that it is one that science can only refine, not replace.

3. Ullin Place invokes conceptual analysis in defense of the behaviorist position, which holds that folk psychology is a language rather than a theory, but one that, for a variety of reasons, is unsuitable for scientific purposes.

4. Graham Richards defends what may be thought of as the Wittgensteinian position, which denies that folk psychology is a theory, but differs from the behaviorist in thinking that no alternative scientific language is possible that could act as a substitute.

The section concludes with a discussion by Elizabeth Valentine of the issues raised in the four preceding papers.

Ullin T. Place

The Falsity of Folk Theories:
Implications for Psychology and Philosophy

Nick Chater

Mike Oaksford

We assume that commonsense knowledge, including our commonsense understanding of human behavior, is organized into theories. After considering certain difficulties in finding out about these theories, we argue that folk theories are analogous to bad scientific theories, and that the ontology of common sense is on a par with epicycles or the *yin* and *yang*. That is, folk theories are false, and their entities that they postulate do not exist. We consider various possible replies to our arguments, and suggest that the underlying reason that folk theories are bad science is that common sense must deal with matters that do not yield to scientific analysis. We draw out some philosophical and psychological implications of our position.

INTRODUCTION

It has become increasingly popular to assume that everyday, commonsense knowledge is organized into theories. In philosophy, it has become standard to conceive of our commonsense beliefs about the mind as a theory: folk psychology (e.g., Fodor, 1987; Stich, 1983). In developmental psychology there has also been much discussion of the child's theory of mind (e.g., Leslie, 1987; Perner, 1991; Wellman, 1990), and more generally there has been an emphasis on children as theorizers (e.g.,

Carey, 1988; Karmiloff-Smith, 1988). In social psychology, there has been much study of "lay theories" of a wide variety of domains (e.g., Furnham, 1987), and everyday thought has been compared extensively to scientific theorizing (Nisbett & Ross, 1980). The psychology of concepts has increasingly stressed that concepts are theoretically embedded (e.g., Medin & Wattenmaker, 1987; Murphy & Medin, 1985). In artificial intelligence, commonsense ideas concerning the everyday world have been formalized as axiomatic theories, where inference is supported by formal logical methods, or some variant (Charniak & McDermott, 1985; McCarthy & Hayes, 1969).

Most researchers who view knowledge as organized into folk theories shy away from trying to give a precise account of exactly what a theory is. Viewing theories as made of knowledge does not amount to a precise and specific doctrine, it seems, but rather to an emphasis on an analogy between the structure of everyday knowledge in individual cognitive agents, and the structure of knowledge in science, from which talk of theories is borrowed. We believe that this analogy is a valuable one, and that when taken seriously it yields significant conclusions for folk psychology and cognitive science.

The structure of the paper is as follows. We first consider the problem of how folk theories can be

known, stressing that natural language does not give direct access to them. Nonetheless, we suggest that it is possible to judge folk theories in broad terms by looking at the explanations to which they give rise, and we present a range of arguments to show that these explanations fair poorly when judged by the standards applied to explanation in science. We conclude that folk theories are false. We then consider the status of the ontologies of our putative folk theories and argue that from the point of view of scientific enquiry, they should naturally be viewed as on a par with terms of other false theories, such as phlogiston or epicycles. We conclude that the entities described by folk ontologies do not exist. Until this point, the discussion appears to take rather a dim view of folk theories, when compared to science; we attempt to correct this impression by stressing the differing roles of scientific and folk theorizing. Finally, we briefly draw out some philosophical and psychological implications of our position.

WHAT DO WE KNOW ABOUT FOLK THEORIES?

One of the most pressing and problematic points of difference between folk and scientific theories is that folk theories are not explicitly articulated for public consumption, but appear to be buried in the individual's cognitive innards. This means that the folk theories that guide thought and action must somehow be inferred from what agents do or say.

Naively, we might hope that speakers can simply tell us what their underlying theories are, so that if, for example, people tell us that they believe that people usually act in their own best interests, then this is likely to be part of their underlying folk theory of human behavior.

A first difficulty with this naive picture is that social psychologists have persistently found that people's reports of their underlying beliefs do not readily cohere into a single picture of the world, but often reflect a wide variety of conflicting points of view (e.g., Potter & Wetherall, 1987). This has led to the view that the ideas expressed in linguistic behavior are better thought of as constructed for a specific purpose, dependent on the particular occasion, rather than as direct reflections of an underlying fund of knowledge.

A second difficulty is that it is not clear to what extent we are able to verbalize commonsense knowledge at all. This point has been stressed

across a range of disciplines. For example, the psychology of memory has stressed the importance of implicit information, which cannot be verbalized (Schacter, 1987). Coming from a very different point of view, ethnomethodologists have stressed that shared commonsense assumptions tend to be inaccessible to individuals; ethnomethodological investigation attempts to discover such assumptions by trying to violate them, rather than relying on introspective reports (Garfinkel, 1964; see Place, 1992, for discussion). A final example is given by philosophical enquiry, in which (among other things) intuitions concerning meaning, good and evil, or beauty are taken as starting points for constructing theories in the philosophy of language, ethics, or aesthetics. The very fact that developing philosophical theories that capture such intuitions is so extraordinarily difficult is a testament to the fact that any folk theories underlying these intuitions are not readily available to the investigator.

These problems in articulating our folk theories of the world have been an important stumbling block for artificial intelligence. It has proved to be extremely difficult to specify the knowledge underlying the most mundane aspects of everyday thought. Specifying such knowledge is, of course, a prerequisite for putting such knowledge into a machine, according to standard artificial intelligence methodology. Attempts to formalize apparently constrained aspects of common sense, such as the naive physics of the behavior of fluids, have been instructive. First, it is not possible simply to take verbal descriptions of what people say is the relevant knowledge and embody this in logical axioms, which can be used as the basis for inference. Instead, it has been necessary to attempt to formulate extremely complex underlying theories of the ontology that people are implicitly using and to devise very complex and subtle principles concerning what people know about this ontology and how this knowledge can be used to reason successfully. Such sophistication is required to even begin to build systems that reason about such everyday matters as the spread of spilt coffee and the results of leaving a tap running (e.g., Hayes, 1978, 1984a, 1984b). Needless to say, the formalization of folk psychology and other more complex domains has scarcely even been attempted.

