Many philosophers have maintained that mental predicates are connected by their meanings with the particular appearance and, indeed, physiognomy of human beings, of members of the species *Homo sapiens*. The peculiar non-scientific behaviorism of Ludwig Wittgenstein's treatment of the use of mental terms is the most influential example of this attachment philosophers find between thinking and being a human being. In his exposition of Wittgenstein's aphoristic expression of this view, Norman Malcolm has written, "Since it has nothing like a human face and body, it makes no sense to say of a tree, or an electronic computer, that it is looking or pointing or fetching something (of course one could always *invent* a sense for such expressions) . . . Things which do not have the human form, or anything like it, not merely do not but *cannot* satisfy the criteria for thinking. I am trying to bring out part of what Wittgenstein meant when he said, 'we only say of a human being, and what is like one that it thinks,' and 'The human body is the best picture of the human soul.' "¹ It is in general hard to understand the views of those who would deny mental states to computers whose behavioral repertoires are identical, under stringent conditions of duplication, to those of human beings, except by attributing to such writers the view that at least some conscious states are either definitionally, or at least empirically, inexplicably limited to members of the species *Homo sapiens*, and perhaps to some specially trained members of closely related primate species.² Even the proponents of machine consciousness seem to accept this doctrine, to the extent that they recommend redefinition of mental predicates in the face of their successes in simulation. Insofar as no simulation or duplication of human behavior by a machine can logically oblige us to accord the machine
mental states of the kind we accord ourselves, the meanings of the terms employed to pick out these states must be given in some species-specific manner. To the extent that the question of whether machines can be accorded these states remains intelligibly open, then no matter how idle for practical purposes the question may ultimately become, the definitions of these states must make appeal to the assumption that the states are at least paradigmatically or exclusively states of men, women, children and a few well-trained chimpanzees. Even if we eventually accord mental states to computers, or to apparently sentient nonterrestrial creatures, it will be on the basis of an analogical argument from our own cases. They will be accorded states of desire and beliefs, and their actions will ordinarily be explained by the citation of a general statement like $L$ only because of the similarity of their behavior to that of the "benchmark" behavior of *Homo sapiens*. Of course we can and do accord nonhuman systems purposive or teleological states independent of any analogical appeal to our own behavior and its determinants. But consciousness involves more than mere teleology, or else we would have to accord it unexceptionally to many complex but purely physical systems like thermostats and inertial guidance systems. It is this extra burden of human behavior, beyond its goal-directedness, that we appeal to conscious states in order to explain, and which philosophers frequently describe in terms of intentionality or content of consciousness. Our difficulty in saying what would constitute a machine's having the desire that or the belief that a particular proposition be true accounts for the resistance to according such states to machines. For we know that humans have such states, although we seem equally perplexed about what exactly constitutes such states in our own cases. Perhaps because we cannot account for the undeniable intentionality of our own states, we can only accord such states to nonliving systems on the basis of a strong analogy between their behaviors and ours, that is, by implicitly defining such states in terms of a paradigm of behavior characteristic of *Homo sapiens*. It is no surprise that those, like Wittgenstein, who deny that we attribute mental states to other human beings on analogical grounds because such grounds are too weak to sustain our moral certainty about these attributions, are the very same philosophers who refuse to accord mental states to nonhuman items on the same ground, that the analogical basis is insufficient to sustain the conclusion that nonhumans have intentional states. The argument between these philosophers and those who would accord mental states to machines and to other nonhumans is in large measure one about the prospects for an analysis of intentionality that will free it from conceptual connection to the concept of a human being, or about the conditions of adequacy that an analogical argument of the required kind will have to satisfy. Both parties to the dispute agree that at present we have no way of separating the meanings of mental terms from the notion of *Homo sapiens*.
For our purposes this fact is important, because, as I shall argue, ‘Homo sapiens’ does not name a natural kind, and is not a predicate at all, let alone a qualitative predicate. Rather it is a name, a proper name for a discrete, spatiotemporally bounded particular thing. As such, it is no more likely to figure in a general law than ‘Napoleon Bonaparte’ or ‘Kalamazoo, Michigan’ or ‘Mona Lisa a Gioconda.’ Although the three items mentioned behave in accordance with general laws, the terms employed to mention them do not figure in the expression of laws, nor do any other predicates defined by appeal to these terms. This is a consequence of the requirement that laws contain only qualitative predicates, that their terms neither explicitly nor implicitly mention particular places or things. But if ‘Homo sapiens’, and all other species terms, like ‘Canis familiaris’, ‘Cygnus olor’, are names of spatiotemporally restricted particular things, then neither they nor any term defined explicitly or implicitly by appeal to them will be permissible in general laws either. And of course if notions like ‘desire’, ‘belief’, and ‘action’ are to be defined, as suggested above, either essentially or paradigmatically as human attributes, then they too will be excluded from general laws as failing to be of the purely qualitative sort which nomological generalization requires. And any general statement which employs them will, if true, be at most an accidental generalization. But this conclusion will give us the independent explanation of why $L$, though true and exceptionless, is no law after all, and will provide just the noncircular explanation of the failure to find laws of human behavior which we require: one consistent both with the dictates of empiricism and with the truth of that vast body of singular causal statements that we express, albeit in nonqualitative terms. Moreover, further reflection on what sort of a term ‘Homo sapiens’ is will reveal the character of the lowest-level laws which we can expect to discover about the behavior of human beings. Our first task in pursuing this line of argument is to show that, in general, species terms do not name natural kinds, but particular things, and this will involve an essential excursus into biological theory.

Our task is made more difficult by the fact that throughout the history of logic the names of species have been invariably treated as the names of classes or sets of indefinitely large numbers of individuals. They have been treated as predicates which figure in open sentences like ‘$x$ is a swan’ that can be true of any number of individual objects from zero to infinity. Species names were initially assigned to distinct kinds of individuals, and much science (natural history) consisted exclusively in consigning individual organisms to their appropriate species. Species themselves were and commonly continue to be assumed to be fixed and utterly distinct in their membership, even though in most cases taxonomists have never been able to provide sets of necessary and sufficient conditions for an organism’s being an example of a given species. Moreover, there are remarkably few generalizations of exceptionless sorts or with any counterfactual force about the species isolated in taxonomic re-
search. 'All swans are white' may serve the philosopher as a generalization for purposes of introductory discussion, but it plainly has exceptions that deprive it of nomological status. In fact, some arguments, to the effect that biology possesses no distinctive laws of its own, trade on the inevitable exceptions and restrictions with which the biologist must hedge around his claims about general features of members of any given species. As we shall see, such arguments reflect mistaken notions about the subject matter of biological laws: the general findings of biology and biological theory do not bear on regularities about particular species, like *Canis familiaris* or *Didus ineptus* (the dodo); they bear on laws with which all species are in conformity. The subject matter of biological theory is the behavior of any species, not of particular ones, just as the subject matter of mechanics is the behavior of any body, but not any particular one.

