparing Behavior: Studying Man Studying Animals*, Lawrence Erlbaum Publishers.

6. These are reviewed in J. Hirshliefer (1977), ''Economics from a Biological Viewpoint,'' *Journal of Law and Economics* 1: 1–52.

7. See K. Popper (1973), *Objective Knowledge*, Oxford University Press.

8. A variety of ''selective-retention'' models of learning and of scientific change are reviewed in Donald Campbell (1974), ''Evolutionary Epistemology,'' in P. Schilpp (ed.), *The Philosophy of Karl Popper*, Open Court Publishing. David Hull's *Science as a Process* (University of Chicago Press, 1988) develops some interesting ideas about how evolutionary ideas can be used to explain scientific change.

9. I do not claim that this taxonomy is exhaustive. For example, the spread of an infectious disease may be thought of as a selection process, in which the two states of an individual (''infected'' and ''not infected'') differ in how catching they are. Clearly, this is not a type I process. Arguably, the concept of learning does not permit this process to be placed in type II. Perhaps the taxonomy would be exhaustive, if ''learning'' were replaced by ''phenotypic resemblance not mediated by genetic resemblance.''

10. See, for example, Richard Alexander (1979), *Darwinism and Human Affairs*, University of Washington Press.

11. See D. Lack (1954), *The Optimal Regulation of Animal Numbers*, Oxford University Press.

12. E. Wilson (1975), *Sociobiology: The New Synthesis*, Harvard University Press.

13. See E. Sober (1988), ''What is Evolutionary Altruism?'' New Essays on Philosophy and Biology (*Canadian Journal of Philosophy Supplementary* Volume 14), University of Calgary Press.

14. See Cavalli-Sforza and Feldman, pp. 255–66.

15. The difference between directed and undirected variation is conceptually different from the difference between biased and unbiased transmission. The former concerns the probability that a mutation will arise; the latter has to do with whether it will be passed along.

Directed variation (mutation) can be described as follows. Let u be the probability of mutating from A to a and v be the probability of mutating from a to A. Mutation is directed if (i) $u > v$ and (ii) $u > v$ because $w(a) > w(A)$, where $w(X)$ is the fitness of X.

16. See J. M. Smith (1989), *Did Darwin Get It Right*? Chapman and Hall.

17. See Elliott Sober (1984), *The Nature of Selection*, MIT Press.

18. So the question about the usefulness of these models of cultural evolution to the day-to-day research of social scientists comes to this: Are social scientists good at intuitive population thinking? If they are, then their explanations will not be undermined by precise models of cultural evolution. If they are not, then social scientists should correct their explanations (and the intuitions on which they rely) by studying these models.

XIII Evolutionary Ethics

26 Moral Philosophy as Applied Science

Michael Ruse and Edward O. Wilson

For much of this century, moral philosophy has been constrained by the supposed absolute gap between *is* and *ought*, and the consequent belief that the facts of life cannot of themselves yield an ethical blueprint for future action. For this reason, ethics has sustained an eerie existence largely apart from science. Its most respected interpreters still believe that reasoning about right and wrong can be successful without a knowledge of the brain, the human organ where all the decisions about right and wrong are made. Ethical premises are typically treated in the manner of mathematical propositions: directives supposedly independent of human evolution, with a claim to ideal, eternal truth.

While many substantial gains have been made in our understanding of the nature of moral thought and action, insufficient use has been made of knowledge of the brain and its evolution. Beliefs in extrasomatic moral truths and in an absolute is/ought barrier are wrong. Moral premises relate only to our physical nature and are the result of an idiosyncratic genetic history—a history which is nevertheless powerful and general enough within the human species to form working codes. The time has come to turn moral philosophy into an applied science because, as the geneticist Hermann J. Muller urged in 1959, 100 years without Darwin are enough.[1]

The naturalistic approach to ethics, dating back through Darwin to earlier preevolutionary thinkers, has gained strength with each new advance in biology and the brain sciences. Its contemporary version can be expressed as follows:

Everything human, including the mind and culture, has a material base and originated during the evolution of the human genetic constitution and its interaction with the environment. To say this much is not to deny the great creative power of culture, or to minimize the fact that most causes of human thought and behavior are still poorly understood. The important point is that modern biology can account for many

From *Philosophy* 61 (1986). Reprinted by permission of Cambridge University Press. © Royal Institute of Philosophy.

of the unique properties of the species. Research on the subject is accelerating, quickly enough to lend plausibility to the belief that the human condition can eventually be understood to its foundations, including the sources of moral reasoning.

This accumulating empirical knowledge has profound consequences for moral philosophy. It renders increasingly less tenable the hypothesis that ethical truths are extrasomatic, in other words divinely placed within the brain or else outside the brain awaiting revelation. Of equal importance, there is no evidence to support the view—and a great deal to contravene it—that premises can be identified as global optima favoring the survival of any civilized species, in whatever form or on whatever planet it might appear. Hence external goals are unlikely to be articulated in this more pragmatic sense.

Yet biology shows that internal moral premises do exist and can be defined more precisely. They are immanent in the unique programs of the brain that originated during evolution. Human mental development has proved to be far richer and more structured and idiosyncratic than previously suspected. The constraints on this development are the sources of our strongest feelings of right and wrong, and they are powerful enough to serve as a foundation for ethical codes. But the articulation of enduring codes will depend upon a more detailed knowledge of the mind and human evolution than we now possess. We suggest that it will prove possible to proceed from a knowledge of the material basis of moral feeling to generally accepted rules of conduct. To do so will be to escape—not a minute too soon—from the debilitating absolute distinction between *is* and *ought*.

All populations of organisms evolve through a law-bound causal process, as first described by Charles Darwin in his *Origin of Species*. The modern explanation of this process, known as natural selection, can be briefly summarized as follows. The members of each population vary hereditarily in virtually all traits of anatomy, physiology, and behavior. Individuals possessing certain combinations of traits survive and reproduce better than those with other combinations. As a consequence, the units that specify physical traits—genes and chromosomes—increase in relative frequency within such populations, from one generation to the next.

This change in different traits, which occurs at the level of the entire population, is the essential process of evolution. Although the agents of natural selection act directly on the outward traits and only rarely on the underlying genes and chromosomes, the shifts they cause in the latter have the most important lasting effects. New variation across each population arises through changes in the chemistry of the genes and their relative positions on the chromosomes. Nevertheless, these changes (broadly referred to as mutations) provide only the raw material of evolution. Natural selection, composed of the sum of differential survival and reproduction, for the most part determines the rate and direction of evolution.[2]

Although natural selection implies competition in an abstract sense between different forms of genes occupying the same chromosome positions or between different gene arrangements, pure competition, sometimes caricatured as ''nature red in tooth and claw,'' is but one of several means by which natural selection can operate on the outer traits. In fact, a few species are known whose members do not compete among themselves at all. Depending on circumstances, survival and reproduction can be promoted equally well through the avoidance of predators, more efficient breeding, and improved cooperation with others.[3]

In recent years there have been several much-publicized controversies over the pace of evolution and the universal occurrence of adaptation.[4] These uncertainties should not obscure the key facts about organic evolution: that it occurs as a universal process among all kinds of organisms thus far carefully examined, that the dominant driving force is natural selection, and that the observed major patterns of change are consistent with the known principles of molecular biology and genetics. Such is the view held by the vast majority of the biologists who actually work on heredity and evolution.[5] To say that not all the facts have been explained, to point out that forces and patterns may yet be found that are inconsistent with the central theory—healthy doubts present in any scientific discipline—is by no means to call into question the prevailing explanation of evolution. Only a demonstration of fundamental inconsistency can accomplish that much, and nothing short of a rival explanation can bring the existing theory into full disarray.

There are no such crises. Even Motoo Kimura, the principal architect of the ''neutralist'' theory of genetic diversity—which proposes that most evolution at the molecular level happens through random factors—allows that ''classical evolution theory has demonstrated beyond any doubt that the basic mechanism for adaptive evolution is natural selection acting on variations produced by changes in chromosomes and genes. Such considerations as population size and structure, availability of ecological opportunities, change of environment, life-cycle 'strategies,' interaction with other species, and in some situations kin or possibly group selection play a large role in our understanding of the process.''[6]

Human evolution appears to conform entirely to the modern synthesis of evolutionary theory as just stated. We know now that human ancestors broke from a common line with the great apes as recently as six or seven million years ago, and that at the biochemical level we are today closer relatives of the chimpanzees than the chimpanzees are of gorillas.[7] Furthermore, all that we know about human fossil history, as well as variation in genes and chromosomes among individuals and the key events in the embryonic assembly of the nervous system, is consistent with the prevailing view that natural selection has served as the principal agent in the origin of humanity.

It is true that until recently information on the brain and human evolution was sparse. But knowledge is accelerating, at least as swiftly as the remainder of natural science, about a doubling every ten to fifteen years. Several key developments, made principally during the past twenty years, will prove important to our overall argument for a naturalistic ethic developed as an applied science.

The number of human genes identified by biochemical assay or pedigree analysis is at the time of writing 3,577, with approximately 600 placed to one or the other of the twenty-three pairs of chromosomes.[8] Because the rate at which this number has been accelerating (up from 1,200 in 1977), most of the entire complement of 100,000 or so structural genes may be characterized to some degree within three or four decades.