This work suggests that, in general, the terms of folk theories may not always have correlates in everyday natural language. But terms of folk theo-

ries will be little, if any, easier to understand even if they do happen to be expressed by the words of a natural language. For the mere existence of a natural language label goes no way at all toward explaining the meaning of the term and its relation to the rest of the folk theory. After all, it took enormous theoretical effort to make sense of intuitive notions of "weight" or "set" (in the sense of collection), which do have natural language labels. Whether this effort should be thought of as making coherent previously incoherent ideas, or simply as making explicit what was really being talked about all along is a controversial question, to which we shall return briefly below. In any event, it is clear that even if we are able to identify people's concepts with words of natural language, this does not solve the problem of specifying what these concepts are. Thus, even though we have a natural language label for "chair," "elbow," and "jazz," and have an intuitive sense of what these labels are supposed to signify, it is notoriously difficult to define (Fodor, 1981) or characterize in any way what these terms mean. If knowledge is organized into theories, then explicating such terms involves specifying the particular folk theories in which they figure; and, as we have seen, this is extremely difficult to do.

The upshot of these considerations is that, if common sense consists of folk theories, then the nature of these theories is unknown and likely to be subtle, complex, and only indirectly related to explicit verbal behavior. The problem of discovering the theories underlying commonsense thought seems, therefore, to be analogous to, for example, the problem of discovering the underlying knowledge of language that governs linguistic behavior. Linguistics uses verbal behavior (and grammaticality judgments, and the like) as the starting point for constructing theories of the underlying knowledge involved in language processing. The resulting linguistic theories are highly elaborate and sophisticated and are, of course, entirely inarticulable from the point of view of everyday speakers. It seems likely to be an equally difficult task to tease out the theories underlying commonsense thought; and the nature of such underlying theories is likely to be no more apparent to naive intuition. In particular, the ontology of folk theories cannot be assumed to be limited to the vocabulary of natural language—indeed, restriction to the ontology of natural language appears to be entirely inadequate to formalize commonsense thought, which is what drives

Hayes (1984a) to define notions such as "portal," "enclosure," "directed surface" in attempting to formalize the naive physics of fluids. It seems likely, then, that much of the ontology of folk theories may be no more captured by everyday language than are phonemes, island constraints, or traces.

ARE FOLK THEORIES GOOD SCIENCE?

We have argued that folk theories must be inferred from verbal and other behavior and are not directly accessible by, for example, verbal report. Although the details of such theories are hidden, however, it is nonetheless possible to use the verbal and other behavior to which they give rise to assess how such theories fare when considered as scientific theories. We shall concentrate on assessing the quality of commonsense explanations and assume that the scientific respectability of these explanations is a reasonable reflection of the scientific status of the underlying folk theories. We outline two arguments why folk explanations are very poor by scientific lights. The first argument compares folk and scientific ideas in domains that are well understood by science; the second, more general argument declares that the ineliminably defeasible character of folk explanation is a hallmark of bad science. Given that the underlying folk theories giving rise to these explanations are hidden, it is just possible, of course, that these theories are actually consistent, coherent, well-confirmed and scientifically respectable accounts but that, for some reason, they give rise to verbal explanations that are confused, completely *ad hoc*, and readily succumb to counterexamples. This possibility is sufficiently bizarre, and lacking in any evidential support, that we shall not consider it further, and shall simply judge folk theories by folk explanations.

Let us turn to our two lines of argument.

Where Common Sense and Science Compete

An obvious way to assess how folk theories compare with science is to consider domains that can be described in both folk and scientific terms. In such domains, it may be possible to assess the quality of commonsense thought by directly comparing it against the corresponding scientific account. We shall concentrate on the physical sciences in the

examples below, leaving aside, for the present, the more controversial case of folk psychological explanation.

The development of physics, chemistry, biology, medicine, and so on no doubt originate in folk intuitions. However, in modern accounts of the phenomena of these areas, little or no vestige of this heritage remains. Rather than supplementing and regimenting folk intuitions about dynamics, reactions, the basis of life, and the cure of disease, modern theories have totally discredited and supplanted these accounts.

There are numerous illustrative examples of folk accounts, which even had a measure of scientific respectability but which now appear completely unfounded. In physics, the motion of an artillery shell was commonly conceived of as consisting of a straight line motion along the line of sight of the gun barrel followed by a vertical descent. In chemistry, it was commonly believed that there are few constraints on the ability of substances to transmute from one form to another, which motivated the search for the "philosopher's stone," which would turn base metals into gold. Even after the development of scientific chemistry, our folk taxonomy of substances has little or nothing to do with the periodic table and molecular composition. In biology, the spontaneous generation of life from decaying substances was a prevalent view as recently as the seventeenth and eighteenth centuries. It was thought that flies arose spontaneously from feces and even that signets emerged from rotting logs. Equally, folk accounts of medicine, some of which go under the banner of "alternative" medicine, do not provide a foundation for, but appear completely at variance with, modern Western medicine. For example, the effectiveness of acupuncture is usually justified as bringing the life forces of *yin* and *yang* into balance.

