



Philosophical Review

Folk Psychology is Here to Stay

Author(s): Terence Horgan and James Woodward

Source: *The Philosophical Review*, Vol. 94, No. 2 (Apr., 1985), pp. 197-226

Published by: Duke University Press on behalf of Philosophical Review

Stable URL: <http://www.jstor.org/stable/2185428>

Accessed: 21/02/2009 06:52

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=duke>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.



Duke University Press and *Philosophical Review* are collaborating with JSTOR to digitize, preserve and extend access to *The Philosophical Review*.

<http://www.jstor.org>

FOLK PSYCHOLOGY IS HERE TO STAY

Terence Horgan and James Woodward

Folk psychology is a network of principles which constitutes a sort of common-sense theory about how to explain human behavior. These principles provide a central role to certain propositional attitudes, particularly beliefs and desires. The theory asserts, for example, that if someone desires that *p*, and this desire is not overridden by other desires, and he believes that an action of kind *K* will bring it about that *p*, and he believes that such an action is within his power, and he does not believe that some other kind of action is within his power and is a preferable way to bring it about that *p*, then *ceteris paribus*, the desire and the beliefs will cause him to perform an action of kind *K*. The theory is largely functional, in that the states it postulates are characterized primarily in terms of their causal relations to each other, to perception and other environmental stimuli, and to behavior.

Folk psychology (henceforth FP) is deeply ingrained in our common-sense conception of ourselves as persons. Whatever else a person is, he is supposed to be a rational (at least largely rational) *agent*—that is, a creature whose behavior is systematically caused by, and explainable in terms of, his beliefs, desires, and related propositional attitudes. The wholesale rejection of FP, therefore, would entail a drastic revision of our conceptual scheme. This fact seems to us to constitute a good *prima facie* reason for not discarding FP too quickly in the face of apparent difficulties.

Recently, however, FP has come under fire from two quarters. Paul Churchland (1981) has argued that since FP has been with us for at least twenty-five centuries, and thus is not the product of any deliberate and self-conscious attempt to develop a psychological theory which coheres with the account of *homo sapiens* which the natural sciences provide, there is little reason to suppose that FP is true, or that humans undergo beliefs, desires, and the like. And Stephen Stich (1983) has argued that current work in cognitive science suggests that no events or states posited by a mature cognitive psychology will be identifiable as the events and states posited by FP; Stich maintains that if this turns out to be the case,

then it will show that FP is radically false, and that humans simply do not undergo such mental states as beliefs and desires.

In this paper we shall argue that neither Churchland nor Stich has provided convincing reasons for doubting the integrity of FP. Much of our discussion will be devoted to showing that they each employ an implausibly stringent conception of how FP would have to mesh with lower-level theories in order to be compatible with them. We do not deny the possibility that FP will fail to be compatible with more comprehensive theories; this would happen, for instance, if the correct theoretical psychology turned out to be a version of radical Skinnerian behaviorism. But we maintain that there is no good reason to suppose that it will *actually* happen.

Before proceeding, several preliminaries. First, we shall use the rubric 'event' in a broad sense, to include not only token changes, but also token states and token processes. Thus, non-momentary folk-psychological token states will count as mental events, in our terminology.

Second, we shall take FP to consist of two components: a set of *theoretical principles*, and an *existential thesis*. Many or all of the theoretical principles may be expected to have the general form exemplified by the example in our opening paragraph; that is, they are universal closures of conditional formulas.¹ As such they do not carry any existential import, since they might all be vacuously true. The existential thesis of FP, on the other hand, is the assertion that generally our everyday folk-psychological descriptions of people are true, and that humans generally do undergo the folk-psychological events that we commonly attribute to them. We take it that Churchland and Stich are arguing primarily against the existential