Hundreds of the known genes affect behavior. The great majority do so simply by their effect on general processes of tissue development and metabolism, but a few have been implicated in more focused behavioral traits. For example, a single allele (a variant of one gene), prescribes the rare Lesch–Nyhan syndrome, in which people curse uncontrollably, strike out at others with no provocation, and tear at their own lips and fingers. Another allele at a different chromosome position reduces the ability to perform on certain standard spatial tests but not on the majority of such tests.[9] Still another allele, located tentatively on chromosome 15, induces a specific learning disability.[10]

These various alterations are of course strong and deviant enough to be considered pathological. But they are also precisely the kind usually discovered in the early stages of behavioral genetic analysis for any species. *Drosophila* genetics, for example, first passed through a wave of anatomical and physiological studies directed principally at chromosome structure and mechanics. As in present-day human genetics, the first behavioral mutants discovered were broadly acting and conspicuous, in other words, those easiest to detect and characterize. When behavioral and biochemical studies grew more sophisticated, the cellular basis of gene action was elucidated in the case of a few behaviors, and the new field of *Drosophila* neurogenetics was born. The hereditary bases of subtle behaviors such as orientation to light and learning were discovered somewhat later.[11]

We can expect human behavioral genetics to travel along approximately the same course. Although the links between genes and behavior in human beings are more numerous and the processes involving cognition and decision making far more complex, the whole is nevertheless conducted by cellular machinery precisely assembled under the direction of the human genome (that is, genes considered collectively as a unit). The techniques of gene identification, applied point by point along each of the twenty-three pairs of chromosomes, is beginning to make genetic dissection of human behavior a reality.

Yet to speak of genetic dissection, a strongly reductionist procedure, is not to suggest that the whole of any trait is under the control of a single gene, nor does it deny substantial flexibility in the final product. Individual alleles (gene-variants) can of course affect a trait in striking ways. To take a humble example, the possession of a single allele rather than another on a certain point on one of the chromosome pairs causes the development of an attached earlobe as opposed to a pendulous earlobe. However, it is equally true that a great many alleles at different chromosome positions must work together to assemble the entire earlobe. In parallel fashion, one allele can shift the likelihood that one form of behavior will develop as opposed to another, but many alleles are required to prescribe the ensemble of nerve cells, neurotransmitters, and muscle fibers that orchestrate the behavior in the first place. Hence classical genetic analysis cannot by itself explain all of the underpinnings of human behavior, especially those that involve complex forms of cognition and decision making. For this reason behavioral development viewed as the interaction of genes and environment should also occupy center stage in the discussion of human behavior. The most important advances at this level are being made in the still relatively young field of cognitive psychology.[12]

With this background, let us move at once to the central focus of our discussion: morality. Human beings, all human beings, have a sense of right and wrong, good and bad. Often, although not always, this "moral awareness" is bound up with beliefs about deities, spirits, and other supersensible beings. What is distinctive about moral claims is that they are prescriptive; they lay upon us certain obligations to help and to cooperate with others in various ways. Furthermore, morality is taken to transcend mere personal wishes or desires. "Killing is wrong" conveys more than merely "I don't like killing." For this reason, moral statements are thought to have an objective referent, whether the Will of a Supreme Being or eternal verities perceptible through intuition.

Darwinian biology is often taken as the antithesis of true morality. Something that begins with conflict and ends with personal reproduction seems to have little to do with right and wrong. But to reason along such lines is to ignore a great deal of the content of modern evolutionary biology. A number of causal mechanisms—already well confirmed in the animal world—can yield the kind of cooperation associated with moral behavior. One is so-called kin selection. Genes prescribing cooperation spread through the populations when self-sacrificing acts are directed at relatives, so that they (not the cooperators) are benefited, and the genes they share with the cooperators by common descent are increased in later generations. Another such cooperation-causing mechanism is "reciprocal altruism." As its name implies, this involves transactions (which can occur between nonrelatives) in which aid given is offset by the expectation of aid received. Such mutual assistance can be extended to a whole group, whose individual members contribute to a general pool and (as needed) draw from the pool.[13]

Sociobiologists (evolutionists concerned with social behavior) speak of acts mediated by such mechanisms as "altruistic." It must be recognized that this is now a technical biological term, and does not necessarily imply conscious free giving and receiving. Nevertheless, the empirical evidence suggests that cooperation between human beings was brought about by the same evolutionary mechanisms as those just cited. To include conscious, reflective beings is to go beyond the biological sense of altruism into the realm of genuine nonmetaphorical altruism. We do not claim that people are either unthinking genetic robots or that they cooperate only when the expected genetic returns can be calculated in advance. Rather, human beings function better if they are deceived by their genes into thinking that there is a disinterested objective morality binding upon them, which all should obey. We help others because it is "right" to help them and because we know that they are inwardly compelled to reciprocate in equal measure. What Darwinian evolutionary theory shows is that this sense of "right" and the corresponding sense of "wrong," feelings we take to be above individual desire and in some fashion outside biology, are in fact brought about by ultimately biological processes.

Such are the empirical claims. How exactly is biology supposed to exert its will on conscious, free beings? At one extreme, it is possible to conceive of a moral code produced entirely by the accidents of history. Cognition and moral sensitivity might evolve somewhere in some imaginary species in a wholly unbiased manner, creating the organic equivalent of an all-purpose computer. In such a blank-slate species, moral rules were contrived some time in the past, and the exact historical origin might now be lost in the mists of time. If proto-humans evolved in this manner, individuals that thought up and followed rules ensuring an ideal level of cooperation then survived and reproduced, and all others fell by the wayside.

However, before we consider the evidence, it is important to realize that any such even-handed device must also be completely gene based and tightly controlled, because an exact genetic prescription is needed to produce perfect openness to any moral rule, whether successful or not. The human thinking organ must be indifferently open to a belief such as "killing is wrong" or "killing is right," as well as to any consequences arising from conformity or deviation. Both a very specialized prescription and an elaborate cellular machinery are needed to achieve this remarkable result. In fact, the blank-slate brain might require a cranial space many times that actually possessed by human beings. Even then a slight deviation in the many feedback loops and hierarchical controls would shift cognition and preference back into a biased state. In short, there appears to be no escape from the biological foundation of mind.

It can be stated with equal confidence that nothing like all-purpose cognition occurred during human evolution. The evidence from both genetic and cognitive studies demonstrates decisively that the human brain is not a *tabula rasa*. Conversely, neither is the brain (and the consequent ability to think) genetically determined in the strict sense. No genotype is known that dictates a single behavior, precluding reflection and the capacity to choose from among alternative behaviors belonging to the same category. The human brain is something in between: a swift and directed learner that picks up certain bits of information quickly and easily, steers around others, and leans toward a surprisingly few choices out of the vast array that can be imagined.

This quality can be made more explicit by saying that human thinking is under the influence of "epigenetic rules," genetically based processes of development that predispose the individual to adopt one or a few forms of behaviors as opposed to others. The rules are rooted in the physiological processes leading from the genes to thought and action.[14] The empirical heart of our discussion is that we think morally because we are subject to appropriate epigenetic rules. These predispose us to think that certain courses of action are right and certain courses of action are wrong. The rules certainly do not lock people blindly into certain behaviors. But because they give the illusion of objectivity to morality, they lift us above immediate wants to actions which (unknown to us) ultimately serve our genetic best interests.

The full sequence in the origin of morality is therefore evidently the following: ensembles of genes have evolved through mutation and selection within a intensely social existence over tens of thousands of years; they prescribe epigenetic rules of mental development peculiar to the human species; under the influence of the rules certain choices are made from among those conceivable and available to the culture; and finally the choices are narrowed and hardened through contractual agreements and sanctification.

In a phrase, societies feel their way across the fields of culture with a rough biological map. Enduring codes are not created whole from absolute premises but inductively, in the manner of common law, with the aid of repeated experience, by emotion and consensus, through an expansion of knowledge and experience guided by the epigenetic rules of mental development, during which people sift the options and come to agree upon and to legitimate certain norms and directions.[15]

Only recently have the epigenetic rules of mental development and their adaptive roles become accepted research topics for evolutionary biology. It should therefore not be surprising that to date the best understood examples of epigenetic rules are of little immediate concern to moral philosophers. Yet what such examples achieve is to draw us from the realm of speculative philosophy into the center of ongoing scientific research. They provide the stepping stones to a more empirical basis of moral reasoning.

One of the most fully explored epigenetic rules concerns the constraint on color vision that affects the cultural evolution of color vocabularies. People see variation in the *intensity* of light (as opposed to color) the way one might intuitively expect to see it. That is, if the level of illumination is raised gradually, from dark to brightly lit, the transition is perceived as gradual. But if the *wavelength* is changed gradually, from a monochromatic purple all across the visible spectrum to a monochromatic red, the shift is not perceived as a continuum. Rather, the full range is thought to comprise four basic colors (blue, green, yellow, red), each persisting across a broad band of wavelengths and giving way through ambiguous intermediate color through narrow bands on either side. The physiological basis of this beautiful deception is partly known. There are three kinds of cones in the retina and four kinds of cells in the lateral geniculate nuclei of the visual pathways leading to the optical cortex. Although probably not wholly responsible, both sets of cells play a role in the coding of wavelength so that it is perceived in a discrete rather than continuous form. Also, some of the genetic basis of the cellular structure is known. Color-blindness alleles on two positions in the X-chromosome cause particular deviations in wavelength perception.

