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. 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).
FOLK PSYCHOLOGY IS HERE TO STAY

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. 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.
FOLK PSYCHOLOGY IS HERE TO STAY

(henceforth CS). Unlike Churchland, Stich does not assume that FP must be reducible to more comprehensive lower-level theories in order to be compatible with them. We shall say more presently about the way he thinks FP must fit with these theories.

Stich offers two arguments against the compatibility of FP and CS; we shall examine these in this section and the next. The first argument purports to show that the overall causal organization of the cognitive system probably does not conform with the causal organization which FP ascribes to it. The argument runs as follows.

Events which satisfy a given sortal predicate of the form "... is a belief that p" are supposed to have typical behavioral effects of both verbal and non-verbal kinds. On the verbal side, the events in this class are ones which typically cause the subject, under appropriate elicitation conditions, to utter an assertion that p. On the non-verbal side, these events are ones which, in combination with a subject's other beliefs, desires, and the like, typically cause the subject to perform those actions which FP says are appropriate to the combination of that belief with those other propositional attitudes. But recent experimental evidence suggests, according to Stich, that the psychological events which control non-verbal behavior are essentially independent of those which control verbal behavior—and hence that the cognitive system simply does not contain events which, taken singly, occupy the causal role which FP assigns to beliefs. If these experimental results prove generalizable, and if CS subsequently develops in the direction of positing separate, largely independent, cognitive subsystems for the control of verbal and non-verbal behavior respectively, then we will be forced to conclude, argues Stich, that there are no such things as beliefs.

One of his central empirical examples is a study in attribution theory, performed by Storms and Nisbett (1970). He describes its first phase this way:

Storms and Nisbett... asked insomniac subjects to record the time they went to bed and the time they finally fell asleep. After several days of record keeping, one group of subjects (the "arousal" group) was given a placebo pill to take fifteen minutes before going to bed. They were told that the pill would produce rapid heart rate, breathing irregularities, bodily warmth and alertness, which are just the typical symptoms of insomnia. A second group of subjects (the "relaxation" group) was told that the pills would produce the opposite symptoms:
lowered heart rate, breathing rate, body temperature and alertness. Attribution theory predicts that the arousal group subjects would get to sleep faster on the nights they took the pills, because they would attribute their symptoms to the pills rather than to the emotionally laden thoughts that were running through their minds. It also predicts that subjects in the relaxation group will take longer to get to sleep. Since their symptoms persist despite having taken a pill intended to relieve the symptoms, they will infer that their emotionally laden thoughts must be particularly disturbing to them. And this belief will upset them further, making it all that much harder to get to sleep. Remarkably enough, both of these predictions were borne out. Arousal group subjects got to sleep 28 percent faster on the nights they took the pill, while relaxation subjects took 42 percent longer to get to sleep (Stich, 1983, p. 232).

What Stich finds particularly significant is the second phase of this study. After the completion of the initial insomnia experiments, the members of the arousal group were informed that they had gotten to sleep more quickly after taking the pill, and the members of the relaxation group were informed that they had taken longer to fall asleep. They were asked why this happened, and Nisbett and Wilson report the following pattern of responses:

Arousal subjects typically replied that they usually found it easier to get to sleep later in the week, or that they had taken an exam that had worried them but had done well on it and could now relax, or that problems with a roommate or girlfriend seemed on their way to resolution. Relaxation subjects were able to find similar sorts of reasons to explain their increased sleeplessness. When subjects were asked if they had thought about the pills at all before getting to sleep, they almost uniformly insisted that after taking the pills they had completely forgotten about them. When asked if it had occurred to them that the pill might be producing (or counteracting) the arousal symptoms, they reiterated their insistence that they had not thought about the pills at all after taking them. Finally, the experimental hypothesis and the postulated attribution process were described in detail. Subjects showed no recognition of the hypothesized process and . . . made little pretense of believing that any of the subjects could have gone through such processes (Nisbett and Wilson, 1977, p. 238).

It is very likely, given the data from the first phase of the study, that the cognitive mechanisms which controlled the subjects’ verbal responses in the second phase were largely distinct from the cog-
nitive mechanisms which influenced their actual sleep patterns. And in numerous other studies in the literature of attribution theory and cognitive dissonance theory, the data support a similar conclusion: the mechanisms which control an initial piece of non-verbal behavior are largely distinct from the mechanisms which control the subject’s subsequent verbal accounts of the reasons for that behavior.7

Stich, if we understand his argument correctly, draws three further conclusions. (1) In cases of the sort described, there is no cogent and consistent way to ascribe beliefs and desires; for FP typically attributes both verbal and non-verbal behavioral effects to particular beliefs and desires, but in these cases the cognitive causes of the non-verbal behavior are distinct from the cognitive causes of the verbal behavior, and hence neither kind of cause can comfortably be identified with a belief or desire. (2) It is likely that in general our verbal behavior is controlled by cognitive mechanisms different from those that control our non-verbal behavior; for the Storms-Nisbett pattern emerges in a broad range of studies in attribution theory and dissonance theory. From (1) and (2) he concludes: (3) It is likely that FP is radically false, that is, that humans do not undergo beliefs and desires.

We do not dispute the contention that in a surprising number of cases, as revealed by studies in attribution theory and dissonance theory, the mental states and processes which cause an initial item of non-verbal behavior are distinct from the states and processes which cause a subject’s subsequent remarks about the etiology of that behavior. But we deny that either (1) or (2) is warranted by this contention. And without (1) or (2), of course the argument for (3) collapses.