These examples show that commonsense conceptions of the world, while they may provide a historically important starting point for scientific investigation, are typically superseded and, crucially, dismissed as false. In particular, the intuitive notions of "impulse" and "natural place" that underwrote naive understandings of ballistics are no longer considered to make sense. Similarly, the philosopher's stone and the alchemical conception of transmutation are not thought to refer to any aspect of the real world. Equally, modern biology does not countenance the possibility of spontaneous generation. Modern Western medicine claims

that the postulates of "alternative" accounts, such as "life force," "yin," and "yang" do not exist. Notice that modern science does not simply contend that the categories of folk science happen to have no members. Since the entire standpoint of folk theory is rejected, it becomes difficult or impossible to conceive of what it would be to encounter an example of such putative categories—the categories are simply rejected wholesale as completely nonsensical. (Of course, some of the vocabulary of false naive theories may survive, for example, "impulse" above, but be construed very differently.)

Backing up such historical considerations are experimental studies of folk beliefs about scientific matters. Modern students of physics are prone to reveal a bizarre conception of basic physical principles (McCloskey, 1983). For example, when asked to describe the trajectory along which a ball will travel after being released from constrained spiral motion, a common response is that it continues in a spiral motion, rather than traveling in a straight line (Kaiser, McCloskey, & Proffitt, 1986). Furthermore, such misconceptions are remarkably difficult to change by instruction (Carey, 1985, 1986; Gentner & Stevens, 1983; West & Pines, 1985; see Kuhn, 1989, for discussion). It is remarkable that we are able to navigate our way through a complex world so successfully, when our explicitly held beliefs about its structure seem to be consistently and dramatically off-target.

Now, if folk theories appear to be bad science in domains that are scientifically well understood, there seems little reason to suppose they will fare better in domains that are scientifically poorly understood. It seems reasonable to assume that domains that have resisted scientific analysis are likely to be especially complex; hence, in these domains, folk theories are even less likely to provide a scientifically respectable analysis. In particular, folk psychology, along with folk economics, folk sociology, and folk theories concerned with tables, cars, music, and shopping, are all likely to prove to be scientifically ill-founded.

The Defeasibility of Commonsense Generalizations

In the previous section, we did not consider folk psychology directly, but drew morals from the firmer ground of explanation of physical phenomena. In this section, by contrast, we start by consid-

ering folk psychological explanation and extend our conclusions to folk theories in general. Consider this schematic folk psychological generalization:

1. If you desire D and have the belief B that action A will lead to D, then you will perform action A.

This can be filled out as, for example, a useful rule for parents: "If a child desires ice cream and has the belief that tidying her room will lead to her being given ice cream, then she will tidy her room." However, bitter experience indicates that this, like other specific instances of the schema, admits of many counterexamples. For example, the generalization will not hold if the child believes that ice cream will be forthcoming in any case, because her parents are weak-willed. Equally, she may believe that the room is so untidy that it is not worth the effort, that there is a fierce dog in the room, that her big sister will take the ice cream anyway, or that there is an alternative, more desirable, action available, such as watching a favorite television program, going swimming, and so on. For this specific instantiation of the generalization, it is clear that there will be no way of ruling out all these possibilities one by one. There cannot be a clause ruling out the possibility of swimming, one for watching television, one for each of the possible dangers that might be encountered in the bedroom, and so on. The only hope of ruling out such possibilities without specifying an exhaustive and presumably indefinitely long list of exceptions is to attempt to save the folk psychological generalization at the schematic level (Chater & Oaksford, 1990; Oaksford & Chater, 1991).

However, at the schematic level, too, it is hard to imagine how the appropriate modification can be achieved. One possibility, which takes account of the counterexamples above, might be that the generalization should read:

2. If you desire D and you have no other more pressing desire D' and have the belief B that action A will lead to D, and that it will not lead to any unwelcome consequences, and that D will not be satisfied if A is not performed, and that you are able to exploit D if it occurs, then you will perform action A.

But this, of course, succumbs to further counterexamples. There may well be a less pressing desire

D', which can be achieved by action A', which is less arduous than action A. In this case, it may be judged not worth going to the extra trouble of performing A, even though the desire D is the most pressing. Further, it must be *possible* to perform A, and the agent must believe that A can be performed (the child will not attempt to tidy the room if she believes that the door is locked and that she cannot get a key). Clearly, further elaboration of this generalization by adding extra clauses of the same kind will not help because further counterexamples can always be generated.

The fact that folk psychological generalizations succumb so readily to counterexamples is recognized in that they are usually stated as holding *ceteris paribus* or "everything else being equal." That is, the situations in the counterexamples described above are viewed as situations in which all other things are not equal. Of course, the use of such a locution does not remove the problem of counterexamples, but simply changes the problem from that of adding conditions to refine the original generalization to that of specifying the conditions under which all other things are equal. This is, of course, simply the original problem in a different guise.

If it is impossible to reconstruct folk psychological generalizations so that they are true, then surely folk psychology must be rejected as a false account of human behavior. In consequence, the postulates of the theory—beliefs, desires, and their kin—should be treated as incoherent. Beliefs and desires will not figure in a scientific account of mind any more than the *yin* and *yang* figure in modern Western medicine. Putting the point bluntly, folk psychology is false and beliefs and desires do not exist.

While this kind of argument is reasonably familiar with regard to folk psychology (see Fodor, 1991; Schiffer, 1987, 1991), which has been the center of intense debate (e.g., Churchland, 1986; Fodor, 1987; Kitcher, 1984), parallel arguments appear to be equally persuasive with regard to other folk theories, to which much less philosophical attention has been devoted.