The following experiment demonstrated the effect of this biological constraint on the formation of color vocabularies. The native speakers of twenty languages from around the world were asked to place their color terms in a standard chart that displays the full visible color spectrum across varying shades of brightness. Despite the independent origins of many of the languages, which included Arabic, Ibidio, Thai, and Tzeltal, the terms placed together fall into four distinct clusters corresponding to the basic colors. Very few were located in the ambiguous intermediate zones.

A second experiment then revealed the force of the epigenetic rule governing this cultural convergence. Prior to European contact the Dani people of New Guinea possessed a very small color vocabulary. One group of volunteers was taught a newly invented Dani-like set of color terms placed variously on the four principal hue categories (blue, green, yellow, red). A second group was taught a similar vocabulary placed off-center, away from the main clusters formed by other languages. The first group of volunteers, those given the ''natural'' vocabulary, learned about twice as quickly as those given the off-center, less natural terms. Dani volunteers also selected these terms more readily when allowed to make a choice between the two sets.[16]

So far as we have been able to determine, all categories of cognition and behavior investigated to the present time show developmental biases. More precisely, whenever development has been investigated with reference to choice under conditions as free as possible of purely experimental influence, subjects automatically favored certain choices over others. Some of these epigenetic biases are moderate to very strong, as in the case of color vocabulary. Others are relatively weak. But all are sufficiently marked to exert a detectable influence on cultural evolution.

Examples of such deep biases included the optimum degree of redundancy in geometric design; facial expressions used to denote the basic emotions of fear, loathing, anger, surprise, and happiness; descending degrees of preference for sucrose, fructose, and other sugars; the particular facial expressions used to respond to various distasteful substances; and various fears, including the fear-of-strangers response in children. One of the most instructive cases is provided by the phobias. These intense reactions are most readily acquired against snakes, spiders, high places, running water, tight enclosures, and other ancient perils of mankind for which epigenetic rules can be expected to evolve through natural selection. In contrast, phobias very rarely appear in response to automobiles, guns, electric sockets, and other truly dangerous objects in modern life, for which the human species has not yet had time to adapt through genetic change.

Epigenetic rules have also been demonstrated in more complicated forms of mental development, including language acquisition, predication in logic, and the way in which objects are ordered and counted during the first steps in mathematical reasoning.[17]

We do not wish to exaggerate the current status of this area of cognitive science. The understanding of mental development is still rudimentary in comparison with that of most other aspects of human biology. But enough is known to see the broad outlines of complex processes. Moreover, new techniques are constantly being developed to explore the physical basis of mental activity. For example, arousal can be measured by the degree of alpha wave blockage, allowing comparisons of the impact of different visual designs. Electroencephalograms of an advanced design are used to monitor moment-by-moment activity over the entire surface of the brain. In a wholly different procedure, radioactive isotopes and tomography are combined to locate sites of enhanced metabolic activity. Such probes have revealed the areas of the brain used in specific mental operations, including the recall of melodies, the visualization of notes on a musical staff, and silent reading and counting.[18] There seems to be no theoretical reason why such techniques cannot be improved eventually to address emotions, more complex reasoning, and decision making. There is similarly no reason why metabolic activity of the brain cannot be mapped in chimpanzees and other animals as they solve problems and initiate action, permitting the comparison of mental activity in human beings with that in lower species.

But what of morality? We have spoken of color perception, phobias, and other less value-laden forms of cognition. We argue that moral reasoning is likewise molded and constrained by epigenetic rules. Already biologists and behavioral scientists are moving directly into that area of human experience producing the dictates of right and wrong. Consider the avoidance of brother-sister incest, a negative choice made by the great majority of people around the world. By incest in this case is meant full sexual attraction and intercourse, and not merely exploratory play among children. When such rare matings do occur, lowered genetic fitness is the result. The level of homozygosity (a matching of like genes) in the children is much higher, and they suffer a correspondingly greater mortality and frequency of crippling syndromes due to the fact that some of the homozygous paris of genes are defective. Yet this biological cause and effect is not widely perceived in most societies, especially those with little or no scientific knowledge of heredity. What causes the avoidance instead is a sensitive period between birth and approximately six years. When children this age are exposed to each other under conditions of close proximity (both "use the same potty," as one anthropologist put it) they are unable to form strong sexual bonds during adolescence or later. The inhibition persists even when the pairs are biologically unrelated and encouraged to marry. Such a circumstance occurred, for example, when children from different families were raised together in Israeli kibbutzim and in Chinese households practicing minor marriages.[19]

A widely accepted interpretation of the chain of causation in the case of brother-sister incest avoidance is as follows. Lowered genetic fitness due to inbreeding led to the evolution of the juvenile sensitive period by means of natural selection; the inhibition experienced at sexual maturity led to prohibitions and cautionary myths against incest or (in many societies) merely a shared feeling that the practice is inappropriate. Formal incest taboos are the cultural reinforcement of the automatic inhibition, an example of the way culture is shaped by biology. But these various surface manifestations need not be consulted in order to formulate a more robust technique of moral reasoning. What matters in this case is the juvenile inhibition: the measures of its strength and universality, and a deeper understanding of why it came into being during the genetic evolution of the brain.

Sibling incest is one of several such cases showing that a tight and formal connection can be made between biological evolution and cultural change. Models of sociobiology have now been extended to include the full co-evolutionary circuit, from genes affecting the direction of cultural change to natural selection shifting the frequencies of these genes, and back again to open new channels for cultural evolution. The models also predict the pattern of cultural diversity resulting from a given genotype distributed uniformly through the human species. It has just been seen how the avoidance of brother-sister incest arises from a strong negative bias and a relative indifference to the preferences of others. The quantitative models incorporating these parameters yield a narrow range of cultural diversity, with a single peak at or near complete rejection on the part of the members of most societies. A rapidly declining percentage of societies possess higher rates of acceptance. If the bias is made less in the model than the developmental data indicate, the mode of this frequency curve (that is, the frequency of societies whose members display different percentages of acceptance) shifts from one end of the acceptance scale toward its center. If individuals are considerably more responsive to the preferences of others, the frequency curve breaks into two modes.[20]

Such simulations, employing the principles of population genetics as well as methods derived from statistical mechanics, are still necessarily crude and applicable only to the simplest forms of culture. But like behavioral genetics and the radionuclide-tomography mapping of brain activity, they give a fair idea of the kind of knowledge that is possible with increasing sophistication in theory and technique. The theory of the co-evolution of genes and culture can be used further to understand the origin and meaning of the epigenetic rules, including those that affect moral reasoning.

This completes the empirical case. To summarize, there is solid factual evidence for the existence of epigenetic rules—constraints rooted in our evolutionary biology that affect the way we think. The incest example shows that these rules, directly related to adaptive advantage, extend into the moral sphere. And the hypothesis of morality as a product of pure culture is refuted by the growing evidence of the co-evolution of genes and culture.

This perception of co-evolution is, of course, only a beginning. Prohibitions on intercourse with siblings hardly exhaust the human moral dimension. Philosophical reasoning based upon more empirical information is required to give a full evolutionary account of the phenomena of interest: philosophers' hands reaching down, as it were, to grasp the hands of biologists reaching up. Surely some of the moral premises articulated through ethical inquiry lie close to real epigenetic rules. For instance, the contractarians' emphasis on fairness and justice looks much like the result of rules brought about by reciprocal altruism, as indeed one distinguished supporter of that philosophy has already noted.[21]

We believe that implicit in the scientific interpretation of moral behavior is a conclusion of central importance to philosophy, namely, that there can be no genuinely objective external ethical premises. Everything that we know about the evolutionary process indicates that no such extrasomatic guides exist. Let us define ethics in the ordinary sense, as the area of thought and action governed by a sense of obligation—a feeling that there are certain standards one ought to live up to. In order not to prejudge the issue, let us also make no further assumptions about content. It follows from what we understand in the most general way about organic evolution that ethical premises are likely to differ from one intelligent species to another. The reason is that choices are made on the basis of emotion and reason directed to these ends, and the ethical premises composed of emotion and reason arise from the epigenetic rules of mental development. These rules are in turn the idiosyncratic products of the genetic history of the species and as such were shaped by particular regimes of natural selection. For many generations—more than enough for evolutionary change to occur—they favored the survival of individuals who practiced them. Feelings of happiness, which stem from positive reinforcers of the brain and other elements that compose the epigenetic rules, are the enabling devices that led to such right action.

It is easy to conceive of an alien intelligent species evolving rules its members consider highly moral but which are repugnant to human beings, such as cannibalism, incest, the love of darkness and decay, parricide, and the mutual eating of feces. Many animal species perform some or all of these things, with gusto and in order to survive. If human beings had evolved from a stock other than savanna-dwelling, bipedal, carnivorous man-apes we might do the same, feeling inwardly certain that such behaviors are natural and correct. In short, ethical premises are the peculiar products of genetic history, and they can be understood solely as mechanisms that are adaptive for the species that possess them. It follows that the ethical code of one species cannot be translated into that of another. No abstract moral principles exist outside the particular nature of individual species.