¹Actually, we regard the example in the first paragraph as a schema which yields a whole range of instances when various sentences are substituted for the letter 'p' and various sortal predicates are substituted for the dummy phrase 'of kind K'. (The word 'someone', though, functions as a quantificational term; under appropriate regimentation, it would go over into a universal quantifier whose scope is the whole schema.) We prefer to think of predicates of the form "... believes that p" as what Quine (1970) calls *attitudinatives*—i.e., complex one-place predicates constructed by appending a predicate-forming operator ('believes that') to a sentence. On this view, propositional attitudes have no "objects," since they are not relational states. For further discussion see Horgan (forthcoming).

thesis of FP; i.e., they are claiming that our everyday folk-psychological ascriptions are radically false, and that there simply do not exist such things as beliefs, desires, and the rest. Thus their argument, as we understand it, leaves open the possibility that the theoretical principles of FP are true but merely vacuously so.

Third, we are not necessarily claiming that FP is fully correct in every respect, or that there is no room to correct or improve FP on the basis of new developments in cognitive science or neuroscience. Rather, we are claiming that FP's theoretical principles are *by and large* correct, and that everyday folk-psychological ascriptions are often true.

Fourth, we want to dissociate ourselves from one currently influential strategy for insulating FP from potential scientific falsification—viz., the instrumentalism of Daniel Dennett (1978, 1981). He says, of beliefs and desires, that these “putative . . . states” can be relegated “to the role of idealized fictions in an action-predicting, action-explaining calculus” (1978, p. 30). They are not what Reichenbach calls “illata—posited theoretical entities”; instead, he maintains, they are “abstracta—calculation-bound entities or logical constructs” (1981, p. 13), whose status is analogous to components in a parallelogram of forces (1981, p. 20). In short, he evidently holds that they are instrumentalistic fictions, and hence that they are compatible with virtually anything we might discover in cognitive science or neuroscience. We reject Dennett's instrumentalism. We maintain that FP, in addition to providing a useful framework for prediction, also provides genuine *causal explanations*. Although an instrumentalistic attitude toward the intentional idioms of FP is compatible with the mere predictive use of these idioms, it simply is not compatible with their explanatory use, or with talk of beliefs and desires as causes. Accordingly, FP requires a defense more vigorous than Dennett's instrumentalism.

I

Churchland's (1981) argument against the compatibility of FP and neuroscience rests on three considerations. First, “FP suffers explanatory failures on an epic scale” (p. 76). Second, “it has been stagnant for at least twenty-five centuries” (p. 76). And third, “its

intentional categories stand magnificently alone, without any visible prospect of reduction" to neuroscience (p. 75). Irreducibility is the main consideration, and it is allegedly reinforced by the other two points: "A successful reduction cannot be ruled out, in my view, but FP's explanatory impotence and long stagnation inspire little faith that its categories will find themselves neatly reflected in the framework of neuroscience" (p. 75).

Let us consider each of Churchland's three points in turn. In elaboration of the first point, he writes:

As examples of central and important mental phenomena that remain largely or wholly mysterious within the framework of FP, consider the nature and dynamics of mental illness, the faculty of creative imagination . . . the nature and psychological functions of sleep. . . the common ability to catch an outfield fly ball on the run. . . the internal construction of a 3-D visual image. . . the rich variety of perceptual illusions. . . the miracle of memory. . . the nature of the learning process itself . . . (p. 73).

There are at least two important respects in which this passage is misleading. First, while FP itself may have little to say about the matters Churchland mentions, theories based on concepts deriving from FP have a good deal to say about them. For example, cognitive psychologists have developed extensive and detailed theories about visual perception, memory, and learning that employ concepts recognizably like the folk-psychological concepts of belief, desire, judgment, etc.² The versions of attribution theory and cognitive dissonance theory considered below in connection with Stich are important cases of theories of this kind. That all such theories are unexplanatory is most implausible, and in any case requires detailed empirical argument of a sort Churchland does not provide.