Of course, the most important theory of biological science is Darwin's theory of natural selection. This theory tells us that as a result of hereditary variation among the members of a species and selection over this variation, species evolve. Subsequent work revealed that there was a special unit of heredity whose behavior accounts both for variation through mutation and inheritance through replication: the gene. The individual organism is, of course, the immediate unit of selection; it is the item on which environmental pressures operate, and whose survival and reproduction determine the character of subsequent members of the species. The crucial feature of Darwinian theory is that the unit of evolution is the species; it is they which evolve, and their evolution consists in changes in the relative proportions with which their members in successive generations manifest the varying hereditary characteristics, or phenotypes, determined by changes in the units of mutation and in the units of selection. This explains both why biologists can provide no necessary and sufficient conditions for various particular species, and why there seem to be no exceptionless generalizations about particular species. Since species evolve, there is no trait which jointly meets the requirement of being hereditary and therefore essential to the species in the way required for defining the particular species, and at the same time either necessary or sufficient for being a member of the species. Similarly, the apparently general statement that 'all polar bears are white' is not a lawlike consequence of evolutionary theory, although if true, its truth can be explained by citing generalizations about the existence of adaptive variations in any species which is exposed to generally described environmental conditions and which survives in those conditions over the long run. Since, according to the theory of natural selection, species evolve, it follows that they should be treated, not as classes whose members satisfy some fixed set of conditions—not even a vague cluster of them—but as lineages, lines of descent, strings of imperfect copies of predecessors, among which there may not even be the manifestation of a set of central and distinctive, let alone necessary
and sufficient, common properties. A kind or sort for which no properties whatever could be definitional is no kind at all, and a kind which remains unchanged while any or all of the defining properties of its instances change over time is equally hard to comprehend. But neither of these difficulties arises for a particular object which may change its properties over time, may evolve, and about which the question of defining properties does not come up. More importantly, biological theory dictates the accordance to species of an individuation and a unity which is unintelligible except on the assumption that species are particulars. Thus, although the disappearance of all atoms of atomic number 79 does not entail the disappearance of a space in the periodic table of the elements, but only its temporary emptiness, the disappearance of all members of a species entails its extinction, the utter disappearance of the species; the appearance of new organisms qualitatively indistinguishable from the extinguished species' members does not, in biological theory, constitute the reappearance of the same species, but represents an entirely new one, just because it did not arise in any line of descent from the old one (which, having no issue, became extinct). Moreover, just as individual organisms, which are themselves particulars, may undergo vast changes in genotypic and phenotypic properties, may divide into two or fuse with another into one, may continue to exist while generating new individuals, so too, evolutionary theory requires that species be capable of all these things. Thus, as binary fission makes two particular organisms out of one, geographic or other environmental isolation can make members of the same lineage reproductively isolated after a long enough period of separation, so that whether their appearance is similar or not, they must be classed as new and different species. Indeed, this is how, on the view of some biologists, speciation proceeds. Where genetic change makes both branches reproductively isolated from the original lineage, the original species may be said to become extinct. When substantial continuity and the possibility of reproduction between members of the original lineage and one of the subsequent lines remains, the phenomenon is akin to that of an individual generating a new individual while continuing to exist as a unit. Similarly, introgression may obtain between two species, sometimes creating a third or interconnecting the lines of descent constituting each species. Now, although examples of one or more kinds of things may do any or all of these things, the kinds themselves cannot. As Plato so clearly recognized, kinds are immutable; it is only things which change. Individuals are spatiotemporally bound; they are discrete entities with a location and a history, even though it is sometimes difficult to plot their boundaries and their beginnings or endings. So too with species. Temporal boundaries among species are marked by extinction, or by the production of sterile offspring, or by the onset of reproduction among an isolated small group of founder organisms, or again by the process of polyploidy, in which miosis—the multiplication of genetic material—occurs with-
out cell division, doubling and thereby changing the chromosomal material that determines reproductive possibilities. Spatial boundaries are simply given by the distribution of members of the species. Spatial continuity is reflected in reproduction and in other sorts of species-specific behavior. Naturally, species do not have all the unity and spatiotemporal continuity of the usual example of a particular object, like a table or a chair, but they certainly have enough to be so classified, when we consider the vagaries of individuation for such particulars as nations or cultures or even organisms with peculiar biological potentials. Most important, species have the uniqueness that is necessary for being individual items and sufficient for not being general classes or kinds of items. Qualitative similarity up to any degree of completeness is neither necessary nor sufficient to determine whether two organisms are members of the same species (as it should be were species kinds or general classes); what is necessary for such determination is that the two organisms be links in the same spatiotemporally restricted chain of genetic inheritance. Without such a connection between the two organisms, there cannot be an evolutionary path between them, and since species are the units in which such paths are laid out, the organisms cannot be members of the same species.⁵

The conceptual status of species names as names of spatiotemporally extended particulars is reflected and reinforced in the character of biological laws, empirical generalizations, and statements of accidental universality or finite scope. Thus, the apparently general statements that all beavers build dams, even if true about each and every beaver, or that all swans are white, do not count as biology's candidates for nomological respectability, nor are they even empirical generalizations of that subject. For, properly understood, they are singular statements which predicate properties to the components of a spatiotemporally restricted particular. The supposition that species are kinds gives such statements the appearance of universality in form, while the existence or physical possibility of exceptions as well as the restriction of their domains seems to deprive them of nomological force. That is why they are sometimes treated as accidental generalizations, and why more than one philosopher of science has denied that biology has any distinctive laws, and has asserted that it simply applies nomological generalizations from the "harder" sciences. But not only are such statements not the laws of biology, they are not even its rough empirical generalizations (general statements with exceptions, but with enough nomological force to permit their explanation by the real laws of biological theory). Among the empirical generalizations of biology are the following statements, which mention no species whatever, but are general claims about all species which are roughly correct:

- Unspecialized species tend to avoid extinction longer than specialized ones.
- Body size tends to increase during the evolution of a species.
Contemporary species living in the same environment tend to change in analogous ways.

In colder regions the members of warmblooded species are larger than members in warmer regions, though their extremities are proportionately smaller.

These are examples of the lowest-level generalizations of biology; they manifest exceptions, like other empirical generalizations, but these exceptions can be explained by appeal to the same more general claims that help explain the exception-ridden generalizations themselves. And these higher-level claims are the laws of biology. These empirical generalizations, unlike statements about swans and beavers, can be expected to obtain, ceteris paribus, on other life-supporting planets, and indeed on planets which manifest “life” that is so different from our own as to be unrecognizable to the anthropocentric eye. If statements like these are the lowest-level generalizations of biology, then the natural kinds of biology cannot be the particular species that the taxonomist constructs, for these laws mention no particular species. Rather, they mention the kind that every particular species is an example of, by virtue of having the property of being a species. If particular species like Ursus ursus or Canis lupus are not natural kinds, since they do not figure in even the lowest-level generalizations of biology, what alternative is there for them besides being treated as names of particulars? We cannot treat these terms as we might treat “phlogiston,” as terms which denote nothing and whose use in singular statements is at best misleading and at worst nonsensical. Species names are not kind-terms which have been superseded in the course of scientific advance. They do enable us to usefully individuate and relate the vast number of individual organisms that populate our planet, and they enable us conveniently to show the hereditary relations between various organisms, their evolutionary distance from one another, and the kinds of environmental factors that make for differences between them. How can we retain the advantages which the use of species terms provides, consistent with the recognition that they do not reflect natural kinds? Certainly not by according them the status of nonnatural kind, like the nonnatural kind “table.” Rather, we may accomplish this by treating species names as names of things instead of kinds, natural things which are named in an organized way that reveals much systematic, biologically fruitful information about them while recognizing their particularity and individuality.

If a statement about the color of swans is not even an empirical generalization, and the claim that body size increases through the evolution of a species is no more than such a generalization, what are examples of biological laws? Good examples are provided by the Mendelian laws of genetics and the laws of segregation and independent assortment of genes. But these laws, like many others in functional biology, are well on their way to reduction to biochemical and chemical laws, and are consequently beginning to lose their
claim to distinctively biological status. More autonomous, and for our purposes, more important examples of biological laws are provided by the principles of evolutionary theory. Such principles are especially significant in the present context, both because their character further establishes the status of species concepts and because we know that these laws govern human phenomena as much as they govern the behavior of any other species, so that the laws of human behavior must minimally be consistent with them. Consider the following four general statements, which, it has been argued, axiomatically represent the content of the theory of natural selection:

1. There is an upper limit to the number of organisms in any generation of a species.
2. Each organism has a certain quantity of fitness with respect to its particular environment.
3. If $D$ is a physically or behaviorally homogeneous subclass of a species, and is superior in fitness to the rest of the members of the species for sufficiently many generations, then the proportion of $D$ in the species will increase.
4. In every generation of a species not on the verge of extinction there is a subclass, $D$, which is superior to the rest of the members of the species for long enough to ensure that $D$ will increase relative to the species, and will retain sufficient superiority to continue to increase, unless it comes to constitute all the living members of the whole species at some time.