It is thus entirely correct to say that ethical laws can be changed, at the deepest level, by genetic evolution. This is obviously quite inconsistent with the notion of morality as a set of objective, eternal verities. Morality is rooted in contingent human nature, through and through.

Nor is it possible to uphold the true objectivity of morality by believing in the existence of an ultimate code, such that what is considered right corresponds to what is truly right—that the thoughts produced by the epigenetic rules parallel external premises.[22] The evolutionary explanation makes the objective morality redundant, for even if external ethical premises did not exist, we would go on thinking about right and wrong in the way that we do. And surely, redundancy is the last predicate that an objective morality can possess. Furthermore, what reason is there to presume that our present state of evolution puts us in correspondence with ultimate truths? If there are genuine external ethical premises, perhaps cannibalism is obligatory.

Thoughtful people often turn away from naturalistic ethics because of a belief that it takes the goodwill out of cooperation and reduces righteousness to a mechanical process. Biological "altruism" supposedly can never yield genuine altruism. This concern is based on a half truth. True morality, in other words behavior that most or all people can agree is moral, does consist in the readiness to do the "right" thing even at some personal cost. As pointed out, human beings do not calculate the ultimate effect of every given act on the survival of their own genes or those of close relatives. They are more than just gene replicators. They define each problem, weigh the options, and act in a manner conforming to a well-defined set of beliefs—with integrity, we like to say, and honor, and decency. People are willing to suppress their own desires for a while in order to behave correctly.

That much is true, but to treat such qualifications as objections to naturalistic ethics is to miss the entire force of the empirical argument. There is every reason to believe that most human behavior does protect the individual, as well as the family and the tribe and, ultimately, the genes common to all of these units. The advantage extends to acts generally considered to be moral and selfless. A person functions more efficiently in the social setting if he obeys the generally accepted moral code of his society than if he follows moment-by-moment egocentric calculations. This proposition has been well documented in the case of preliterate societies, of the kind in which human beings lived during evolutionary time. While far from perfect, the correlation is close enough to support the biological view that the epigenetic rules evolved by natural selection.[23]

It should not be forgotten that altruistic behavior is most often directed at close relatives, who possess many of the same genes as the altruist and perpetuate them through collateral descent. Beyond the circle of kinship, altruistic acts are typically reciprocal in nature, performed with the expectation of future reward either in this world or afterward. Note, however, that the expectation does not necessarily employ a crude demand for returns, which would be antithetical to true morality. Rather, I expect you (or God) to help me because it is right for you (or God) to help me, just as it was right for me to help you (or obey God). The reciprocation occurs in the name of morality. When people stop reciprocating, we tend to regard them as outside the moral framework. They are "sociopathic" or "no better than animals."

The very concept of morality—as opposed to mere moral decisions taken from time to time—imparts efficiency to the adaptively correct action. Moral feeling is the shortcut taken by the mind to make the best choices quickly. So we select a certain action and not another because we feel that it is "right," in other words, it satisfies the norms of our society or religion and thence, ultimately, the epigenetic rules and their prescribing genes. To recognize this linkage does not diminish the validity and robustness of the end result. Because moral consistency feeds mental coherence, it retains power even when understood to have a purely material basis.

For the same reason there is little to fear from moral relativism. A common argument raised against the materialist view of human nature is that if ethical premises are not objective and external to mankind, the individual is free to pick his own code of conduct regardless of the effect on others. Hence philosophy for the philosophers and religion for the rest, as in the Averrhoist doctrine. But our growing knowledge of evolution suggests that this is not at all the case. The epigenetic rules of mental development are relative only to the species. They are not relative to the individual. It is easy to imagine another form of intelligent life with nonhuman rules of mental development and therefore a radically different ethic. Human cultures, in contrast, tend to converge in their morality in the manner expected when a largely similar array of epigenetic rules meet a largely similar array of behavioral choices. This would not be the case if human beings differed greatly from one another in the genetic basis of their mental development.

Indeed, the materialist view of the origin of morality is probably less threatening to moral practice than a religious or otherwise nonmaterialistic view, for when moral beliefs are studied empirically, they are less likely to deceive. Bigotry declines because individuals cannot in any sense regard themselves as belonging to a chosen groups or as the sole bearers of revealed truth. The quest for scientific understanding replaces the hajj and the holy grail. Will it acquire a similar passion? That depends upon the value people place upon themselves, as opposed to their imagined rulers in the realms of the supernatural and the eternal.

Nevertheless, because ours is an empirical position, we do not exclude the possibility that some differences might exist between large groups in the epigenetic rules governing moral awareness. Already there is related work suggesting that the genes can cause broad social differences between groups—or, more precisely, that the frequency of genes affecting social behavior can shift across geographic regions.

An interesting example now being investigated is variation in alcohol consumption and the conventions of social drinking. Alcohol (ethanol) is broken down in two steps, first to acetaldehyde by the enzyme alcohol dehydrogenase and then to acetic acid by the enzyme acetaldehyde dehydrogenase. The reaction to alcohol depends substantially on the rate at which ethanol is converted into these two products. Acetaldehyde causes facial flushing, dizziness, slurring of words, and sometimes nausea. Hence the reaction to drinking depends substantially on the concentration of acetaldehyde in the blood, and this is determined by the efficiency of the two enzymes. The efficiency of the enzymes depends in turn on their chemical structure, which is prescribed by genes that vary within populations. In particular, two alleles (gene forms) are known for one of the loci (chromosome sites of the genes) encoding alcohol dehydrogenase, and two are known for a locus encoding acetaldehyde dehydrogenase. These various alleles produce enzymes that are either fast or slow in converting their target substances. Thus one combination of alleles causes a very slow conversion from ethanol to acetic acid, another the reverse, and so on through the four possibilities.

Independent evidence has suggested that the susceptibility to alcohol addiction is under partial genetic control. The tendency now appears to be substantially although not exclusively affected by the combination of genes determining the rates of ethanol and acetaldehyde conversion. Individuals who accumulate moderate levels of acetaldehyde are more likely to become addicted than those who sustain low levels. The propensity is especially marked in individuals who metabolize both ethanol and acetaldehyde rapidly and hence are more likely to consume large quantities to maintain a moderate acetaldehyde titer.

Differences among human populations also exist. Most caucasoids have slow ethanol and acetaldehyde conversion rates, and thus are able to sustain moderately high drinking levels while alone or in social gatherings. In contrast, most Chinese and Japanese convert ethanol rapidly and acetaldehyde slowly and thus build up acetaldehyde levels quickly. They reach intoxication levels with the consumption of a relatively small amount of alcohol.

Statistical differences in prevalent drinking habits are well known between the two cultures, with Europeans and North Americans favoring the consumption of relatively large amounts of alcohol during informal gatherings and eastern Asiatics favoring the consumption of smaller amounts on chiefly ceremonial occasions. The divergence would now seem not to be wholly a matter of historical accident but to stem from biological differences as well. Of course a great deal remains to be learned concerning the metabolism of alcohol and its effects on behavior, but enough is known to illustrate the potential of the interaction of varying genetic material and the environment to create cultural diversity.[24]

It is likely that such genetic variation accounts for only a minute fraction of cultural diversity. It can be shown that a large amount of the diversity can arise purely from the statistical scatter due to differing choices made by genetically identical individuals, creating patterns that are at least partially predictable from a knowledge of the underlying universal bias.[25] We wish only to establish that, contrary to prevailing opinion in social theory but in concert with the findings of evolutionary biology, cultural diversity can in some cases be enhanced by genetic diversity. It is wrong to exclude *a priori* the possibility that biology plays a causal role in the differences in moral attitude among different societies. Yet even this complication gives no warrant for extreme moral relativism. Morality functions within groups and now increasingly across groups, and the similarities between all human beings appear to be far greater then any differences.

The last barrier against naturalistic ethics may well be a lingering belief in the absolute distinction between *is* and *ought*. Note that we say ''absolute.'' There can be no question that *is* and *ought* differ in meaning, but this distinction in no way invalidates the evolutionary approach. We started with Hume's own belief that morality rests ultimately on sentiments and feelings. But then we used the evolutionary argument to discount the possibility of an objective, external reference for morality. Moral codes are seen instead to be created by culture under the biasing influence of the epigenetic rules and legitimated by the illusion of objectivity. The more fully this process is understood, the sounder and more enduring can be the agreements.

Thus the explanation of a phenomenon such as biased color vision or altruistic feelings does not lead automatically to the prescription of the phenomenon as an ethical guide. But this explanation, the *is* statement, underlies the reasoning used to create moral codes. Whether a behavior is deeply ingrained in the epigenetic rules, whether it is adaptive or nonadaptive in modern societies, whether it is linked to other forms of behavior under the influence of separate developmental rules: all these qualities can enter the foundation of the moral codes. Of equal importance, the means by which the codes are created, entailing the estimation of consequences and the settling upon contractual arrangements, are cognitive processes and real events no less than the more elementary elements they examine.