Secondly, Churchland's argument seems to impose the *a priori* demand that any successful psychological theory account for a certain pre-established range of phenomena, and do so in a unified way. Arguments of this general type deserve to be treated with skepticism and caution. The history of science is full of examples in which our pre-theoretical expectations about which phenomena it is reasonable to expect a theory to account for or group together

²For visual perception, see, e.g. Gregory (1970).

have turned out to be quite misleading. For example, the demand was frequently imposed on early optical theories that they account for facts which we would now recognize as having to do with the physiology or psychology of vision; this had a deleterious effect on early optical theorizing. Similar examples can readily be found in the history of chemistry.³

The general point is that reasonable judgments about which phenomena a theory of some general type should be expected to account for require considerable theoretical knowledge; when our theoretical knowledge is relatively primitive, as it is with regard to many psychological phenomena, such judgments can go seriously astray. There is no good reason, *a priori*, to expect that a theory like FP, designed primarily to explain common human actions in terms of beliefs, desires, and the like, should also account for phenomena having to do with visual perception, sleep, or complicated muscular coordination. The truth about the latter phenomena may simply be very different from the truth about the former.

What about Churchland's second argument, viz., that FP has remained stagnant for centuries? To begin with, it seems to us at least arguable that FP has indeed changed in significant and empirically progressive ways over the centuries, rather than stagnating. For example, it is a plausible conjecture that Europeans in the 18th or 19th centuries were much more likely to explain human behavior in terms of character types with enduring personality traits than 20th century Europeans, who often appeal instead to "situational" factors. (Certainly this difference is dramatically evident in 18th and 20th century literature; contrast, say, Jane Austen and John Barth.)⁴ Another example of empirically progressive change, perhaps, is the greater willingness, in contemporary culture, to appeal to unconscious beliefs and motivations.

³For example, eighteenth century chemical theories attempted to explain such properties of metals as their shininess and ductility by appeal to the same factors which were also thought to explain the compound-forming behavior of metals. Chemical theories such as Lavoisier's focused just on compounds, and originally were criticized for their failure to provide also a unified explanation of metallic shininess and ductility.

⁴For some striking evidence that situational theories are more empirically adequate, and hence that this change has been a progressive one, see Nisbett and Ross (1980).

Another reason to question the “empirical unprogressiveness” argument is that cognitive psychological theories employing belief-like and desire-like events have led to a number of novel and surprising predictions, which have borne out by experiment. (We discuss some pertinent examples below. For other striking cases the reader is referred to Nisbett and Ross (1980).) Yet Churchland seems to argue as though the (alleged) empirical unprogressiveness of FP is a good reason for taking any theory modelled on FP to be false.⁵ This is rather like arguing that any sophisticated physical theory employing central forces must be false on the grounds that the ordinary person’s notions of pushing and pulling have been empirically unprogressive.

Furthermore, the standard of “empirical progressiveness” is not very useful in assessing a theory like FP anyway. The typical user of FP is interested in applying a pre-existing theory to make particular causal judgments about particular instances of human behavior, not in formulating new causal generalizations. He is a consumer of causal generalizations, not an inventor of them. In this respect he resembles the historian, the detective, or the person who makes ordinary singular causal judgments about inanimate objects. It is not appropriate, we submit, to assess these activities using a standard explicitly designed to assess theories that aim at formulating novel causal generalizations.

This point emerges clearly when one realizes that much of the implicit theory behind many ordinary (but non-psychological) particular causal judgments has presumably changed very slowly, if at all, over the past thousand years. Both we and our ancestors judge that the impact of the rock caused the shattering of the pot, that the lack of water caused the camel to die, that a very sharp blow on the head caused A’s death, that heat causes water to boil, etc. None of these judgments are part of a (swiftly) empirically progressive theory, yet it seems ludicrous to conclude (on those grounds alone) that they are probably false. A similar point can be made about much (although by no means all) of the implicit causal theory employed

⁵Thus his critical remarks on Fodor (1975), and in general on cognitive psychological theories that take information to be stored in sentential form; cf. Churchland (1981, pp. 78 ff.).

by historians. These examples serve to remind us that not all folk theorizing is now regarded as radically false.