Consider the falsity of the following commonsense generalizations: "All chairs have legs," "All birds can fly," "If you turn the key, the car starts." Armchairs do not have legs, swivel chairs have a central column; ostriches, penguins, and injured birds cannot fly; if the battery is dead the car will not start. Perhaps these generalizations are false

because they are formulated with insufficient precision—perhaps not *all* birds can fly, but all uninjured garden birds can fly. But what about very young birds, very old birds, birds tangled up in nets, birds in extremely cold weather or high winds, birds with clipped wings, and so on. Refining further, we may say that "all birds can fly" means that every bird will have or has had the ability to fly at some time in its life—according to this reading, very young, very old, and entangled birds count as flyers; and transient meteorological conditions may be ignored; and perhaps having clipped wings counts as an injury. What, then, of very young abandoned birds, destined to starve before they learn to fly; genetically abnormal birds; caged birds, and so on. The terms used in these attempts to refine the generalization themselves equally require refinement in order to save the original generalization from counterexamples. What counts as a garden bird—a turkey in a run at the end of the garden? an ostrich in an African garden? There seems to be no end to this refinement—every term adduced in refining the original generalization itself requires refinement.

This phenomenon has been given many different labels in different areas of cognitive theory, from cognitive psychology and philosophy to artificial intelligence. Folk theories are said to: be context-sensitive (Barsalou, 1987); hold only relative to some background conditions (Barwise & Perry, 1983); be defeasible (Minsky, 1975), admit exceptions (Holland, Holyoak, Nisbett, & Thagard, 1986), lack generality (Goodman, 1983), have intention-relative categories (Winograd & Flores, 1986). These are many ways of saying that every commonsense generalization, just like the generalizations of folk psychology, succumb to endless counterexamples.

We have focused above on the fact that folk generalizations have counterexamples. On the reasonably standard assumption that every good scientific law is without exceptions, this immediately implies that folk theories are bad science. But it has been argued that scientific laws are quite generally defeasible but not thereby false. An extreme version of this view has been advocated by Cartwright (1983), who argues that the "phenomenological" laws of science, which are defeasible, are the only candidates for truth and that "deep" putatively exceptionless laws should be rejected as false.

Even independently of the inference from counterexamples to falsehood, however, folk theories,

when judged as scientific theories, are woefully inadequate—they correspond to bad science rather than good science. This inadequacy has a number of aspects, including the inchoate poorly articulated nature of such theories, internal inconsistency, the *ad hoc* character of explanation, lack of predictive power, and so on. These properties are evident in the commonsense explanations considered above.

So our reaction of folk theories does not presuppose that the laws of a good scientific theory do not admit some exceptions. We merely require that there be some distinction between good and bad science and that common sense falls into the latter category. In view of this, the discussion could stop here. However, while our arguments do not hinge on the issue of defeasibility, we actually believe that it is central to a proper understanding of the distinction between good theory and bad theory. In particular, we view the abundance of exceptions to the laws of common sense as the diagnosis for its other ailments. An abundance of exceptions goes hand in hand with an abundance of predictive and explanatory failures, the invocation of *ad hoc* rules to account for these exceptions, and an inability to retain theoretical consistency in the face of endless counterexamples. Hence, although for the purposes of our argument we need not be committed to drawing a distinction between the defeasibility of commonsense generalizations and good scientific laws, we actually believe maintaining such a distinction to be very important.

In any case, the defeasibility of scientific laws does not offer a means of maintaining the truth of folk theories. Whether or not defeasibility can ever be entirely eliminated within scientific theories, it is uncontroversial that defeasibility should be *minimized*. The degree of defeasibility (in conjunction with other factors such as breadth of coverage and simplicity) is a crucial measure of theoretical adequacy. As we have already argued, the generalizations of folk theories are defeasible through and through. So on this score, folk theories will always be ranked at the bottom.

In this section, we have argued that folk theories are false. We now argue that this means that the entities of folk theories do not exist.

FOLK ENTITIES DO NOT EXIST

If common sense is organized into theories, then commonsense categories correspond to the mean-

ing of theoretical terms. If folk theories are bad science, then, *prima facie* at least, it seems that their ontologies should also be rejected.

After all, because we assume that modern chemistry is true, we assume that oxygen is ontologically respectable—what it refers to is determined by chemistry. On the other hand, since we reject alchemy, we assume that "phlogiston" does not apply to anything; it is a term without reference. That is, terms of the internal theories underlying commonsense knowledge are analogous to phlogiston rather than to oxygen. In plain terms, the upshot may be stated: The referents of the terms of folk theories do not exist.

If this formulation seems shocking (and we have found in presentations that it certainly does!), we recommend an alternative: that folk ontologies cannot be used as the basis for scientific explanation. This leaves open the possibility of making sense of some notion of existence in some extrascientific sense. Given the problems involved in making sense of just one kind of existence, the postulation that there are two or even more kinds does not strike us as attractive. We shall continue to use the shocking formulation below.

The conclusion that folk entities do not exist (or its milder variant) appears to be too rapid, however. It may be objected that surely the adoption of a new scientific theory cannot automatically mean the wholesale rejection of the ontology of the previous theory. After all, the vocabulary of one theory is typically largely preserved in subsequent theories. During a scientific revolution, when one theory replaces another, many of the terms in the discredited theory will be preserved in the new theory. For example, modern relativistic physics retains mass, momentum, and so on from Newtonian mechanics, although rejecting such constructs as the luminiferous aether (the medium through which light waves were supposed to travel). Or consider the stability of the term "electron" over the vagaries of the development of twentieth-century physics (Hacking, 1983). Such examples suggest that some terms of false theories may refer after all. It might be argued that "table" and "chair" may be more like "mass" and "momentum" than "epicycle" and "phlogiston." Therefore, it may seem entirely plausible that a significant fragment of folk ontology may exist despite the falsity of the folk theories in which they are embedded.

One version of this position is that a term may refer not because the theory in which it currently plays a role is true, but because some future theory in which it will one day figure is true. This line appears to be advocated by Putnam (1975), who argues that "gold" was a referential term even before the chemical composition of gold was known, and that this coherence is underwritten by the truth of modern chemistry. This suggestion presupposes that a theoretical term may continue to have the same reference when the theory in which it is embedded changes; this is a controversial thesis. Moreover, adversion to a future true theory is of no avail in attempting to maintain the coherence of folk categories, since it seems, to put it mildly, unlikely that "table," "chair," and "eating ice cream" will feature in any future scientific theory.