No major subject is more important or relatively more neglected at the present time than moral philosophy. If viewed as a pure instrument of the humanities, it seems heavily worked, culminating a long and distinguished history. But if viewed as an applied science in addition to being a branch of philosophy, it is no better than rudimentary. This estimation is not meant to be derogatory. On the contrary, moral reasoning offers an exciting potential for empirical research and a new understanding of human behavior, providing biologists and psychologists join in its development. Diverse kinds of empirical information, best obtained through collaboration, are required to advance the subject significantly. As in twentieth-century science, the time of the solitary scholar pronouncing new systems in philosophy seems to have passed.

The very weakness of moral reasoning can be taken as a cause for optimism. By comparision with the financial support given other intellectual endeavors directly related to human welfare, moral philosophy is a starveling field. The current expenditure on health-related biology in the United States at the present time exceeds $3 billion. Support has been sustained at that level or close to it for over two decades, with the result that the fundamental processes of heredity and much of the molecular machinery of the cell have been elucidated. And yet a huge amount remains to be done: the cause of cancer is only partly understood, while the mechanisms by which cells differentiate and assemble into tissues and organs are still largely unknown. In contrast, the current support of research on subjects directly related to moral reasoning, including the key issues in neurobiology, cognitive development, and sociobiology, is probably less than 1 percent of that allocated to health-related biology. Given the complexities of the subject, it is not surprising that very little has been learned about the physical basis of morality—so little, in fact, that its entire validity can still be questioned by critics. We have argued that not only is the subject valid, but it offers what economists call increasing returns to scale. Small absolute increments in effort will yield large relative returns in concrete results. With this promise in mind, we will close with a brief characterization of several of the key problems of ethical studies as we see them.

First, only a few processes in mental development have been worked out in enough detail to measure the degree of bias in the epigenetic rules. The linkage from genes to cellular structure and thence to forms of social behavior is understood only partially. In addition, a curious disproportion exists: the human traits regarded as most positive, including altruism and creativity, have been among the least analyzed empirically. Perhaps they are protected by an unconscious taboo causing them to be regarded as matters of the ‘‘spirit’’ too sacred for material analysis.

Second, the interactive effects of cognition also remain largely unstudied. Among them are hierarchies in the expression of epigenetic rules. An extreme example is the suppression of preference in one cognitive category when another is activated. This is the equivalent to the phenomenon in heredity known as epistasis. We know in a very general way that certain desires and emotion-laden beliefs take precedence over others. Tribal loyalty can easily dominate other social bonds, especially when the group is threatened from the outside. Individual sacrifice becomes far more acceptable when it is believed to enhance future generations. The physical basis and relative quantitative strengths of such effects are almost entirely unknown.

Third, there is an equally enticing opportunity to create a comparative ethics, defined as the study of conceivable moral systems that might evolve in other intelligent species. Of course it is likely that even if such systems exist, we will never perceive them directly. But that is beside the point. Theoretical science, defined as the study of all conceivable worlds, imagines nonexistent phenomena in order to classify more precisely those that do exist. So long as we confine ourselves to one rather aberrant primate species (our own), we will find it difficult to identify the qualities of ethical premises that can vary and thus provide more than a narrow perspective in moral studies. The goal is to locate human beings within the space of all possible moral systems, in order to gauge our strengths and weaknesses with greater precision.

Fourth, there are pressing issues arising from the fact that moral reasoning is dependent upon the scale of time. The trouble is that evolution gave us abilities to deal principally with short-term moral problems. ("Save that child!" "Fight that enemy!") But, as we now know, short-term responses can easily lead to long-term catastrophes. What seems optional for the next ten years may be disastrous thereafter. Cutting forests and exhausting nonrenewable energy sources can produce a healthy, vibrant population for one generation—and starvation for the next ten. Perfect solutions probably do not exist for the full range of time in most categories of behavior. To choose what is best for the near future is relatively easy. To choose what is best for the distant future is also relatively easy, providing one is limited to broad generalities. But to choose what is best for both the near and distant futures is forbiddingly difficult, often drawing on internally contradictory sentiments. Only through study will we see how our short-term moral insights fail our long-term needs, and how correctives can be applied to formulate more enduring moral codes.

Notes

1. H. J. Muller is quoted by G. G. Simpson in *This View of Life* (New York: Harcourt, Brace & World, 1964), 36.

2. See the following widely used textbooks: J. Roughgarden, *Theory of Population Genetics and Evolutionary Ecology: An Introduction* (New York: Macmillan, 1979); D. L. Hartl, *Principles of Population Genetics* (Sunderland, Mass.: Sinauer Associates, 1980); R. M. May (ed.), *Theoretical Ecology: Principles and Applications*, 2d ed. (Sunderland, Mass.: Sinauer Associates, 1981); J. R. Krebs and N. B. Davies (eds.), *Behavioral Ecology: An Evolutionary Approach*, 2d ed. (Sunderland, Mass.: Sinauer Associates, 1984).

3. Reviews of the various modes of selection, including forms that direct individuals away from competitive behavior, can be found in E. O. Wilson, *Sociobiology: The New Synthesis* (Cambridge, Mass.: Belknap Press of Harvard University Press, 1975); G. F. Oster and E. O. Wilson, *Caste and Ecology in the Social Insects* (Princeton: Princeton University Press, 1978); S. A. Boorman and P. R. Levitt, *The Genetics of Altruism* (New York: Academic Press, 1980); D. S. Wilson, *The Natural Selection of Populations and Communities* (Menlo Park, Calif.: Benjamin/Cummings, 1980).

4. For example, the debate over "punctuated equilibrium" versus "gradualism" among paleontologists and geneticists. For most biologists, the issue is not the mechanism of evolution but the conditions under which evolution sometimes proceeds rapidly and sometimes slows to a crawl. There is no difficulty in explaining the variation in rates. On the contrary, there is a surplus of plausible explanations, virtually all consistent with neo-Darwinian theory, but insufficient data to choose among them. See, for example, S. J. Gould and N. Eldredge, "Punctuated Equilibria: The Tempo and Mode of Evolution reconsidered," *Paleobiology* 3 (1977), 115–151; and J. R. G. Turner, "'The Hypothesis That Explains Mimetic Resemblance Explains Evolution': The Gradualist-Saltationist Schism," in M. Grene (ed.), *Dimensions of Darwinism* (Cambridge University Press, 1983), 129–169.

5. See note 2.

6. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, 1983).

7. C. G. Sibley and J. E. Ahlquist, "The Phylogeny of the Hominoid Primates, as Indicated by DNA-DNA Hybridization," *Journal of Molecular Evolution* 20 (1984), 2–15.

8. We are grateful to Victor A. McKusick for providing the counts of identified and inferred human genes up to 1984.

9. G. C. Ashton, J. J. Polovina, and S. G. Vandenberg, "Segregation Analysis of Family Data for 15 Tests of Cognitive Ability," *Behavior Genetics* 9 (1979), 329–347.

10. S. D. Smith, W. J. Kimberling, B. F. Pennington, and H. A. Lubs, "Specific Reading Disability: Identification of an Inherited Form through Linkage Analysis," *Science* 219 (1982), 1345–1347.

11. See J. C. Hall and R. J. Greenspan, "Genetic Analysis of *Drosophila* Neurobiology," *Annual Review of Genetics* 13 (1979), 127–195.

12. See, for example, the recent analysis by J. R. Anderson, *The Architecture of Cognition* (Cambridge, Mass.: Harvard University Press, 1983).

13. See note 3.

14. The evidence for biased epigenetic rules of mental development is summarized in C. J. Lumsden and E. O. Wilson, *Genes, Mind, and Culture* (Cambridge, Mass.: Harvard University Press, 1981) and *Promethean Fire: Reflections on the Origin of Mind* (Cambridge, Mass.: Harvard University Press, 1983).

15. A new discipline of decision making is being developed in cognitive psychology based upon the natural means—one can correctly say the epigenetic rules—by which people choose among alternatives and reach agreements. See, for example, A. Tversky and D. Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science* 211 (1981), 453–458; and R. Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984).

16. E. Rosch, "Natural Categories," *Cognitive psychology* 4 (1973), 328–350.

17. The epigenetic rules of cognitive development analysed through the year 1980 are reviewed by C. J. Lumsden and E. O. Wilson, op. cit.

18. N. A. Lassen, D. H. Ingvar and E. Skinhøj, ''Brain Function and Blood Flow,'' *Scientific American* 239 (1978), 62–71.

19. A. P. Wolf and C. S. Huang, *Marriage and Adoption in China, 1845–1945* (Stanford University Press, 1980); J. Shepher, *Incest: A Biosocial View* (New York: Academic Press, 1983); P. L. van den Berghe, ''Human Inbreeding Avoidance: Culture in Nature,'' *Behavioral and Brain Sciences* 6 (1983), 91–123.

20. C. J. Lumsden and E. O. Wilson, op. cit. See also the précis of *Genes, Mind, and Culture* and commentaries on the book by twenty-three authors in *Behavioral and Brain Sciences* 5 (1982), 1–37.

21. J. Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), 502–503.

22. This is the argument proposed by R. Nozick in *Philosophical Explanations* (Cambridge, Mass.: Belknap Press of Harvard University Press, 1981) in order to escape the implications of sociobiology.