Unlike the merely empirical generalizations mentioned above, these principles are supposed to be exceptionless, although there is difficulty giving specific content to the notions of "sufficiently many generations" and "long enough time to ensure increase." The important points are again that these biological laws mention no particular species, and that they treat species as particular evolving lineages, not as types or classes or kinds of organisms. We cannot deduce anything about the evolution of any particular species from these laws, nor can we predict anything about the future of a given species from these laws alone. Some philosophers have criticized evolutionary theory on this score, and even condemned it as an empty and unfalsifiable account bereft of explanatory power; other philosophers have moved in the opposite direction on the tracks of the same argument and claimed that since evolutionary theory makes no predictions, and since it is clearly a theory of great explanatory power, it must be false that there is any parallel between explanation and prediction, contrary to empiricists’ views. Both sorts of philosophers’ complaints reflect a misunderstanding of evolutionary theory, and of the status of claims about particular species. They mistakenly suppose that statements about the properties of species or their members are general statements, and that therefore they should follow from a theory about the evolution of species in general. Since the statements that the theory alone
enables us to infer about particular species are either true because (close to) tautologous or so vague as to be unassessable, some philosophers conclude that logical empiricist strictures on scientific explanation are wrong, for this clearly explanatory theory makes no specific predictions. Other philosophers argue that because such methodological strictures are correct, evolutionary theory is a scientifically disreputable enterprise. In fact, the relationship between the theory and particular statements about special species sustains neither of these two views. The theory cannot be expected to issue in the sort of singular statements that are open to test, without the addition of statements of initial conditions, and yet species-specific statements are of just the former sort. On the other hand, with the provision of the required statement of initial conditions, the theory and its derivative laws will enable us to make predictions up to the levels of accuracy of the statements of initial conditions supplied. For example, from more complex mathematically expressed versions of the four laws of evolutionary theory enumerated above, the following general statement is derivable as a theorem:7

If $D_1$ and $D_2$ are distinct species within an environment of cohabitating organisms, $D$, and if $D_1$ is superior to $D_2$ as long as it constitutes less than $e$, a certain portion of $D$, while $D_2$ becomes superior to $D_1$ when $D_1$ comes to constitute more than that proportion of $D$, then the proportion of $D_1$ to $D$ will either stabilize at $e$ or oscillate around $e$.

Now if $D_1$ and $D_2$ are two species related as predator and prey, then our derived law will explain why their populations oscillate around fixed values. And given two particular species, say caribou and wolves, if it can be independently shown that the wolf population depends for existence solely on predation of the caribou, and that survival of caribou depends solely on avoiding the predation of wolves, the theorem tells us that the average fitness of wolves will decrease as their ratio to caribou increases and thus their survival rate decreases, so that there must be a balance between the numbers of these species. In other words, if we can establish these initial conditions in the interrelations between wolves and caribou, the theory will enable us to predict that each of them will eventually exist at or around a fixed level of population, regardless of what population distribution the species began with.

Of course, the biologist, like other natural scientists, is not content with such a generic prediction, with the prediction of the existence of an equilibrium value of population for competitor species; he would like a prediction of the actual value where that value is unknown, and an explanation of that value where it is known. The method of acquiring such predictions and explanations in biological theory has consisted in the construction of a series of mathematical models of increasing realism and sophistication, with a better and better fit both to actual data and to the factors that theory tells us determine fitness and populations. The earliest of these models was developed by A. J. Lotka and V. Volterra,8 and took the following form:
\[
\frac{dH}{dt} = (a_1 - b_1P)H
\]
\[
\frac{dP}{dt} = (-a_2 + b_2H)P,
\]
where \( H \) is the size of the predated or host species, and \( P \) is the size of the predator or parasite species, \( a_1 \) is the net growth rate per individual of the host or predated species in the absence of predation, and \( b_1 \) is a parameter measuring the predation rate per individual of the predator species. Our derived law tells us that there are values for \( H \) and \( P \) above which their derivatives with respect to time are zero, and this restriction together with the values of the parameters will enable us to generate a precise prediction from evolutionary theory about the population level of given species. This early model, of course, fails to take many evolutionary forces into account, and subsequent models attempt to correct this lack of realism in several directions. For example, we may add variables to the model in order to account for the presence and degree of available cover for prey, or where predation is a function of the food needs of predators instead simply of the density of prey. Alternatively, more general models have been devised, and these have been compared to such natural predator-prey systems as house sparrows and sparrow hawks, muskrat and mink, snowshoe hare and lynx, mule deer and cougar, white-tailed deer and wolf, moose and wolf, and bighorn sheep and wolf.\(^9\) Since the number of such models is potentially very large and the data do not always enable us to make a decisive choice between them, it is a matter of interest to know what the general constraints are on these models by virtue of the law, derived from evolutionary theory, that phenomena modeled must reflect stable equilibria. This too has been a subject of theoretical interest, and it has been argued that the four following assumptions are necessary and sufficient for the required equilibrium:\(^10\)

1. Increasing size in parasite or predator species reduces its own and its host's or prey's growth rate.
2. Increasing size in prey or host species reduces its own growth rate, but increases that of the parasite or predator.
3. For both species there are minimum sizes at which they both have positive growth rates in all circumstances.
4. Each species has a maximum size at which its growth rate sinks to zero.

One feature of this aspect of biological theory that is of special interest in comparison with work in social science is that its concern for the general conditions of population equilibrium, and for the production of proofs of the existence of equilibrium, mirrors the traditional concern of economics for the specification of conditions of general equilibrium in economies, and for the proofs of the existence of this equilibrium. That both economics and biological theory should search for such conditions, and manifest interests
in such proofs, of course reflects the extremal character of their most important theories, utility maximization and natural selection, respectively. The striking difference is that biological theory embodies a much more well-confirmed set of theoretical assumptions—the four axioms of evolutionary theory—and partly in consequence, provides not only formal conditions for general equilibrium but the actual numerical values of population levels to reasonable degrees of accuracy for a variety of actual species at which this equilibrium is reached. All this shows that biology’s general assumptions are far better candidates for nomological standing than those of economics or any of the social sciences, and that the concepts that figure in its assumptions stand a far better chance of being natural kinds than do those of contemporary social science.

Like the economist’s models, the biologist’s models make highly unrealistic assumptions. Thus, consider the following list of assumptions actually employed in and characteristic of models constructed to account for the relationships among different species:

a. The species under investigation occupies a spatially homogeneous environment and conditions are temporally constant.
b. The system is closed, so that interacting species are not reinforced by immigrations or depleted by emigration.
c. Each species responds to changes in its own and others’ sizes instantly, without delay.
d. Variations in the age structure of the members of species do not occur or can be disregarded.
e. The interaction coefficient between each pair of species is unaffected by changes in the species composition of the remainder of the community.
f. The genetic properties and hence the competitive abilities of a species are independent of its size.

Not only will the economist be comfortable with such unrealistic assumptions, but he will recognize some as his own and the rest as parallel to his own. Thus, a and b are assumed in conventional economic theory; c is assumed with respect to decision units and markets; d is assumed to hold for all consumers; and e and f have their duals in assumptions about returns to scale and substitution effects among productive factors. This parallel is highly significant, for it shows that the failures of economic theory to provide predictive results at least as well confirmed and practically useful as those of theoretical biology cannot be attributed solely to its employment of unrealistic assumptions. For these same assumptions figure in the construction of models that we know to be predictively fruitful, and have all the explanatory power of the well-established general theory that they are derived from. The failures of economic theory must be sought at least additionally,
and perhaps exclusively, in the character of its fundamental general statements, instead of in its unrealistic boundary conditions. We know that every one of the six assumptions enumerated above is false and that each of the causal forces whose efficacy they deny actually does help determine the value of the level of population towards which the size of a species tends, so that the models constructed with these assumptions can only provide estimates of this value. But we also know, independently of the models, and of the assumptions, that there is such a value. That is, we can be as certain about the existence of such a value as we are of the general laws of the theory of natural selection from which its existence is deducible. By contrast, we cannot say whether the equilibria levels of prices and production generated by the economist's models, with assumptions like \( a \) through \( f \), are estimates of a value that we know independently to exist because we do not have the same degree of assurance about the assumptions of maximization that play for the economist the role which natural selection of the fittest plays for the biologist. This is a parallel to which we shall return, but for present purposes it is important to note that in biological theory particular species play the same role that individual agents, markets, and industries play in economic theory. Thus, just as economic theory names no agents, firms, markets, or industries, biological theory names no species. Just as economic theory purports to explain the behavior of instances of the kinds agent, market, and industry, so biological theory purports to explain the behavior of the kind species. Just as we would not expect a generalization about General Motors or IBM to follow from the central principles of economic theory alone, neither should we expect to find a generalization about kangaroos or lemmings to follow from evolutionary theory alone. And just as we would not expect an exceptionless statement about Imperial Chemical Industries or all of its employees to figure as a law of nature, derived or not, we cannot expect a statement about giraffes' necks or elephants' trunks to be laws either, no matter how exceptionless their claims are about members of these species. To see all this, one may again turn to the assumptions enumerated above and to their parallels in economic theory. If species were kinds instead of spatiotemporally extended and limited particulars, the parallel between these assumptions and the assumptions economists make about spatiotemporally extended particulars like agents and firms would not be so manifest.