There is, however, a more radical way of retaining ontology and rejecting theory: by denying that the coherence of terms is dependent on the truth of a theory in which they may be embedded. This position, entity realism, views entities as prior to theories about them. If it is denied that ontology is determined by an embedding theory, some other account of how ontologies are fixed is required. Two possibilities have been advanced, one that applies specifically to *biological* categories and can offer a defense of folk ontologies if they can be treated in the same way; and one that applies more generally. Very roughly, the first approach individuates entities *historically* and the second individuates entities by their *effects*. We now consider these in turn and argue that they do not change our conclusion that, like phlogiston and epicycles, the entities of folk theories do not exist.

Individuation by History

Millikan (1986) aims to explain what makes biological categories coherent, without assuming that coherence must be guaranteed by embedding in a true theory, for the now familiar reason that biological generalizations, like folk generalizations, typically have counterexamples. She notes that generalizations about, for example, hearts, like the generalizations about birds, tables, and so on that we considered earlier, seem to admit of countless exceptions: "A heart…may be large or small (elephant or mouse), three-chambered or four-chambered etc., and it may *also* be diseased or malformed or excised from the body that once contained it, hence liable to pump blood" (Millikan, 1986).

For Millikan, counterexamples to biological generalizations pose no threat to the coherence of biological categories, since coherence is judged by other historical–functional standards to which we shall turn presently. If an appropriate alternative basis for ontological coherence can be found for biology, this might be applied to folk theories—indeed, Millikan suggests that folk psychological terms should be construed as biological categories.

Millikan's approach is complex, but can be illustrated by example:

> A heart…falls in the category *heart*, first, because it was produced by mechanisms that have proliferated during their evolutionary history in part because they were producing items which managed to circulate blood efficiently in the species that contained them, thus aiding the proliferation of that species. It is a *heart*, second, because it was produced by such mechanisms in accordance with an explanation that approximated, to some undefined degree, a Normal explanation for production of the majority of Normal hearts of that species. By a "Normal explanation" I mean the sort of explanation that historically accounted for production of the majority of Normal hearts of that species. And by a "Normal heart," I mean a heart that matches in the relevant respects the majority of hearts that, during the history of that species, managed to pump blood efficiently enough to aid survival and reproduction. (Millikan, 1986, p. 51)

This approach turns out, however, to be extremely liberal. Suppose, for example, you are skeptical about the laws of Freudian psychoanalytic theory and doubt the relationship between the failure to resolve certain conflicts that arise at specific psychosexual stages and consequent specific forms of neurosis. On the orthodox view, this would entail a similar skepticism with respect to the Freudian categories of, for example, the Oedipus complex, the id, the ego and the superego. However, according to Millikan's account, these categories can be maintained in the face of skepticism concerning the laws in which they figure. Consider how the term *superego* may be grounded in the same way as Millikan grounds heart:

> A superego…falls in the category *superego*, first, because it was produced by mechanisms that have proliferated during their evolutionary history in part because they were producing items which managed successfully to resolve psychosexual conflicts in the species that contained them, thus aiding the proliferation of that species. It is a *superego*, second, because it was produced by such mechanisms in accordance with an explana-

tion that approximated, to some undefined degree, a Normal explanation for production of the majority of Normal superegos of that species. By a "Normal explanation" I mean the sort of explanation that historically accounted for production of the majority of Normal superegos of that species. And by a "Normal superego," I mean a superego that matches in the relevant respects the majority of superegos that, during the history of that species, managed to successfully resolve sufficient psychosexual conflicts to aid survival and reproduction.

For the Freudian theorist such a line of argument might seem to be extremely appealing. For it appears to establish the coherence of the fundamental categories of Freudian theory, even though the laws of Freudian theory may not hold; similar arguments appear to establish the *yin* and *yang* as respectable entities, whose purpose is setting life forces in balance. This liberalization of the criterion for ontological commitment appears to allow the grounding of the terms of false theories of all sorts, folk psychology and other folk theories perhaps included. But if the categories of folk theories exist only in the sense that the *yin* and the *yang* exist, this sense of existence is surely too weak to be of any interest.

It might seem that there is a crucial difference between Millikan's grounding of *heart* and the apparently analogous grounding for *superego* or the *yin* and *yang*, however: that the first explanation seems to be intuitively plausible and the second does not. For example, it seems entirely plausible that the heart has proliferated because of the survival-related benefits of pumping blood; it may seem far less plausible that the superego has proliferated because of the survival-related benefits of resolving psychosexual conflicts; and it seems entirely implausible that the *yin* and *yang* have proliferated because of the survival-related benefits of balancing the life forces.

In what, however, does the intuition that these cases are different consist? Why shouldn't the superego, the *yin* and the *yang* proliferate because of their survival-related benefits? The most obvious reply is that this is because, unlike hearts, they don't exist. Yet existence is the very issue which the historical account is supposed to decide, so this appeal is illegitimate. A further suggestion may be that while 'heart' is a biologically respectable category, the superego, the *yin* and the *yang* are not. But the historical account is intended to distinguish genuine biological categories from bogus biologi-

cal categories, so this appeal too begs the question. Finally it could be suggested that 'heart' plays a role in some true biological theory, whereas the superego plays a role in an at best highly controversial theory, and the yin and the yang are parts of a radically false folk theory of medicine. Yet the historical account is intended to provide an alternative to this appeal to the truth of the embedding theory and thus cannot rely upon it. In sum, it seems that the historical account is entirely neutral between purportedly genuine and presumably bogus categories. Hence it cannot be used to demonstrate that folk ontologies have a legitimate basis, despite the falsity of folk theories.