23. See note 16.

24. E. Jones and C. Aoki, ''Genetic and Cultural Factors in Alcohol Use'' (submitted to *Science*).

25. C. J. Lumsden and E. O. Wilson, op. cit., who show the way to predict cultural diversity caused by random choice patterns in different societies.

27 Four Ways of "Biologicizing" Ethics

Philip Kitcher

I

In 1975, E. O. Wilson invited his readers to consider "the possibility that the time has come for ethics to be removed temporarily from the hands of the philosophers and biologicized" (Wilson 1975: 562). There should be no doubting Wilson's seriousness of purpose.[1] His writings from 1975 to the present demonstrate his conviction that nonscientific, humanistic approaches to moral questions are indecisive and uninformed, that these questions are too important for scholars to neglect, and that biology, particularly the branches of evolutionary theory and neuroscience that Wilson hopes to bring under a sociobiological umbrella, can provide much-needed guidance. Nevertheless, I believe that Wilson's discussions of ethics, those that he has ventured alone and those undertaken in collaboration first with the mathematical physicist Charles Lumsden and later with the philosopher Michael Ruse, are deeply confused through failure to distinguish a number of quite different projects. My aim in this chapter is to separate those projects, showing how Wilson and his co-workers slide from uncontroversial truisms to provocative falsehoods.

Ideas about "biologicizing" ethics are by no means new, nor are Wilson's suggestions the only proposals that attract contemporary attention.[2] By the same token, the distinctions that I shall offer are related to categories that many of those philosophers Wilson seeks to enlighten will find very familiar. Nonetheless, by developing the distinctions in the context of Wilson's discussions of ethics, I hope to formulate a map on which would-be sociobiological ethicists can locate themselves and to identify questions that they would do well to answer.

From K. Bayertz (ed.), *Evolution und Ethik*. Reclam, 1993.

II

How do you "biologicize" ethics? There appear to be four possible endeavors:

1. Sociobiology has the task of explaining how people have come to acquire ethical concepts, to make ethical judgments about themselves and others, and to formulate systems of ethical principles.
2. Sociobiology can teach us facts about human beings that, in conjunction with moral principles that we already accept, can be used to derive normative principles that we had not yet appreciated.
3. Sociobiology can explain what ethics is all about and can settle traditional questions about the objectivity of ethics. In short, sociobiology is the key to metaethics.
4. Sociobiology can lead us to revise our system of ethical principles, not simply by leading us to accept new derivative statements—as in number 2 above—but by teaching us new fundamental normative principles. In short, sociobiology is not just a source of facts but a source of norms.

Wilson appears to accept all four projects, with his sense of urgency that ethics is too important to be left to the "merely wise" (1978:7) giving special prominence to endeavor 4. (Endeavors 2 and 4 have the most direct impact on human concerns, with endeavor 4 the more important because of its potential for fundamental changes in prevailing moral attitudes. The possibility of such changes seems to lie behind the closing sentences of Ruse and Wilson 1986.) With respect to some of these projects, the evolutionary parts of sociobiology appear most pertinent; in other instances, neurophysiological investigations, particularly the exploration of the limbic system, come to the fore.

Relatives of endeavors 1 and 2 have long been recognized as legitimate tasks. Human ethical practices have histories, and it is perfectly appropriate to inquire about the details of those histories. Presumably, if we could trace the history sufficiently far back into the past, we would discern the coevolution of genes and culture, the framing of social institutions, and the introduction of norms. It is quite possible, however, that evolutionary biology would play only a very limited role in the story. All that natural selection may have done is to equip us with the capacity for various social arrangements and the capacity to understand and to formulate ethical rules. Recognizing that not every trait we care to focus on need have been the target of natural selection, we shall no longer be tempted to argue that any respectable history of our ethical behavior must identify some selective advantage for those beings who first adopted a system of ethical precepts. Perhaps the history of ethical thinking instantiates one of those coevolutionary models that show cultural selection's interfering with natural selection (Boyd and Richerson 1985). Perhaps what is selected is some very general capacity for learning and acting that is manifested in various aspects of human behavior (Kitcher 1990).

Nothing is wrong with endeavor 1, so long as it is not articulated in too simplistic a fashion and so long as it is not overinterpreted. The reminders of the last paragraph are intended to forestall the crudest forms of neo-Darwinian development of this endeavor. The dangers of overinterpretation, however, need more detailed charting. There is a recurrent tendency in Wilson's writings to draw unwarranted conclusions from the uncontroversial premise that our ability to make ethical judgments has a history, including, ultimately, an evolutionary history. After announcing that "everything human, including the mind and culture, has a material base and originated during the evolution of the human genetic constitution and its interaction with the environment" (Ruse and Wilson 1986: 173), the authors assert that "accumulating empirical knowledge" of human evolution "has profound consequences for moral philosophy" (174). For that knowledge "renders increasingly less tenable the hypothesis that ethical truths are extrasomatic, in other words divinely placed within the brain or else outside the brain awaiting revelation" (174). Ruse and Wilson thus seem to conclude that the legitimacy of endeavor 1 dooms the idea of moral objectivity.

That this reasoning is fallacious is evident once we consider other systems of human belief. Plainly, we have capacities for making judgments in mathematics, physics, biology, and other areas of inquiry. These capacities, too, have historical explanations, including, ultimately, evolutionary components. Reasoning in parallel fashion to Ruse and Wilson, we could thus infer that objective truth in mathematics, physics, and biology is a delusion and that we cannot do *any* science without "knowledge of the brain, the human organ where all decisions . . . are made" (173).

What motivates Wilson (and his collaborators Ruse and Lumsden) is, I think, a sense that ethics is different from arithmetic or statics. In the latter instances, we could think of history (including our evolutionary history) bequeathing to us a capacity to learn. That capacity is activated in our encounters with nature, and we arrive at objectively true beliefs about what nature is like. Since they do not see how a similar account could work in the case of moral belief, Wilson, Ruse, and Lumsden suppose that their argument does not generalize to a denunciation of the possibility of objective knowledge. This particular type of skepticism about the possibility of objectivity in ethics is revealed in the following passage: "But the philosophers and theologians have not yet shown us how the final ethical truths will be recognized as things apart from the idiosyncratic development of the human mind" (Lumsden and Wilson 1983: 182–183).

There is an important challenge to those who maintain the objectivity of ethics, a challenge that begins by questioning how we obtain ethical knowledge. Evaluating that challenge is a complex matter I shall take up in connection with project 3. However, unless Wilson has independent arguments for resolving questions in metaethics, the simple move from the legitimacy of endeavor 1 to the "profound consequences for moral philosophy" is a blunder. The "profound consequences" result not from any novel information provided by recent evolutionary theory but from arguments that deny the possibility of assimilating moral beliefs to other kinds of judgments.

III

Like endeavor 1, endeavor 2 does not demand the removal of ethics from the hands of the philosophers. Ethicists have long appreciated the idea that facts about human beings, or about other parts of nature, might lead us to elaborate our fundamental ethical principles in previously unanticipated ways. Cardcarrying Utilitarians who defend the view that morally correct actions are those that promote the greatest happiness of the greatest number, who suppose that those to be counted are presently existing human beings, and who identify happiness with states of physical and psychological well-being will derive concrete ethical precepts by learning how the maximization of happiness can actually be achieved. But sociobiology has no monopoly here. Numerous types of empirical investigations might provide relevant information and might contribute to a profitable division of labor between philosophers and others.

Consider, for example, a family of problems with which Wilson, quite rightly, has been much concerned. There are numerous instances in which members of small communities will be able to feed, clothe, house, and educate themselves and their children far more successfully if a practice of degrading the natural environment is permitted. Empirical information of a variety of types is required for responsible ethical judgment. What alternative opportunities are open to members of the community if the practice is banned? What economic consequences would ensue? What are the ecological implications of the practice? All these are questions that have to be answered. Yet while amassing answers is a prerequisite for moral decision, there are also issues that apparently have to be resolved by pondering fundamental ethical principles. How should we assess the different kinds of value (unspoiled environments, flourishing families) that figure in this situation? Whose interests, rights, or well-being deserve to be counted?

Endeavors like the second one are already being pursued, especially by workers in medical ethics and in environmental ethics. It might be suggested that sociobiology has a particularly important contribution to make to this general enterprise, because it can reveal to us our deepest and most entrenched desires. By recognizing those desires, we can obtain a fuller understanding of human happiness and thus apply our fundamental ethical principles in a more enlightened way. Perhaps. However, as I have argued at great length (Kitcher 1985), the most prominent sociobiological attempts to fathom the springs of human nature are deeply flawed, and remedying the deficiencies requires integrating evolutionary ideas with neuroscience, psychology, and various parts of social science (see Kitcher 1987a, 1987b, 1988, 1990). In any event, recognizing the legitimacy of endeavor 2 underscores the need to evaluate the different desires and interests of different people (and, possibly, of other organisms), and we have so far found no reason to think that sociobiology can discharge that quintessentially moral task.