We may conclude from this long excursion that we cannot embrace the truth of the theory of natural selection, in its present character at least, and continue to treat particular species as kinds. Of course, our venture into biological theory has touched on other matters as well, and has involved discussion of matters that will have further ramifications for the argument of this book; otherwise, so extended and technical a discussion would have been out of place. Nevertheless, it is crucial that we be strongly convinced that "Paramecium causatum," "Drosophila melanogaster," and "Homo
"Homo sapiens" are not purely qualitative predicates but name spatiotemporally restricted particulars. It immediately follows from this that we can expect no laws of paramecium behavior, or fruitfly behavior. This is no surprise, for no one has ever expected that there are regularities in nature that are manifested exclusively and uniformly by each and every paramecium or fruitfly. No one has ever expected to detect generalizations about all and only beavers or sunfish, orangutans or baboons. Had such expectations arisen, semi-autonomous sciences of beaver behavior, say, might have appeared, spawned their own journals, launched arguments for their substantive and methodological autonomy, and pursued research programs calculated to establish a systematic hypothetico-deductive theory of beaver behavior and an estimate of the values of the parameters that the theory claims to limit the behavior of members of this species. Not only is the very description of such a science preposterous, but the serious labors of ethologists who study small numbers of one and only one species clearly reflect the fact that among biologists there is no misunderstanding of the status of ethological findings as constituting an autonomous body of nomological generalizations about members of a natural kind, or even preliminary steps to such a body of general laws. And yet the prospects for a science of human behavior, for a body of nomological generalizations manifested exclusively and uniformly by the behavior of each and every member of the species Homo sapiens, are no greater than they are for a science of beaver behavior. And if human behavior is the subject matter of the social sciences, then their claims to substantive and methodological autonomy from the natural sciences and from one another, their research programs devoted to the production of laws of human behavior, or to its economic, or sociological, or indeed psychological aspects, are just as vain as those of a science of beaver behavior.

We may express this claim in terms of the problems of Chapter 5, their solutions, and the limitations on these solutions there noted. There it was claimed that we may formally pass through the horns of the trilemma posed by empiricism's demand for laws of human action, their absence in the social sciences, and the truth of most of our singular judgements about the causes of particular actions, by the hypothesis that the kinds in which we classify the causes in the singular statements are not natural ones. Therefore, it is to be expected that no law relating them to actions will be forthcoming, even though the singular statements are true (they correctly pair actions with their causes, misleadingly described), and even though there are laws (expressed in terms of kinds utterly different from those of common sense) underlying the singular statements. The circumvention of the trilemma proposed was described as formal, because in and by itself it offers what may at best be described as a mere logical possibility. Without some independent evidence for the truth of the hypotheses that underwrite this purely formal, logical possibility, we will lack the conviction that it represents more than such an abstract possibility.
Our excursion into biological theory has provided some of the requisite independent evidence, for it has shown that distinctive laws of human behavior are impossible because they would be about the members of a spatio-temporally restricted particular and would have to be expressed in terms that do not pass the test of being purely qualitative. Yet we have also seen that there are some laws governing human behavior not couched in ordinary terms: minimally, the laws of the theory of natural selection which describe some aspects of the behavior of any species and its members, including *Homo sapiens*. Thus, we have independent evidence that human behavior is law-governed, but that there are no laws of distinctively human behavior, as opposed, say, to mammalian behavior or simply animal behavior. Our excursion provides more evidence for our solution to the trilemma than this, and in fact converts it from the formal solution of a philosophical puzzle into the outline of an independently grounded explanation for the failures of the social sciences.

Recall that in Chapter 5, *L*, our exceptionless general statement relating actions and their reasons, was denied nomological status because the isolation of its constituent terms from correlation with indubitably natural kinds precluded *L*’s nomological entrenchment and consigned it to the status of accidental generalization. Now we can see, again independently of our formal solution to the trilemma, that *L* must indeed be at best an accidental generalization. For *L* asserts the co-occurrence of states of members of the species *Homo sapiens*, states which can only be defined in terms of their exemplification by members of this species, that is, by components of a spatiotemporally restricted particular individual. In fact, this result, that *L* must be an accidental generalization, is independent of the argument that its exceptionlessness coupled with the complexities of neurophysiology jointly insulate it from theoretical entrenchment. For this conclusion now can be seen to turn on recognizing the character of species names and embracing the view that our only current characterizations of the states, events, and conditions mentioned in *L* make inevitable reference to a particular species. If we accept these claims, not only does it turn out to be no surprise that the kinds of events, states, and conditions *L* appeals to are not correlated with a small number of natural kinds described in the biochemical language of neurophysiology, but it would be a surprise—indeed, something little short of a miracle—to discover that there are a finite number of states that jointly satisfy a set of characterizations which are spatiotemporally restricted and another set which do reflect natural kinds. A correlation between classes of states of belief and desire on the one hand and a causally homogeneous class of neurophysiological states would be as surprising as the discovery that each and every winner of the Irish Sweepstakes lottery had an additional chromosome beyond the forty-six we are normally endowed with. Either case would be a puzzle worthy of considerable research. In effect, our discovery about the conceptual status of the term *Homo sapiens* provides us
with an independent reason to believe that \( L \) is an accidental generalization at best, and with an explanation of why it cannot be nomologically entrenched that is equally independent of the fact that so treating it enables us to circumvent the philosophical and methodological trilemma facing us.

Of course, there are laws governing human beings and their behavior— for example, the laws of the theory of natural selection. And there are laws governing the relationships among the particular events, states, and conditions which the antecedents and consequent of \( L \) refer to: the laws of physiology and neurophysiology, of biochemistry, chemistry, and physics. But \( L \) is not deducible from these laws, nor will \( L \) enable us to deduce any further nomological generalizations from these laws, because \( L \) implicitly mentions a particular thing with restricted spatiotemporal location: the species \textit{Homo sapiens}.

The explanation of the failure of the social sciences to have as yet uncovered any general laws need not, therefore, involve the inapplicability of empirical methods in the study of human behavior, nor the denial that most of our singular statements about it are true, nor the greater complexities and observational inaccessibility of the research object of the social sciences. The fault lies with a mistaken belief that the truth of the singular statements, which turns on their successful \textit{reference} to the causes of human action, implies the truth of some nomological generalization which grounds the causal connections on the manifestation of the properties \textit{attributed} to their relata by the singular statements. This is the hypothetical that motivates the search for a law of human action. We may surrender it without surrendering its antecedent, and without surrendering the search for laws of human behavior. We need only recognize that such laws will not express connections between beliefs, desires, and actions, as we ordinarily describe them. How will they express the connections reflected in our true singular statements? This is in effect the question: What are the natural kinds under which the causes and effects described in our ordinary terms really do fall? As such, it is a question that can be answered neither by philosophical analysis nor by empirically untested speculation. The natural kinds under which the particular pairs of causes and effects fall can only be read off the laws of nature in which they figure. Since our own assurance about any of the lawlike statements we embrace is at best inductive, we shall never acquire any greater certainty about just what are these natural kinds into which particular reasons and actions fall.