Individuation by Effects

While Millikan's approach to individuating entities theory-independently seems too liberal, an alternative approach, developed in the philosophy of science, is motivated by the stability of theoretical terms, such as electron, in the context of dramatically changing scientific theories. As mentioned above, in the last hundred years there have been a wide variety of very different scientific accounts of the electron. Nonetheless, it seems natural to view all of these accounts as theories *of the electron.* That is, while theories have come and gone, entities seem to have remained the same.

Hacking (1983) suggests that what is common between the same entity in different theories is its effects. For example, in theories as different as the plum pudding model of the atom and contemporary particle physics, the electron is held to propel a vane in a vacuum (the "electron wind"); be sensitive to both electrical and magnetic fields (as evidenced, for example, in the Maltese cross experiment); produce, on average, a three-centimeter track in a cloud chamber, and so on. That is, although theories about electrons have changed considerably, the set of effects that electrons have been taken to explain has remained relatively stable.

However, while in some cases the set of effects that a theoretical account attempts to explain has a real basis, in other cases more than one entity or property explains what was erroneously supposed to be a set of phenomena with a coherent basis. For example, the putative negative weight of phlogiston could be used to explain both the gain in mass of materials after burning (since phlogiston was released) and the fact that hot air balloons rise (by

trapping phlogiston released from burning). However, these phenomena have very different origins. The first is explained by oxidization during burning and the second is explained by the expansion of air when heated. Because the set of phenomena that phlogiston was postulated to explain turned out to *fractionate* in just this way the preservation of the term "phlogiston" would have been rather confusing from the point of view of Priestley's account of bleaching and burning. Thus, a new term "oxygen" was used to refer to the postulated entity, which explained the gain in mass of materials after burning.

The criterion of individuation by effects appears to apply equally well to entities that are dismissed by modern science as to entities that have been accepted. That is, it could explain the stability of the term "phlogiston" over hundreds of years of chemical theorizing just as well as it explains the preservation of the term "electron" over the last hundred years. It is, therefore, entirely neutral with regard to the existence of the entities postulated. In particular, it will apply to folk terms whether they refer or not, and hence provides no defense of the coherence of folk entities.

In the previous section, we argued that folk theories are bad science; in this section, we have argued that the ontologies of folk theories are not scientifically respectable. These conclusions appear to cast common sense in a very poor light; the next section aims to correct this impression. Folk theories, while poor science, are remarkably successful at helping us make sense of and act in a world that is far too complex for scientific analysis to be tractable.

DIFFERING GOALS: THE ART OF THE SOLVABLE VERSUS COPING WITH COMPLEXITY

Despite the notional goal of explaining all aspects of the natural world, in practice, science is, to use Medawar's famous dictum, the art of the solvable. That is, scientists seek out and explore just those areas where theories can be built, tested, and applied; they shy away from areas that presently appear to be intractable to scientific methods. The ability to choose to focus on tractable matters and ignore the intractable marks an important difference between science and common sense. Folk theories must allow us to make the best possible sense of our everyday world and guide our actions as suc-

cessfully as possible; to do this they must face up to the full complexity of the everyday world, which, we suggest, science rightly prefers to avoid.

Most aspects of our everyday world are simply too complex, and too downright messy, to be the basis of science; there simply is no clear-cut theory of the behavior of everyday objects, of the changing patterns of food supply, of the nature and degree of various types of danger, or, most challenging of all, of human nature itself. From the point of view of science, each of the domains is criss-crossed by a myriad of different causal paths, most of which are little understood by science; furthermore, the complexity of these causes, and their interactions, makes such matters inherently resistent to scientific analysis. Consider, for example, the problem of predicting the likely effects of falling down the stairs: the range of relevant biological and physical factors—exact layout of the stairs, shape of body, clothing worn, etc.—make scientific study quite impossible. The scientist may choose to pick apart these multiple causes, studying gravitation, blood flow, bone strength, and so on, independently, without ever having to put all these factors back together to deal with a specific case of falling.

The agent faced with the problem of successfully coping with the baffling complexity of the everyday world has no such luxury. Folk theories must provide rough and ready advice—that here falling is dangerous and extreme care must be taken; that there, it is not so dangerous and it is safe to hurry, and so on. Our folk understanding of mind provides another good example. Human behavior appears to be generated by an extraordinarily complex mix of factors, both psychological and biological, upon which scientific psychology and biology have made only partial inroads. Yet folk theories allow us to make rough and ready assessments of how and why people behave; and when it comes to guiding action appropriately, such theories, for all their faults, are much better than nothing.

In general, then, folk theories must deal with aspects of the world that science avoids as intractable, i.e., it must deal with domains in which good science is more or less impossible, and rough and ready generalization must suffice. Thus, the fact that folk explanations do not stand up to scientific scrutiny should not be viewed as a criticism of folk theories; it is an inevitable consequence of the fact that folk theories must venture where science cannot. If we are right, then the very domains that folk

theories must cover, where scientific analysis is impossible, means that folk theories will inevitably be bad science; and that the ontology of common sense will not be scientifically respectable.

We now turn to briefly consider some of the implications of this perspective for the study of mind.

CONSEQUENCES AND CONCLUSIONS

We have argued that folk theories are false and that the entities they postulate do not exist. If we are right in equating the objects and relations of common sense with the ontology of false scientific theories, then folk ontologies do not carve nature at the joints any more than Ptolemaic astronomy. Just as the theory of epicycles was a remarkable *product* of human attempts to make sense of the astronomical world, so our everyday categories "chair," "home," and "friend" represent remarkable *products* of human attempts to understand the everyday world of artifacts, dwellings, and human relationships. The character of common sense is perhaps obscured because we are so close to its objects; but just because we make friends, build homes, and manufacture chairs does not lessen the individual and social achievement of creating the folk theories in which these terms are embedded. This is the heart of the thesis of this chapter—common sense is an *explanandum,* not an *explanans.* A science of cognition must explain the basis of our folk theories and cannot use folk theories as its foundation.