IV

Wilson's claims about the status of ethical statements are extremely hard to understand. It is plain that he rejects the notion that moral principles are objective because they encapsulate the desires or commands of a deity (a metaethical theory whose credentials have been doubtful ever since Plato's *Euthyphro*). Much of the time he writes as though sociobiology settled the issue of the objectivity of ethics negatively. An early formulation suggests a simple form of emotivism:

> Like everyone else, philosophers measure their personal emotional responses to various alternatives as though consulting a hidden oracle. That oracle resides within the deep emotional centers of the brain, most probably within the limbic system, a complex array of neurons and hormone-secreting cells located just below the "thinking" portion of the cerebral cortex. Human emotional responses and the more general ethical practices based on them have been programmed to a substantial degree by natural selection over thousands of generations. (1978:6)

Stripped of references to the neural machinery, the account Wilson adopts is a very simple one. The content of ethical statements is exhausted by reformulating them in terms of our emotional reactions. Those who assent to, "Killing innocent children is morally wrong," are doing no more than reporting on a feeling of repugnance, just as they might express gastronomic revulsion. The same type of metaethics is suggested in more recent passages, for example, in the denial that "ethical truths are extrasomatic" which I have already quoted.

Yet there are internal indications and explicit formulations that belie interpreting Wilson as a simple emotivist. Ruse and Wilson appear to support the claim that "'killing is wrong' conveys more than merely 'I don't like killing'" (1986:178). Moreover, shortly after denying that ethical truths are extrasomatic, they suggest that "our strongest feelings of right and wrong" will serve as "a foundation for ethical codes" (173), and their paper concludes with the visionary hope that study will enable us to see "how our short-term moral insights fail our long-term needs, and how correctives can be applied to formulate more enduring moral codes" (192). As I interpret them, they believe that some of our inclinations and disinclinations, and the moral judgments in which they are embodied, betray our deepest desires and needs and that the task of formulating an "objective" ("enduring," "corrected") morality is to identify these desires and needs, embracing principles that express them.

Even in Wilson's earlier writings, he sounds themes that clash with any simple emotivist metaethics. For example, he acknowledges his commitment to different sets of "moral standards" for different populations and different groups within the same population (1975:564). Population variation raises obvious difficulties for emotivism.

On emotivist grounds, deviants who respond to the "limbic oracle" by wilfully torturing children must be seen as akin to those who have bizarre gastronomic preferences. The rest of us may be revolted, and our revulsion may even lead us to interfere. Yet if pressed to defend ourselves, emotivism forces us to concede that there is no standpoint from which our actions can be judged as objectively more worthy than the deeds we try to restrain. The deviants follow their hypothalamic imperative, and we follow ours.

I suspect that Wilson (as well as Lumsden and Ruse) is genuinely torn between two positions. One hews a hard line on ethical objectivity, drawing the "profound consequence" that there is no "extrasomatic" source of ethical truth and accepting an emotivist metaethics. Unfortunately, this position makes nonsense of Wilson's project of using biological insights to fashion an improved moral code and also leads to the unpalatable conclusion that there are no grounds for judging those whom we see as morally perverse. The second position gives priority to certain desires, which are to be uncovered through sociobiological investigation and are to be the foundation of improved moral codes, but it fails to explain what normative standard gives these desires priority or how that standard is grounded in biology. In my judgment, much of the confusion in Wilson's writings comes from oscillating between these two positions.

I shall close this section with a brief look at the line of argument that seems to lurk behind Wilson's emotivist leanings. The challenge for anyone who advocates the objectivity of ethics is to explain in what this objectivity consists. Skeptics can reason as follows: If ethical maxims are to be objective, then they must be objectively true or objectively false. If they are objectively true or objectively false, then they must be true or false in virtue of their correspondence with (or failure to correspond with) the moral order, a realm of abstract objects (values) that persists apart from the natural order. Not only is it highly doubtful that there is any such order, but, even if there were, it is utterly mysterious how we might ever come to recognize it. Apparently we would be forced to posit some ethical intuition by means of which we become aware of the fundamental moral facts. It would then be necessary to explain how this intuition works, and we would also be required to fit the moral order and the ethical intuition into a naturalistic picture of ourselves.

The denial of "extrasomatic" sources of moral truth rests, I think, on this type of skeptical argument, an argument that threatens to drive a wedge between the acquisition of our ethical beliefs and the acquisition of beliefs about physics or biology (see the discussion of endeavor 1 above). Interestingly, an exactly parallel argument can be developed to question the objectivity of mathematics. Since few philosophers are willing to sacrifice the idea of mathematical objectivity, the philosophy of mathematics contains a number of resources for responding to that skeptical parallel. Extreme Platonists accept the skeptic's suggestion that objectivity requires an abstract mathematical order, and they try to show directly how access to this order is possible, even on naturalistic grounds. Others assert the objectivity of mathematics without claiming that mathematical statements are objectively true or false. Yet others may develop an account of mathematical truth that does not presuppose the existence of abstract objects, and still others allow abstract objects but try to dispense with mathematical intuition.

Analogous moves are available in the ethical case. For example, we can sustain the idea that some statements are objectively justified without supposing that such statements are true. Or we can abandon the correspondence theory of truth for ethical statements in favor of the view that an ethical statement is true if it would be accepted by a rational being who proceeded in a particular way. Alternatively, it is possible to accept the thesis that there is a moral order but understand this moral order in naturalistic terms, proposing, for example, with the Utilitarians, that moral goodness is to be equated with the maximization of human happiness and that moral rightness consists in the promotion of the moral good. Yet another option is to claim that there are indeed nonnatural values but that these are accessible to us in a thoroughly familiar way—for example, through our perception of people and their actions. Finally, the defender of ethical objectivity may accept all the baggage that the skeptic assembles and try to give a naturalistic account of the phenomena that skeptics take to be incomprehensible.

I hope that even this brief outline of possibilities makes it clear how a quick argument for emotivist metaethics simply ignores a host of metaethical alternatives—indeed the main alternatives that the ''merely wise'' have canvassed in the history of ethical theory. Nothing in recent evolutionary biology or neuroscience forecloses these alternatives. Hence, if endeavor 3 rests on the idea that sociobiology yields a quick proof of emotivist metaethics, this project is utterly mistaken.

On the other hand, if Wilson and his co-workers intend to offer some rival metaethical theory, one that would accord with their suggestions that sociobiology might generate better (''more enduring'') moral codes, then they must explain what this metaethical theory is and how it is supported by biological findings. In the absence of any such explanations, we should dismiss endeavor 3 as deeply confused.

V

In the search for new normative principles, project 4, it is not clear whether Wilson intends to promise or to deliver. His early writing sketches the improved morality that would emerge from biological analysis.

> In the beginning the new ethicists will want to ponder the cardinal value of the survival of human genes in the form of a common pool over generations. Few persons realize the true consequences of the dissolving action of sexual reproduction and the corresponding unimportance of ''lines'' of descent. The DNA of an individual is made up of about equal contributions of all the ancestors in

> any given generation, and it will be divided about equally among all descendants at any future moment. . . . The individual is an evanescent combination of genes drawn from this pool, one whose hereditary material will soon be dissolved back into it. (1978 : 196–197)

I interpret Wilson as claiming that there is a fundamental ethical principle, which we can formulate as follows:

W: Human beings should do whatever is required to ensure the survival of a common gene pool for *Homo sapiens*.

He also maintains that this principle is not derived from any higher-level moral statement but is entirely justified by certain facts about sexual reproduction. Wilson has little time for the view that there is a fallacy in inferring values from facts (1980a : 431; 1980b : 68) or for the "absolute distinction between *is* and *ought*" (Ruse and Wilson 1986 : 174). It appears, then, that there is supposed to be a good argument to W from a premise about the facts of sex:

S: The DNA of any individual human being is derived from many people in earlier generations and, if the person reproduces, will be distributed among many people in future generations.

I shall consider both the argument from S to W and the correctness of W.

Plainly, one cannot deduce W from S. Almost as obviously, no standard type of inductive or statistical argument will sanction this transition. As a last resort, one might propose that W provides the best explanation for S and is therefore acceptable on the grounds of S, but the momentary charm of this idea vanishes once we recognize that S is explained by genetics, not by ethical theory.

There are numerous ways to add ethical premises so as to license the transition from S to W, but making these additions only support the uncontroversial enterprise 2, not the search for fundamental moral principles undertaken under the aegis of endeavor 4. Without the additions, the inference is so blatantly fallacious that we can only wonder why Wilson thinks that he can transcend traditional criticisms of the practice of inferring values from facts.

The faults of Wilson's method are reflected in the character of the fundamental moral principle he identifies. That principle, W, enjoins actions that appear morally suspect (to say the least). Imagine a stereotypical postholocaust situation in which the survival of the human gene pool depends on copulation between two people. Suppose, for whatever reason, that one of the parties is unwilling to copulate with the other. (This might result from resentment at past cruel treatment, from recognition of the miserable lives that offspring would have to lead, from sickness, of whatever.) Under these circustances, W requires the willing party to coerce the unwilling person, using whatever extremes of force are necessary—perhaps even allowing for the murder of those who attempt to defend the reluctant one. There is an evident conflict between these consequences of W and other ethical principles, particularly those that emphasize the rights and autonomy of individuals. Moreover, the scenario can be developed so as to entail enormous misery for future descendants of the critical pair, thus flouting utilitarian standards of moral correctness. Faced with such difficulties for W, there is little consolation in the thought that our DNA was derived from many people and will be dispersed among many people in whatever future generations there may be. At stake are the relative values of the right to existence of future generations (possibly under dreadful conditions) and the right to self-determination of those now living. The biological facts of reproduction do not give us any information about that relationship.