Of course the physicalist thinks he knows the most \textit{general} natural kinds into which states of human beings and their movements fall: these are just the natural kinds which can be read off the laws of physics, and they subsume the kinds of events, states, and conditions in which humans figure, just because humans are nothing but complex physical systems. Moreover, by virtue of his commitment to the actuality of the reduction of chemistry to
physics, and to the potentiality of reducing biology to chemistry and eventually to physics, the physicalist thinks he knows not only the most general natural kind into which human behavior and its determinants fall, but also the narrower kinds reflected in the distinctive laws of chemistry and biology. For human beings are not only physical systems, they are more specifically complex structures of organic chemical material that satisfy laws governing all biological species. The interesting question for the physicalist is not whether there are natural kinds into which human behavior falls, or what some of these kinds are, but rather what is the narrowest set of kinds which subsume this behavior; in other words, what are the lowest-level laws we can expect to find in our search for the explanation of human behavior. This question is not merely interesting. Answering it is crucial to our expectations for and standards of systematic explanation, reliable prediction, and effective control of human behavior. If the lowest-level laws governing the behavior of people were the laws of physics, if the narrowest natural kinds under which states of belief and desire and the events we call actions are subsumed are just those that can be described in terms of predicates from mechanics, electromagnetism, etc., then we cannot expect to be able to explain or predict the occurrence of a particular human action any more fully than we can explain or predict the movements of a ball on a roulette wheel. We can explain why the ball landed on 28 ex post facto, by attributing values of mechanical variables to the ball and the wheel at some prior time, and then sketching out how these initial conditions, together with the laws of physics, determined the final location of the ball on the roulette wheel. And although in principle there is no limit beyond those of quantum mechanical uncertainty to the exactness of our explanation, in fact we could not have predicted the result by any means that would be of practicable use by a gambler. That is why the roulette wheel is used for gambling. Similarly, if the lowest-level laws governing human behavior were the same ones that govern the behavior of roulette wheels and balls, our explanations of particular human actions could not be expected to be any more than schematic, appealing to the operation of the laws of physics on unspecified values of position and momentum of bodies, with unspecified coefficients of elasticity or tensile strength or thermometric expansion, and so on. And our predictions would require descriptions of initial conditions so esoteric in their demands on physical knowledge and so variable from individual to individual that they would be without practicable significance for any purpose currently envisioned by social scientists and those who employ their findings and theories. And the case would be little different if the narrowest natural kinds subsuming the states, events, and conditions in which humans figure were those of some highly restricted branch of organic chemistry.

It is the desire to explain in more than merely schematic, hand-waving fashion and, even more important, to be able to predict and control to
within practically useful limits the behavior of individuals and groups that leads to the search for laws governing this behavior that are of a far lower level than those of a restricted branch of organic chemistry. The failure to find such laws at the level of belief, desire, and action which the history of social science reflects, and which the conceptual dependence of these variables on a particular species explains, rules out generalizations embodying kinds as narrow as those of belief, desire, and action. A certain amount of work by social and behavioral scientists can be understood as reflecting a recognition of this state of affairs. For example, the research program of operant conditioning associated with the name of B. F. Skinner is clearly one which eschews beliefs, desires, and actions as the kinds about which laws of human and nonhuman behavior can be formulated. Similarly, the attempt to formulate an account of human behavior by inference from successful computer simulations of salient aspects of this behavior can be construed as the attempt to find practically useful terms, less species-specific than "reason" or "action," in which to describe the explananda phenomena, terms which we have some independent reason to believe figure in laws of nature that can be independently established. In the case of Skinnerian conditioning, the laws are alleged to find confirmation in the behavior of all living systems beyond the simplest. In the case of computer simulation, the warrant for the generalizations is to be found more speculatively in information science, and ultimately in our physical understanding of computer simulation of human functions. Consideration of these two research programs in the light of the findings of this chapter will reveal the issues on which their success must hinge, and will also reflect the significance of the claims of this chapter.

Explanations of human action that employ principles from operant learning theory are, like those that figure in ordinary contexts, teleological in form; they explain the occurrence of an event by appeal to its effects on the organism. Unlike the teleological explanations of common sense, however, the forces to which they appeal are not characterized in ways that require them to be conscious states of human beings. Accordingly, the explanatory principles of operant conditioning theory are as applicable to members of any other species as they are to our own. In fact, the principles were first formulated with respect to the behavior of laboratory animals, and only extended in their application to human action by virtue of highly controversial assumptions about that behavior. Without commenting on the merits of those assumptions it is clear that the hypothesis of what operant theorists call instrumental conditioning has enhanced considerably our predictive and manipulative powers with respect to the behavior of children, schizophrenics, and normal individuals in laboratory settings. The leading principle of operant conditioning was first advanced by E. L. Thorndike in the late nineteenth century, and was called "the law of effect." According to this law, and its subsequent amplifications, a positive reinforcement will
increase the probability (or the frequency or the intensity) of the recurrence of the kind of behavior which it follows; its omission will decrease the probability (or the frequency or intensity) of the behavior; the elimination of a negative reinforcer will increase the probability (or intensity or frequency) of the behavior emitted on the occasion of elimination; and a punisher will decrease the probability (or intensity or frequency) of the kind of behavior which it follows. This law is highly generic in that it does not specify the units of the behavior which it purports to explain; nor can we determine without considerable further detailed research whether in individual cases what is explained is the probability of a kind of behavior, or the relative intensity of a particular incident of behavior, or the frequency of emission per unit time; nor does the law specify the delay times between events; most important, in applying it to the explanation of particular cases we must provide a specific characterization of reinforcement and punishment independent of the law of effect. Similarly, if this general claim is to have any chance of the sort of theoretical entrenchment a nomological statement bears, reinforcement and punishment will have to be given general specifications independent of the law of effect—lawlike correlations to natural kinds of states, conditions, and events which characterize the organisms in question and their relations to the environment. In other words, to be acceptable as a law that governs the behavior of any organism, human or otherwise, the law of effect will have to pass the very test which our candidate $L$, of Chapter 5, failed. The demand that it do so is tantamount to the requirement, imposed by operant psychologists upon themselves with as much force as their opponents impose it, that the law of effect, in its particular versions, be testable. These opponents are constantly chiding the proponents of the law with the complaint that the only general characterizations available for reinforcer and punisher make the law of effect an empty tautology. This criticism has some warrant in texts that define "reinforcer" as "any stimulus which, if presented (or withdrawn) contingent on an operant, increases (decreases) the probability of the occurrence of that operant." Such a definition does trivialize the law of effect, but examination of actual practice among psychologists reveals that when they employ the law, they do take steps to specify reinforcers independently. In fact, psychologists do not appeal to the law of effect in its generic form at all in their actual research, even though it bulks large in the textbook expositions of this work. They are interested in formulating and testing a law of effect for, say, pigeons, key-pecking, and food pellets, with a lag time of exactly 0.5 seconds, and a reinforcement schedule of a highly specified type; or again, the law of effect for rats, maze-running, and electric shocks; or monkeys, problem-solving, and free-play opportunities. Reinforcers, the behavior emitted, the rate of reinforcement—all these are specified independently of the law of effect for each of the experiments undertaken. Employing species-specific versions of
the law, psychologists have been able to produce replicable experiments confirming generalizations about the constancy of lag times between reinforcements and behavior, rates of response to given reinforcers, rates of extinction of behavior in the absence of reinforcements, degrees of discrimination abilities generated by conditioning, and variations of rates of learning for different schedules of reinforcement. These generalizations have enabled psychologists to shape, control, and predict the occurrences of particular kinds of behavior for a wide variety of species, including *Homo sapiens*, at a level of accuracy that far exceeds the power of any available alternative hypotheses.