This view has significant consequences for the theory of meaning, whether for natural language or for mental states: It undercuts the project of devising a theory of reference for the terms of natural language, as this project is traditionally conceived. Typically, the problem is viewed as that of specifying some naturalistic relation between, for example, the symbol "chair" or "ice cream" and actual chairs and ice cream. There are a number of suggestions about how this "naturalization" problem can be solved. The crudest suggestion is that the appropriate relation is that the tokening of symbols is caused by encounters with their referents, or that symbol-tokenings *correlate* with such encounters. Causal theories of reference (e.g., Kripke, 1972; Plantinga, 1974; Putnam, 1975) and informational semantics (Dretske, 1981; Fodor, 1987, 1990; Stampe, 1977) have devised extremely sophisti-

cated versions of these views. But if commonsense categories are incoherent, then there are no chairs or ice cream. *A fortiori*, the tokening of the symbols "chair" and "ice cream" cannot be caused or correlated with instances of chairs and ice cream, since there are none. A causal/correlational story is no more appropriate for commonsense categories than it would be for explaining the meaning of "phlogiston" and "epicycle." Quite generally, any view that attempts to explain the meaning of commonsense terms as a relation to the corresponding category in the environment is simply not applicable—the naturalization problem for everyday folk terms cannot, in principle, be solved.

Since this argument applies just as much to mental states as to natural language, this view also poses problems for any representational theory of mind that specifies the content of mental representations in terms of commonsense, folk categories. For example, any theory that assumes that mental representations correspond to the contents of propositional attitudes is ruled out immediately, since there is no coherent folk ontology to which the contents of the attitudes can map. In particular, this constitutes a rather nonstandard attack on folk psychology as a basis for scientific psychology. Typically, folk psychology is attacked directly that it postulates entities, beliefs and desires that do not exist. According to our more general arguments to the falsity of folk theories the contents of propositional attitudes are equally in doubt. Hence folk psychology is doubly vulnerable, if folk theories are put into doubt: It is vulnerable first because the integrity of the contents of folk psychology presuppose the truth of other folk theories; and it is vulnerable because folk psychology is itself a folk theory.

A practical consequence of this additional line of attack on folk psychology is that a putative scientific psychology cannot merely reject the attitudes while retaining their contents to act as the interpretations of the representations it postulates. So practical work in knowledge representation in cognitive science and artificial intelligence, which is typically neutral with respect to the nature of the attitudes, nevertheless must be rejected, since they retain the folk ontology of tables, chairs, and so on. Such considerations apply just as much to most connectionist approaches to knowledge representation, where states of networks are interpreted in terms of folk ontologies (e.g., see papers in McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986). Notice that this applies to "dis-

tributed" as well as "localist" connectionist representation. A distributed representation of an object in a connectionist network modeling commonsense inference still relies on a featural decomposition such that the feature nodes of the network correspond to the types of which the object represented is a token. The types that provide the interpretation of the feature nodes are typically the categories of our folk ontologies. Hence on the current position interpreting features is as pressing a problem for connectionism as interpreting the predicate symbols of the knowledge representation language is for traditional AI (Christiansen & Chater, 1992, 1993).

In the light of these considerations it is perhaps not surprising that the areas of cognitive science and cognitive psychology in which most progress has been made are those that do not involve knowledge-rich inferential processes. That is, progress is only really apparent in those areas which, from a philosophical standpoint, as Davies (1992) has pointed out, are not really *cognitive* domains at all. Fodor (1983) captures the distinction very neatly. He divides the cognitive system into informationally encapsulated input (and output) "modules," on the one hand, and informationally unencapsulated central processes on the other. Central processes are explicitly identified as those involving knowledge-rich inferential processes of belief fixation and revision, i.e., precisely the processes for which our theories postulate inference over representations whose content is given in terms of our folk ontology. Fodor argues that progress in the cognitive sciences has only been and is only likely to be forthcoming for the informationally encapsulated input and output modules.

One way of viewing this diagnosis of lack of progress is that cognitive science has failed to resolve the problems that beset behaviorism. Behaviorists eschewed an introspectionist methodology and imposed rigorous strictures on psychological practice and theory. In particular, they demanded that stimulus and response be physicalistically rather than intentionally characterized. However, as Fodor (1968), Chomsky (1959), and other pioneers of cognitive science observed, in behaviorist theorizing (e.g., Skinner, 1957), such physicalistic characterizations were in practice supplanted by inadvertent use of intentional terminology—in particular, the stimulus and response were not described in the terms of physical (or other) science, but rather in terms of the experi-

menter's commonsense understanding of the task. Description of the conditioned stimulus as a pencil and the conditioned response as, say, the act of writing is description in terms of folk theories rather than physics. The cognitivist response was to attempt to legitimize this nonphysicalistic, intentional vocabulary. This presupposed that such everyday vocabulary can be naturalistically grounded as a relation between mental representations and the world. Since the domain of this intentional vocabulary is folk ontology, and since the entities of folk ontology do not exist, naturalization is impossible. "The assumption that scientific psychology can be based on the principal *product* of psychological processes—commonsense theories and commonsense ontologies—whether implicit, as in behaviorism, or explicit, as in contemporary cognitive science, is unsustainable.

REFERENCES

Barsalou, L. W. (1987). The instability of graded structures: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development* (pp. 101–140). Cambridge: Cambridge University Press.

Barwise, J., & Perry, J. (1983). *Situations and attitudes.* Cambridge, MA: MIT Press.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press.

Carey, S. (1986). Cognitive science and science education. *American Psychologist, 41,* 1123–1130.

Carey, S. (1988). Conceptual differences between children and adults. *Mind & Language, 3,* 167–181.

Cartwright, N. (1983). *How the laws of physics lie.* Oxford: Oxford University Press.