This brings us to Churchland's third, and most fundamental, argument for the alleged incommensurability of FP with neuroscience: viz., the likely irreducibility of the former to the latter. An ideal intertheoretic reduction, as he describes it, has two main features:

First, it provides us with a set of rules—"correspondence rules" or "bridge laws," as the standard vernacular has it—which effect a mapping of the terms of the old theory (T_o) onto a subset of the expressions of the new or reducing theory (T_n). These rules guide the application of those selected expressions of T_n in the following way: we are free to make singular applications of those expressions in all those cases where we normally make singular applications of their correspondence-rule doppelgangers in T_o

Second, and equally important, a successful reduction ideally has the outcome that, under the term mapping effected by the correspondence rules, the central principles of T_o (those of semantic and systematic importance) are mapped onto general sentences of T_n that are *theorems* of T_n (1979, p. 81).

We certainly agree that an ideal, or approximately ideal, reduction of FP to natural science would be *one* way of salvaging FP. And we also agree that such a reduction—indeed, even a species-specific reduction—is an unlikely prospect, given that FP is at least twenty-five centuries old and hence obviously was not formulated with an eye toward smooth term-by-term absorption into 20th century science. (A non-species-specific reduction is even less likely, because if FP is true of humans then it can equally well be true of Martians whose physico-chemical composition is vastly different from our own—so different that there are no theoretically interesting physical descriptions that can subsume both the physico-chemical properties which "realize" FP in humans and the corresponding physico-chemical properties in Martians.)

But even if FP cannot be reduced to lower-level theories, and even if lower-level theories can themselves provide a marvelous account of the nature and behavior of *homo sapiens*, it simply does not follow that FP is radically false, or that humans do not undergo the intentional events it posits. Churchland's eliminative mate-

rialism is not the only viable naturalistic alternative to reductive materialism. Another important alternative is the non-reductive, non-eliminative materialism of Donald Davidson (1970, 1973, 1974).

Davidson advocates a thesis which asserts that every concrete mental event is identical to some concrete neurological event, but which does not assert (indeed, denies) that there are systematic bridge laws linking mental event-*types*, or properties, with neurological event-*types*. He calls this view *anomalous monism*; it is a form of monism because it posits psychophysical identities, and it is "anomalous" because it rejects reductive bridge laws (or reductive type-type identities).⁶

The availability of anomalous monism as an alternative to reductive materialism makes it clear that even if FP is not reducible to neuroscience, nevertheless the token mental events posited by FP might well exist, and might well bear all the causal relations to each other, to sensation, and to behavior which FP says they do.

Churchland never mentions Davidson's version of the identity theory—a very odd fact, given its enormous influence and its obvious relevance to his argument. Instead he argues directly from the premise that FP probably is not reducible to neuroscience to the conclusion that FP probably is false. So his argument is fallacious, in light of token-token identity theory as an alternative possible account of the relation between FP and neuroscience. He is just mistaken to assume that FP must be reducible to neuroscience in order to be compatible with it.

II

Let us now consider Stich's reasons for claiming that FP probably will not prove compatible with a developed cognitive science

⁶In order to elevate anomalous monism into a full-fledged version of materialism, one must add to it an account of the metaphysical status of mental state-types (properties) *vis à vis* physico-chemical state-types. The appropriate doctrine, we think, is one also propounded by Davidson (1970, 1974): viz., that mental properties are *supervenient* upon physical ones. Several philosophers recently have developed this idea, arguing that materialism should incorporate some sort of supervenience thesis. Cf. Kim (1978, 1982); Haugeland (1982); Horgan (1981b, 1982b); and Lewis (1983). Also see the papers collected in the Spindel issue of *The Southern Journal of Philosophy*, 22, 1984.