These successes have certainly suggested to operant psychologists and to those who employ their findings that the law of effect, at least in its specific forms, provides a far better guess as to (one of) the lowest-level nomological generalization governing human behavior than might any principle couched in terms of the variables of belief and desire. Such theorists offer as the narrowest natural kind into which human behavior falls the operant, and as the narrowest kinds into which its determinant falls, the reinforcer and the punisher. Their candidates have a clear advantage over reasons and actions if only because their cross-species applicability seems to guarantee their purely qualitative character, and thus does not exclude the possibility of laws formulated in their terms on formal grounds alone. Furthermore, the assumption that some version of the law of effect operates at the level of human behavior is strengthened by its relation to at least some of the true singular statements we make about particular actions and their causes in our beliefs and desires. The law of effect will explain why a particular state of belief and desire resulted in a given action on the assumption that the particular states and action are jointly examples of a kind of operant which has been previously reinforced. Naturally, this kind is not to be characterized in the terms which describe the belief, desire, and action in question as such, nor will subsequent, similarly described co-occurrences of beliefs, desires, and actions of the same or other individuals instantiate the same kind of operant manifested on this occasion, although they too might be explained by the law of effect as an instance of another previously reinforced operant. If all those agents whose states instantiate $L$ do so because the agents satisfy a vast, disorganized, and heterogeneous class of variously reinforced operants, then we shall have an even stronger reason to suppose that there are no expressible laws relating reasons and actions, and that $L$ is at best an accidental generalization. Under these circumstances our inabilitys to predict actions, given knowledge of reasons, and to control them, given manipulation of reasons, will become clear, for they will reflect the heterogeneity of the schedules, types, latencies, and other variables of reinforcement obscured in the commonality of our ordinary descriptive discriminations of operant states, in terms of beliefs, desires, and actions.
But although these assumptions about the powers of operant theory to explain human behavior give flesh to our own account of the failures of the conventional hypotheses to do so, they do not give any independent ground to believe that this theory actually has such powers. And the fear that it does not, that it may be as much of a dead end as the reason-action explanatory model has proved to be, continues to lurk in the repeated charges that the law of effect is an empty tautology, because reinforcement cannot be given a general enough independent characterization to sustain its explanatory force. Despite the local successes of fitting the behavior of various species to smooth curves of reinforcement for situation- and species-specific reinforcers, the law of effect can function as a nomological statement explaining all of these cases as common reflections of the operation of the same mechanism only if we can find a characterization of reinforcement common to all the species-specific behaviors the theory purports to explain, a characterization independent of the law of effect itself. Failure to do this will make the cost of preserving the narrow operant explanations of the behavior of particular organisms in particular experimental and natural settings nothing less than the surrender of the law of effect as the unifying general law behind them all. In other words, without such a specification, the experimental findings will turn out to have the same status as the true singular causal statements of commonsense psychology: true in their reference, but indeterminate or false in their causal attributions. The actual direction of considerable research in experimental psychology reflects this concern. Once psychologists satisfy themselves that a piece of behavior is emitted because of reinforcement of other instances of its kind, they begin to ask why the particular item identified as the reinforcer has the demonstrated effect; psychologists, like other scientists, want to know what the mechanism beneath their lowest-level generalizations is. If they cannot find one, then the status of their generalizations is called into question. At present, it appears that no uniform mechanism accounts for reinforcement across the range of species and settings to which it has been applied. The earliest and most obvious mechanism hypothesized was that all reinforcers satisfy a biological need of one sort or another. Yet, laboratory animals can be trained by administration of saccharin as a reinforcement, and this is a substance which moves through the body practically unchanged and serves no biological need normally construed. Naturally, if operant behavior is reinforced by saccharin-feeding, there must be some mechanism to account for this, but it is not that of fulfilling narrowly biological needs. A related but somewhat broader theory of reinforcement hypothesizes that reinforcers all reduce tension of some sort in organisms. It has been demonstrated, however, that animals can be trained by reinforcers which it is highly implausible to suppose reduce any tensions. The most widely known of such experiments involves the operant training of monkeys to press a bar by reinforcing the behavior with the opening of a
window on their cages that allows them to watch activity in the laboratory. It is hard to say what tensions are thereby reduced. More strikingly, monkeys can be reinforced for puzzle-solving behavior with the solution of the puzzle as sole reward. In an experiment which might be supposed to get at the neurological basis of all reinforcement, James Olds showed that the electrical stimulation of an area of the rat’s midbrain provides it with a most powerful reinforcer. The stimulation is of such intensity and frequency that Olds came to call this area of the brain the pleasure center, and to speculate that all reinforcers are causally connected to the excitation of these regions.¹³ This sort of result is just what the physicalist would expect, and insofar as it might be supposed that operant theory applies to all overt, nonreflex behavior of an organism, neurophysiological correlation is the only sort of independent specification potentially available to provide independent specifications for reinforcers, as it is in the case of the states of agents that figure in L. A general expectation, however, is a far cry from a practically useful independent specification. Moreover, the same neurophysiological structural complexities and functional redundancies that bedevil attempts to specify beliefs and desires in terms of brain states provide obstacles to the provision of neurophysiological correlates for reinforcers as well. The obstacles are not so formidable for the behavioral psychologist, however, because his typology of operant and reinforcement is clearly applicable beyond the species Homo sapiens, and therefore enables him to construct experiments on simpler systems with evidential bearing on hypotheses about more complex ones. The crucial point is that such experiments must ultimately eventuate in manageable small neurophysiological correlates for reasonably broad categories of reinforcers, or they too will turn out to be nonnatural kinds as systematically superfluous as reasons and actions. Pending the successful outcome of such experiments, the law of effect remains at best a candidate for the lowest-level entrenched nomological generalization governing human behavior.

A different sort of research program in psychology which hopes to provide systematic theories of human behavior and the human cognition resulting in human behavior is that associated with investigations of artificial intelligence and the computer simulation of human behavior. Here the strategy, very roughly, is to specify a set of circumstances in which a human being will produce a certain action, and then to write a program for a machine which, given the circumstances as inputs, will generate the description of the action as an output. For instance, faced with a problem in chess, the human being will decide upon and make a certain move. If a computer can be programmed to generate the same move in the same chess problem, then, it is claimed, there are some circumstances under which the program on which the computer operates serves as the explanation of why, under the circumstances, the human performed the action to be explained.
A distinction is often made between work in artificial intelligence and computer simulation. In the former arena, the aim is to design machines and, more usually, programs that will enable computers to perform activities performed by humans, especially activities which are laborious, routine, and highly time-consuming, although perhaps requiring a great deal of simple calculation, correlation, or other sorts of “number-crunching.” Thus, it may be the aim of research in artificial intelligence to design a program that will permit a machine to quickly and accurately complete a quantity of accounting or inventory-taking or multiple regressions or mail-sorting that a human could do, but only in a long period of time, and with the likelihood of a large number of errors produced by boredom, fatigue, inattention, and the like. In writing a program for a machine to take over such functions, we obviously focus on means and powers of the machine that humans do not have; and typically, the program enables the machine to complete tasks by using methods utterly different from those which humans use. Artificial-intelligence research often aims to produce machines that can perform tasks which require from humans the exercise of intelligence, but which can be performed by machines as quickly or more quickly, or more accurately, or more cheaply, by means different from those we use, and which individually need not be described as the exercise of anything like human intelligence. But besides routine jobs that require intelligence of humans but can be performed by computers employing processes that reflect no “intelligence,” we may also attempt to program a computer to play chess, a highly nonroutine activity. An artificial intelligence program for a chess-playing machine will usually trade on its brute force powers to apply a few very simple rules to search out the best move to a breadth and depth of alternative move and countermove far beyond human powers of memory and imagination. This, of course, is not the way we play chess: we apply a large number of very complex rules of choice to a much smaller and narrower range of alternative move and countermove possibilities. The differences in strategy obviously reflect differences in strength and weakness between humans and computers. Computers can do simpler calculations faster and remember them; we can do more complex calculations than we have yet been able to articulate and to program; or perhaps, we do not play chess by calculation at all, but by the employment of pattern-recognition capacities beyond the powers of current computers. It cannot yet be said whether artificial-intelligence research will eventuate in a computer with the chess powers of the best human players.

Unlike artificial intelligence, work in computer simulation has a more specific aim than merely building machines that can produce the same “output” as humans, for a given “input.” In studies of computer simulation the aim is to produce programs with inputs and outputs which are the same as in human cases and in which the program reflects the “same” processes that we employ to generate the given outputs. Although simulation of human
behavior has been judged successful in certain areas, most aspects of this behavior have remained recalcitrant to computer modeling. Attempts to simulate human behavior, however, raise the questions of just what constitutes successful simulation and what sort of understanding of the simulated behavior does such simulation provide. It seems clear that successful simulation of at least some sorts of human behavior need not duplicate all or even most aspects of that behavior. Thus, a successful simulation of chess-playing behavior need not be one in which the machine actually moves the pieces on the board; it would be sufficient for the machine simply to print out a directive to move a specified piece to a particular location, the same piece to the same place that the human, whose behavior is simulated, would move it. It would be a constraint on success that the data fed into the computer be the same as that on the basis of which the agent acts, including the position of the pieces, information about the opponent, the stakes; another important constraint is the requirement that the time scale for easy and difficult moves be at least proportionate between computer and human. Conditions on successful simulation are hard to state in general because they will vary from behavior to behavior and over the degree of simulation possible with given levels of technology.