In his more recent writings, Wilson has been less forthright about the principles of "scientific ethics." Biological investigations promise improved moral codes for the future: "Only by penetrating to the physical basis of moral thought and considering its evolutionary meaning will people have the power to control their own lives. They will then be in a better position to choose ethical precepts and the forms of social regulation needed to maintain the precepts" (Lumsden and Wilson 1983:183). Ruse and Wilson are surprisingly reticent in expressing substantive moral principles, apparently preferring to discuss general features of human evolution and results about the perception of colors. Their one example of an ethical maxim is not explicitly formulated, although since it has to do with incest avoidance, it could presumably be stated as, "Do not copulate with your siblings!" (see Ruse and Wilson 1986:183–185; for discussion of human incest avoidance, see Kitcher 1990). If this is a genuine moral principle at all, it is hardly a central one and is certainly not fundamental.

I believe that the deepest problems with the sociobiological ethics recommended by Wilson, Lumsden, and Ruse can be identified by considering how the most fundamental and the most difficult normative questions would be treated. If we focus attention, on the one hand, on John Rawls's principles of justice (proposals about fundamental questions) or on specific claims about the permissibility of abortion (proposals about a very difficult moral question), we discover the need to evaluate the rights, interests, and responsibilities of different parties. Nothing in sociobiological ethics speaks to the issue of how these potentially conflicting sets of rights, interests, and responsibilities are to be weighed. Even if we were confident that sociobiology could expose the deepest human desires, thus showing how the enduring happiness *of a single individual* could be achieved, there would remain the fundamental task of evaluation the competing needs and plans of different people. Sociobiological ethics has a vast hole at its core—a hole that appears as soon as we reflect on the implications of doomsday scenarios for Wilson's principle (W). Nothing in the later writings of Wilson, Lumsden, and Ruse addresses the deficiency.

The gap could easily be plugged by retreating from project 4 to the uncontroversial project 2. Were Wilson a Utilitarian, he could address the question of evaluating competing claims by declaring that the moral good consists in maximizing total human happiness, conceding that this fundamental moral principle stands outside sociobiological ethics but contending that sociobiology, by revealing our evolved desires, shows us the nature of human happiness. As noted above in connection with project 2, there are grounds for wondering if sociobiology can deliver insights about our "deepest desires." In any case, the grafting of sociobiology onto utilitarianism hardly amounts to the fully naturalistic ethics proclaimed in Wilson's rhetoric.

If we try to develop what I take to be Wilson's strongest motivating idea, the appeal to some extrasociobiological principle is forced upon us. Contrasting our "short-term moral problems" with our "long-term needs," Ruse and Wilson hold out the hope that biological investigations, by providing a clearer picture of ourselves, may help us to reform our moral systems (1986: 192). Such reforms would have to be carried out under the guidance of some principle that evaluated the satisfaction of different desires within the life of an individual. Why is the satisfaction of long-term needs preferable to the palliation of the desires of the moment? Standard philosophical answers to this question often presuppose that the correct course is to maximize the total life happiness of the individual, subject perhaps to some system of future discounting. Whether any of those answers is adequate or not, Wilson needs some principle that will play the same evaluative role if his vision of reforming morality is to make sense. Wilson's writings offer no reason for thinking of project 4 as anything other than a blunder, and Wilson's own program of moral reform presupposes the nonbiological ethics whose poverty he so frequently decries.

VI

Having surveyed four ways of "biologicizing" ethics, I shall conclude by posing some questions for the aspiring sociobiological ethicist. The first task for any sociobiological ethics is to be completely clear about which project (or projects) are to be undertaken. Genuine interchange between biology and moral philosophy will be achieved only when eminent biologists take pains to specify what they mean by the "biologicizations" of ethics, using the elementary categories I have delineated here.

Project 1 is relatively close to enterprises that are currently being pursued by biologists and anthropologists. Human capacities for moral reflection are phenotypic traits into whose histories we can reasonably inquire. However, those who seek to construct such histories would do well to ask themselves if they are employing the most sophisticated machinery for articulating coevolutionary processes and whether they are avoiding the adaptationist pitfalls of vulgar Darwinism.

Project 2 is continuous with much valuable work done in normative ethics over the last decades. Using empirical information, philosophers and collaborators from other disciplines have articulated various types of moral theory to address urgent concrete problems. If sociobiological ethicists intend to contribute to this enterprise, they must explicitly acknowledge the need to draw on extrabiological moral principles. They must also reflect on what ethical problems sociobiological information can help to illuminate and on whether human sociobiology is in any position to deliver such information. Although project 2 is a far more modest enterprise than that which Wilson and his collaborators envisage, I am very doubtful (for reasons given in Kitcher 1985, 1990) that human sociobiology is up to it.

Variants of the refrain that "there is no morality apart from biology" lead sociobiologists into the more ambitious project 3. Here it is necessary for the aspiring ethicists to ask themselves if they believe that some moral statements are true, others false. If they do believe in moral truth and falsity, they should be prepared to specify what grounds such truth and falsity. Those who think that moral statements simply record the momentary impulses of the person making the statement should explain how they cope with people who have deviant impulses. On the other hand, if it is supposed that morality consists in the expression of the "deepest" human desires, then it must be shown how, *without appeal to extrabiological moral principles*, certain desires of an individual are taken to be privileged and how the confliction desires of different individuals are adjudicated.

Finally, those who undertake project 4, seeing biology as the source of fundamental normative principles, can best make their case by identifying such principles, by formulating the biological evidence for them, and by revealing clearly the character of the inferences from facts to values. In the absence of commitment to any specific moral principles, pleas that "the naturalistic fallacy has lost a great deal of its force in the last few years" (Wilson 1980a: 431) will ring hollow unless the type of argument leading from biology to morality is plainly identified. What kinds of premises will be used? What species of inference leads from those premises to the intended normative conclusion?

It would be folly for any philosopher to conclude that sociobiology can contribute nothing to ethics. The history of science is full of reminders that initially unpromising ideas sometimes pay off (but there are even more unpromising ideas that earn the right to oblivion). However, if success is to be won, criticisms must be addressed, not ignored. Those inspired by Wilson's vision of a moral code reformed by biology have a great deal of work to do.

Notes

1. Some of Wilson's critics portray him as a frivolous defender of reactionary conservatism (see, for example, Lewontin, Rose, and Kamin 1984). While I agree with several of the substantive points that these critics make against Wilson's version of human sociobiology, I dissent from their assessment of Wilson's motives and commitments. I make the point explicit because some readers of my *Vaulting Ambition* (1985) have mistaken the sometimes scathing tone of that book for a questioning of Wilson's intellectual honesty or of his seriousness. As my title was intended to suggest, I view Wilson and other eminent scientists who have ventured into human sociobiology as

treating important questions in a ham-fisted way because they lack crucial intellectual tools and because they desert the standards of rigor and clarity that are found in their more narrowly scientific work. The tone of my (1985) work stems from the fact that the issues are so important and the treatment of them often so bungled.

2. For historical discussion, see Richards (1986). Richard Alexander (1987) offers an alternative version of sociobiological ethics, while Michael Ruse (1986) develops a position that is closer to that espoused in Wilson's later writings (particularly in Ruse and Wilson 1986).

References

Alexander, Richard. 1987. *The Morality of Biological Systems*. Chicago: Aldine.

Boyd, Robert, and Peter Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Kitcher, Philip. 1985. *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Cambridge: MIT Press.

———. 1987a. Precis of *Vaulting Ambition* and reply to twenty-two commentators ("Confessions of a Curmudgeon"). *Behavioral and Brain Sciences* 10: 61–100.

———. 1987b. "The Transformation of Human Sociobiology." In A. Fine and P. Machamer (eds.), *PSA 1986*, Proceedings of the Philosophy of Science Association, 63–74.

———. 1988. "Imitating Selection." In Sidney Fox and Mae-Wan Ho (eds.), *Metaphors in the New Evolutionary Paradigm*. Chichester: John Wiley and Sons.

———. 1990. "Developmental Decomposition and the Future of Human Behavioral Ecology." *Philosophy of Science* 57: 96–117.

Lewontin, Richard, Stephen Rose, and Leon Kamin, 1984. *Not in Our Genes*. New York: Pantheon Books.

Lumsden, Charles, and Edward O. Wilson. 1983. *Promethean Fire*. Cambridge: Harvard University Press.

Richards, R. 1987. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago: University of Chicago Press.

Ruse, Michael. 1986. *Taking Darwin Seriously*. London: Routledge.

Ruse, Michael, and Edward O. Wilson. 1986. "Moral Philosophy as Applied Science." *Philosophy* 61: 173–192.

Wilson, Edward O. 1975. *Sociobiology: The New Synthesis*. Cambridge: Harvard University Press.

———. 1978. *On Human Nature*. Cambridge: Harvard University Press.

———. 1980a. "The Relation of Science to Theology." *Zygon* 15: 425–434.

———. 1980b. "Comparative Social Theory." *Tanner Lecture*, University of Michigan.

Index