Charniak, E., & McDermott, D. (1985). *An introduction to artificial intelligence.* Reading, MA: Addison-Wesley.

Chater, N., & Oaksford, M. (1990). Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn. *Cognition, 4,* 93–107.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior. Language, 35,* 26–58.

Christiansen, M., & Chater, N. (1992). Connectionism, learning and meaning. *Connection Science, 4,* 227–252.

Christiansen, M., & Chater, N. (1993). Symbol grounding—the emperor's new theory of meaning? *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 155–160). Hillsdale, NJ: Lawrence Erlbaum.

Churchland, P. S. (1986). *Neurophilosophy.* Cambridge, MA: MIT Press.

Davies, M. (1992). Thinking persons and cognitive science. In A. Clark & R. Lutz (Eds.), *Connectionism in context* (pp. 111–122). Berlin: Springer-Verlag.

Dretske, F. (1981). *Knowledge and the flow of information.* Cambridge, MA: MIT Press.

Fodor, J. A. (1968). *Psychological explanation.* New York: Random House.

Fodor, J. A. (1981). The present status of the innateness controversy. In J. A. Fodor (Ed.), *Representations* (chap. 10, pp. 257–316). Cambridge, MA: MIT Press.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind.* Cambridge, MA: MIT Press.

Fodor, J. A. (1990). *A theory of content and other essays.* Cambridge, MA: MIT Press.

Fodor, J. A. (1991). You can fool some of the people all of the time, everything else being equal: Hedged laws and psychological explanations. *Mind, 100,* 18–34.

Furnham, A. (1987). *Lay theories: Everyday understanding of problems in the social sciences.* Oxford: Pergamon Press.

Garfinkel, H. (1964). Studies in the routine grounds of everyday activities. *Social Problems, 11,* 225–250.

Gentner, D., & Stevens, A. (Eds.). (1983). *Mental models.* Hillsdale, NJ: Erlbaum.

Goodman, N. (1983). *Fact, fiction and forecast* (4th ed.). Cambridge, MA: Harvard University Press. (Originally published 1954)

Hacking, I. (1983). *Representing and intervening.* Cambridge: Cambridge University Press.

Hayes, P. (1978). The naive physics manifesto. In D. Michie (Ed.), *Expert systems in the microelectronic age.* Edinburgh: Edinburgh University Press.

Hayes, P. (1984a). Liquids. In J. Hobbs (Ed.), *Formal theories of the commonsense world.* Hillsdale, NJ: Ablex.

Hayes, P. (1984b). The second naive physics manifesto. In J. Hobbs (Ed.), *Formal theories of the commonsense world.* Hillsdale, NJ: Ablex.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning and discovery.* Cambridge, MA: MIT Press.

Kaiser, M. K., McCloskey, M., & Proffitt, D. R. (1986). Development of intuitive theories of motion: Curvilinear motion in the absence of external forces. *Developmental Psychology, 22,* 67–71.

Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind & Language, 1,* 183–196.

Kitcher, P. (1984). In defense of intentional psychology. *Journal of Philosophy, 81,* 89–106.

Kripke, S. (1972). *Naming and necessity.* Oxford: Basil Blackwell.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96,* 674–689.

Leslie, A. M. (1987). Pretense and representation in infancy: The origins of 'theory of mind'. *Psychological Review*, *94*, 412–426.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence 4*. Edinburgh: Edinburgh University Press.

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructures of cognition. Vol. 2: Psychological & biological models*. Cambridge, MA: MIT Press.

McCloskey, M. (1983). Intuitive physics. *Scientific American*, *24*, 122–130.

Medin, D. L., & Wattenmaker, W. D. (1987). Category cohesiveness, theories and cognitive archeology. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 25–62). Cambridge: Cambridge University Press.

Millikan, R. G. (1986). Thought without laws: Cognitive science without content. *Philosophical Review*, *95*, 47–80.

Minsky, M. (1975). Frame-system theory. In R. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. Cambridge, MA: B. B. N.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.

Nisbett, R. C., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgement*. Englewood Cliffs, NJ: Prentice-Hall.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind & Language*, *6*, 1–38,

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Place, U. T. (1992). The role of the ethnomethodological experiment in the empirical investigation of social norms and its application to conceptual analysis. *Philosophy of the Social Sciences*, *22*, 461–474.

Plantinga, A. (1974). *The nature of necessity*. Oxford: Oxford University Press.

Potter, J., & Wetherall, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour*. London: Sage.

Putnam, H. (1975). The meaning of 'meaning'. In *Mind, language & reality: Philosophical papers, Vol. II* (pp. 215–271). Cambridge: Cambridge University Press.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructures of cognition. Vol. 1: Foundations*. Cambridge, MA: MIT Press.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 501–518.

Schiffer, S. (1987). *Remnants of meaning*. Cambridge, MA: MIT Press.

Schiffer, S. (1991). Ceteris paribus laws. *Mind*, *100*, 1–17.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Stampe, D. (1977). Toward a causal theory of linguistic representation. In P. French, T. Euling, & H. Wettstein (Eds.), *Midwest studies in philosophy* (Vol. 2). Minneapolis: University of Minnesota Press.

Stich, S. (1983). *From folk psychology to cognitive science*. Cambridge, MA: MIT Press.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

West, L., & Pines, A. (Eds.). (1985). *Cognitive structure and conceptual change*. Orlando, FL: Academic Press.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*. Reading, MA: Addison-Wesley.

Does Science Underwrite Our Folk Psychology?

Barry C. Smith

Folk psychology is part of common sense. It is that part we use to make sense of our own and other people's words and deeds. We rely on it to predict what people will do and say on particular occasions, and to make them intelligible even when what they do and say surprises us.[1] It comprises a complex body of knowledge about the workings of the mind and its effects on behavior, which in the loosest sense may be called a theory. Viewed in this way it is akin to our commonsense understand-