A question of greater significance is what the simulator hopes to gain by his simulation. The short answer is that the successful simulation of behavior is tantamount to an explanation of it, or at least provides all the resources we require in order to explain it. For while the human represents a black box to the simulator, the machine is, so to speak, a transparent box. Everything about its operation is as well understood as any purely physical system's behavior is: we are acquainted with the laws of physics that govern its behavior, and we know with what program it has been programmed. This knowledge enables us to explain how the machine performs the activities which constitute the simulation and, it is argued, constitute an implicit explanation of the behavior simulated as well. Of course, the conclusion here is a non sequitur, as reflection on results in artificial intelligence reveals. For mere simulation may be a case of artificial intelligence, in which the output of the machine is equivalent to the behavior modeled, but the processes employed to produce it are utterly unlike those humans actually use or in some cases could use. The question therefore arises, Under what conditions will a simulation serve as an explanation of or a guide to the explanation of the behavior simulated? This question has received some philosophical treatment. The answers provided by at least one influential writer are worthy of our attention because the results of this chapter turn out to suggest that highly plausible conditions of the sort he advocates, and that many would take to be indubitably obvious necessary constraints on successful simulation, may in fact be serious obstacles to its success.
In *Psychological Explanation* J. A. Fodor argues that in order for a machine simulation to provide an adequate explanation of the behavior of an organism, it must satisfy at least two conditions: (1) the machine and the organism must be weakly equivalent, that is, the behavioral repertoire of the machine must be identical with the behavioral repertoire of the organism; (2) the machine and the organism must be strongly equivalent, that is, the processes upon which the behavior of the machine is contingent must be of the same type as the processes upon which the behavior of the organism is contingent. Fodor writes: “For an adequate simulation to be an adequate explanation it must be the case both that the behaviors available to the machine correspond to the behaviors available to the organism and that the processes whereby the machine produces behavior simulate the processes whereby the organism does.”

As Fodor notes, the requirement of weak equivalence is simply a variant on a standardly imposed adequacy condition for scientific theories. An adequate theory cannot merely be compatible with the known data, for, as noted in connection with Becker’s economic theory of behavior, there are an indefinite number of theories compatible with a finite amount of data; an adequate theory must also generate relevant counterfactuals about its subject’s behavior. Similarly, a machine must not merely duplicate the actual behavior of an organism in order to simulate it adequately; it must be able to duplicate what the organism could do—that is, counterfactuals true of the organism must be true of the machine or be paired with parallel counterfactuals true of the machine. This is the force of the requirement of weak equivalence: that behavioral repertoires, and not just actual behavior manifested by machine and organism, be qualitatively the same under some relevant descriptions. This requirement can, of course, be satisfied by a machine that evinces only artificial intelligence, and does not simulate any actual behavior or repertoire of behaviors. For simulation, it is required in addition that there be a functional equivalence between the internal processes of the machine and the internal processes of the organism that are causally responsible for their respective, equivalent repertoires. Thus, if the program for a simulator gives the states of its information-processing in a flow chart or a machine table, then there must be a one-to-one correspondence between the stages described in the program’s flow chart and the cognitive stages whereby the organism generated its output.

Since we can produce an indefinitely large number of different programs for any given repertoire, this requirement in effect makes our guide to the discovery of explanatory simulations reports by the organism of its own internal states as it generates the behavior simulated, for at present we have no other source of information about the stages which constrain our program choice. Thus, for instance, when simulators attempted to model problem-solving in logic or geometry, they monitored the verbal reports of subjects
about the stages in their problem-solving and the reasons which led them to make certain decisions about how to proceed. Then they wrote a program for a computer that enabled it to solve problems of the same kind in similar times, and inferred that the program constituted at least the sketch of an explanation of the human behavior simulated and that, in this regard at least, human behavior is explicable in terms of generalizations drawn from information science. Unfortunately, success in simulation has been limited to a narrow range of behavior, and to a fairly simple level of complexity as well, a range far narrower and simpler than that to which the relatively unconstrained methods of artificial intelligence have been applicable. The reason is in part that the requirement of strong equivalence has limited the alternative programs to those reflecting the kinds that humans employ to characterize the factors in their own internal decision processes—beliefs that some propositions are true and desires that others come true. If beliefs and desires are not natural kinds, and computer simulation is constrained to find functional equivalents for them which will generate the same behavioral repertoires in computers as in the subjects simulated, then such simulations are bound to fail. For the absence of laws systematically linking reasons and actions precludes the existence of a behavioral repertoire generated by a causally homogeneous set of internal states of belief and desire. Accordingly, any attempt to generate a repertoire that characterizes its members as examples of a type of action is bound to fail, since that repertoire is not produced in the human case through the exemplification of a small number of natural kinds by the inner states of the organism. We know that our descriptions of the internal states of a computer are couched in natural-kind terms because such descriptions figure in the physical laws that are employed to build computers, and that are confirmed in their operation. It is because we already have a well-confirmed theory to explain the interrelation of such states and their consequences for the output of the machine that simulation is an appealing strategy for explaining hitherto unexplained phenomena like human behavior. But the criteria of weak and strong simulation in effect oblige us to abandon this advantage by requiring us to characterize the repertoire of the machine in terms drawn from the everyday description of behavior as actions of various sorts, and to limit our choice of programs to ones which are in effect functional equivalents of \( L \) and its substitution instances.

If human behavioral repertoires are individuated in terms of types of actions that they exemplify, then the requirement of identity of repertoires condemns computer simulation to the impossible task of finding a causally homogeneous explanation for a heterogeneous class of effects. And if the determinants of these classes are assumed to be beliefs and desires, the requirement of strong equivalence, of identity of processes, condemns computer simulation to search for uniform relations between causally hetero-
geneous internal states of computers. The upshot is that so long as we retain the commonsense view about the nature of human behavior and its causes, the recalcitrance of most of it to simulation will reflect the imposition of these two apparently unexceptionable constraints. In the attachment of simulators to the common assumption about human behavior and to these adequacy conditions on simulation, we have a far more plausible explanation for the tremendous difficulty involved in providing computer simulations than in the metaphysical differences in kind between men and machines. Of course we need not surrender our adequacy conditions if instead we give up the common assumption that behavior is produced by the joint operation of belief and desire. Naturally, the application of the criteria will turn on the provision of other ways of individuating the behavioral repertoires of humans and the internal determinants of the members of these repertoires. But the descriptive language already available to computer science may provide these means. Since we know how to individuate internal states and outputs of computers in a terminology that does reflect natural kinds—the natural kinds of computer science—it follows from the assumption that we can simulate human behavior, that their repertoires and the internal states of humans must be describable in terminology drawn from the language of computer science.

In other words, if we erect as an adequacy criterion for descriptions of human behavior and its causes the requirement that these descriptions employ kind-terminology already entrenched in the laws of computer science, then the events, states, and conditions thus described will stand a chance of subsumption under natural laws which on our diagnosis they cannot now do. In effect, this is to turn the adequacy conditions upside down, treating them as constraints on the description of human internal states and behavior instead of constraints on the description of computer behavior and its determinants. Under this assumption, which preserves the plausibility of the requirements of strong and weak equivalence while freeing them from the dead weight of a fruitless descriptive typology, the recalcitrance of human behavior to computer simulation will clearly be seen to reflect empirical, instead of conceptual, obstacles to this sort of explanation of human behavior. Thus, suppose a machine were built that provided a perfect simulation of human behavior—that is, a robot all of whose behavior was indistinguishable from that of a human's by some such test as Turing's. Or even better, suppose a robot were built with the same physiognomy and appearance as well as behavior of a human being, and was in fact not detected by anyone but his designers to be a robot. In such a case the designers might think themselves to have in their design for the hardware and in their program at least the next best thing to an explanation of the behavior simulated. Given our adequacy conditions, it would follow from this perfect simulation that humans did behave in accordance with the program of this robot and that
their internal and external states were properly described and correctly explained by appeal to the language and directives of this program. Of course, this conclusion will not follow unless all human behavior is duplicated. For unless complete duplication were attained, the range of behavior unduplicated would allow the possibility of differences in repertoire that could be explained by attributing different internal processes and programs to the man and the machine.

The upshot of recognizing the nonqualitative, spatiotemporally restricted character of the terms in which human cognitive states are described is that computer simulators should be encouraged simply to maximize identity of repertoires between man and machines, where these repertoires are to be described in the language adapted to the description of computer behavior, without concerning themselves with the question of whether the programs they hit upon reflect any introspective data currently in hand about human cognitive function. And the reason for this is that what is in hand is pretty useless; and successful simulation suggests that the machine program which generates it represents the best evidence available for the actual character of the cognitive states behind the behavior successfully simulated. The degree to which we can satisfy the requirement of weak equivalence is our only measure of the extent to which we are also satisfying the requirement of strong equivalence in machine simulation.

I have argued that desires, beliefs, and actions are not natural kinds, in the last chapter on the grounds that this claim represents the conclusion of an inference to the best explanation of the failure to discover any laws of human behavior, and in this chapter, independently, by appeal to empirical considerations reflected in biological theory. In the light of this conclusion I have examined two research programs for the explanation of human behavior, both to assess their prospects and to illustrate the significance of the reasoning behind my claims about reasons, actions, and natural kinds. In the next chapter the consequences of these claims for the social sciences in general (and not just for these two branches of them) will be examined in more detail, but we may conclude this chapter by turning to a philosophical controversy of the most basic sort, one already touched on at the outset of this chapter. Our conclusions, it turns out, lend new weight to one side of the dispute.

The issue is one often broached in connection with claims like the ones endorsed immediately above about the prospects for simulating human behavior by computers, and the conclusions about mentality drawn from such prospects. It seems an indubitable fact that mental states have content; in the case of desires and beliefs, they have propositional content. We believe that ..., where a proposition can fill the ellipses, and we desire that some state of affairs be true. States like belief and desire are accordingly called propositional attitudes, and they are described as reflecting the property of
intensionality. Now one test of intensionality is that statements attributing intentional properties to particular items, like psychological states to human beings, are incapable of absorption into the extensional logical apparatus that seems to suffice for analyzing and regimenting the statements and inferences of mathematics and the natural sciences. A simple illustration of this recalcitrance of intentional discourse is provided by the following seemingly valid inference from apparently true premises to a presumably false conclusion:

Lady Astor desired to sail on the largest ship afloat in 1912.
The largest ship afloat in 1912 was identical to the ship that struck an iceberg, sank, and caused Lady Astor’s death.

Therefore,
Lady Astor desired to sail on the ship that struck an iceberg, sank, and caused her death.

Since the conclusion is false, and the inference-form unexceptional, the premises must be ill-formed for purposes of logical manipulation. The trouble is usually diagnosed by noting the intensional character of the first (and sometimes also the second) premise, and such premises reflecting propositional attitudes are excluded from theories which, like those of physics, are regimented in accordance with extensional logic. Now, if computers, or brains, for example, are purely physical systems, and no physical description of their properties and states will generate invalid inferences of the sort illustrated, then they and all their activities can be completely described in propositions that make no recourse to intensionality. In consequence, since psychological states do seem to manifest intensionality, the attribution of such states to computers (which the simulator hopes to make) or to brains (which the neurophysiologist hopes to make) will require an analysis of intentionality that enables us either to attribute it to computers or brains or to translate intensional statements into ones that are not intensional, that do not generate invalid inferences of the form illustrated. Now the first alternative is counterproductive of systematic regimentation of scientific theories, for it would extend the range of theories not amenable to treatment by the resources of extensional logic. But the second seems unlikely of fulfillment. None of the large number of attempts to analyze belief, desire, or any of the other propositional attitudes has eliminated their intensional features, although some have translated one sort of intensional feature into another. Indeed, it has been a widely held philosophical thesis associated, since the nineteenth century, with the name Brentano that intentionality is an essential feature of psychological attitudes, and that therefore any science of such attitudes, like psychology and the other social sciences, must as a matter of logic be autonomous and irreducible to the natural, extensional sciences.
But this conclusion, frequently cited to insure the insulation and irreducibility of mentalistic to physicalistic language, is a double-edged sword. Although no physicalistically motivated account of mental states and events has successfully analyzed away the intentional element of such states and events, the antiphysicalist has been able to do no more than to assert that mental phenomena have this property of intentionality. They have been equally unable to explain why it is that intentionality is limited to this range of phenomena or what this intentionality consists in. It is no answer to either of these questions merely to say that propositional attitudes are intentional because they are conscious states of agents or are related to such states, or that they are intentional because they are psychological or mental, nor will it do to explain their intentionality in terms of their intensionality, in the invalidity of otherwise unexceptionable arguments in which they might figure. The first answer is unsatisfactory because the question, Why are psychological or mental states intentional? is the same question about the intentionality of propositional attitudes all over again. The second answer is similarly unavailing. In asking what intentionality consists in, we want to know why it is intensional, why it generates invalidity in the way it does, not merely that it does generate such anomalies for logic. Of course, the explanation of intentionality or intensionality may not excite the interests of those who cite it to establish the autonomy of the sciences of intentional phenomena. But the existence of such an explanatory dead end for features of so complex a system as a human being is plainly intolerable for the empiricist. Thus, W. V. O. Quine writes, "One may accept the Brentano thesis either as showing the indispensability of intentional idioms and the importance of an autonomous science of intention, or as showing the baselessness of intentional idioms and the emptiness of a science of intention. My attitude, unlike Brentano's, is the second. To accept intentionality at face value is . . . to postulate translation relations as somehow objectively valid though indeterminate in principle relative to the totality of speech dispositions. Such postulation promises little gain in scientific insight if there is no better ground for it than that the supposed translation relations are presupposed by the vernacular of semantics and intention." Quine's argument here is similar to that advanced above in connection with the adequacy conditions of computer simulation, and in fact is but a restriction of that argument from behavior in general to linguistic behavior in particular. Requiring computers to satisfy intentional descriptions which are themselves inexplicable promises no more gain in scientific insight than requiring speakers to satisfy the same intentional properties when they decode each other's linguistic signals. Because of the scientific sterility of intentional notions, Quine continues, "if we are limning the true and ultimate structure of reality, the canonical scheme for us is the austere scheme that knows no . . . propositional attitudes but only the physical constitution and behavior of organisms. . . . If we are venturing
to formulate the fundamental laws of a branch of science, however tentative, this austere idiom is again likely to be the one that suits. 16

How is this dispute to be settled? Brentano and his party have the certainty that comes with direct introspective access to the existence of these intentionally characterized states. Quine and his cohorts have the admission of the inexplicability and therefore systematic sterility of intentionality, as well as its recalcitrance to regimentation in a logic that seems to suffice for all other scientific purposes. Is the argument a standoff? Do the conclusions of this chapter—that for considerations independent of issues controversial between the two parties, propositional attitudes are not natural kinds and cannot figure in laws—tilt the balance in Quine's direction? I should say the answer is yes. For they show, minimally, that Brentano's science of intention will be a science without laws, and therefore no science at all, strictly so called. More arguably, these conclusions and their applications to behaviorism and computer simulation suggest that there may well be natural kinds under which each and every particular propositional attitude falls which are clearly not themselves intentional. Most importantly, the considerations of this chapter help the empiricist and the physicalist explain away the phenomenon of intentionality as without systematic scientific significance in the description of human beings and their behavior; and they do so without appealing to philosophical theories controversial between Brentano's exponents and his opponents. The physical irreducibility of propositional attitudes and the intentionalist's own inability to explain why psychological states are intentional both hinge on the fact that such states when intentionally characterized, involve tacit reference to a spatiotemporal particular (the species *Homo sapiens*) and so cannot figure in synthetic general statements of a lawlike kind. But if the explanation of the apparently general fact that all psychological states or propositional attitudes are intentional requires the citation of other general laws from which the explanandum is derivable, then plainly no such explanation will ever be available. Singular statements cannot be derived from general ones. The behaviorist's persistent inability to analyze intentional states into statements about their extensionally characterized behavioral consequences is explained by the fact that intentional description of the behavior's causes precludes the expression of nonaccidental general statements relating these causes and their effects. For these causes are not characterizable in purely qualitative predicates. Indeed, on our hypothesis, the discovery of a true synthetic statement relating an intentionally characterized psychological kind of state to a nonintensionally characterized kind of behavior would be nothing short of a miracle. This is why behavioral analyses of such concepts as "belief that . . ." or "desire that . . .," or more specifically, "belief that it is raining on 31 August 1996 in the center of Salzburg, Austria" are invariably either open to obvious counterexample, or closed to empirical overthrow altogether. This failure, paralleled in the
similar impossibility of giving a true general neurophysiological characterization of psychological states, is what their intentionality consists in. Because we cannot independently characterize them in consequence of the systematically undescribable connection which each and every psychological state has to some particular brain state and to some particular behavior (within or at the periphery of the body), we are reduced to characterizing them in terms of independent states of affairs (described in the propositions they "contain") with which the intentional states bear no invariable causal connections whatever. Because they bear no such relations, they generate invalid arguments, are irreducibly nonphysical and nonbehavioral, and their attribution to anything whatever is utterly inexplicable. All of this follows from the arguments of this chapter, and suggests, with Quine, not the autonomy of a science composed of such statements, but its impossibility.