Consider (1). Is there really a problem in consistently ascribing beliefs, desires, and other folk-psychological states in light of the phenomena described in the Storms-Nisbett study, for instance? No. For we can appeal to unconscious beliefs, desires, and inferences. Although FP asserts that beliefs and desires normally give rise to their own verbal expression under appropriate elicitation

7For surveys of the relevant literature, see Nisbett and Wilson (1977), and Wilson (forthcoming).
conditions, it does not assert this about unconscious beliefs and desires. On the contrary, part of what it means to say that a mental event is unconscious is that it lacks the usual sorts of direct causal influence over verbal behavior. Thus we have available the following natural and plausible folk-psychological account of the subjects' behavior in the Storms-Nisbett study: their initial non-verbal behavior was caused by unconscious beliefs and inferences, whereas their subsequent verbal behavior was caused by distinct, conscious, beliefs about the likely causes of their initial nonverbal behavior. In short, FP does not break down in such cases, because one has the option—the natural and plausible option—of positing unconscious folk-psychological causes.

There is a temptation, we realize, to identify FP with "what common sense would say," and to take the fact that the Storms/Nisbett results confute our common-sense expectations as automatically falsifying some component of FP. But this temptation should be resisted. Common sense would not postulate the relevant unconscious beliefs and desires. But once we do postulate them, perhaps on the basis of rather subtle-non-verbal behavioral evidence, FP seems to yield the correct predictions about how the subjects will perform in Storms and Nisbett's study.

Indeed, as we understand the views of psychologists like Storms, Nisbett, and Wilson who cite such studies as evidence that verbal and non-verbal behavior often are under separate cognitive controls, this appeal to unconscious folk-psychological causes is precisely the theoretical move they are making concerning such cases. Attribution theory and cognitive dissonance theory give center stage to folk-psychological notions like desire and belief. Accordingly, the dual control thesis is nothing other than the folk-psychological thesis just stated: it is the claim that unconscious beliefs and inferences cause the subjects' initial non-verbal behavior, whereas distinct conscious beliefs (which constitute hypotheses about the causes of their original behavior) cause their subsequent verbal behavior. Notice how Stich himself, in the above-quoted passage, describes the first phase of the Storms-Nisbett study. "Attribution

---

8At any rate, this is what the dual-control thesis amounts to as regards the Storms/Nisbett study. Other kinds of mental events besides beliefs and inferences might sometimes be involved too.
theory," he says, "predicts that subjects in the relaxation group will infer that their emotionally laden thoughts must be particularly disturbing to them. And this belief will upset them further . . ." (emphasis ours). Now Stich may have in mind a way of reinterpretation these claims so that the notions of belief and inference they employ are very different from the FP-notions, but in the absence of such a reinterpretation, his contention that beliefs and belief-generating mechanisms cannot be cogently ascribed to subjects like those of Storms and Nisbett is quite unfounded.

Our construal of the dual-control thesis assumes, of course, that it makes sense to speak of beliefs and other mental events as unconscious. But Storms, Nisbett, and Wilson claim quite explicitly that there can be non-verbal behavioral criteria which warrant the ascription of beliefs and other mental events even when a subject's verbal behavior appears inconsistent with the existence of such events.9

It may well be that the appeal to these criteria—and to unconscious beliefs and inferences generally—constitutes an extension and partial modification of traditional FP; but even if it does, this is hardly a wholesale rejection of folk-psychological notions. On the contrary, the very naturalness of the appeal to unconscious folk-psychological causes reflects the fact that the overall causal architecture posited by FP remains largely intact even when we introduce the conscious/unconscious distinction.

So conclusion (1) should be rejected. This means that even if (2) were accepted, FP would not necessarily be undermined. But conclusion (2) should be rejected in any case. From the fact that unconscious mental mechanisms control our non-verbal behavior in a surprising number of cases, one may not reasonably infer that in general our verbal and non-verbal behavior are under separate cognitive control. The findings of attribution theory and dissonance theory, although they do caution us against excessive confidence in our ability to know ourselves, fall far short of establishing such a sweeping conclusion. In this connection it is useful to examine the remarks of Timothy Wilson (forthcoming), a leading advocate of the idea of "dual cognitive control" over verbal and nonverbal behavior respectively. Stich makes much of Wilson's position,

9See, for instance, Wilson (unpublished), pp. 7 ff.
which he construes as the radical thesis that our own statements concerning the mental events that cause our nonverbal behavior are virtually never caused by those mental events themselves. But this is a mistaken interpretation, in our judgment. Wilson articulates his proposal this way:

In essence the argument is that there are two mental systems: One which mediates behavior (especially unregulated behavior), is largely nonconscious, and is perhaps, the older of the two systems in evolutionary terms. The other, perhaps newer system, is largely conscious, and its function is to attempt to verbalize, explain, and communicate mental states. As argued earlier, people often have direct access to their mental states, and in these cases the verbal system can make direct and accurate reports. When there is limited access, however, the verbal system makes inferences about what these processes and states might be (pp. 18–19).

It seems clear from this passage that Wilson is not suggesting that in general our utterances about our mental events are generated by cognitive events other than those mental events themselves. Rather, he is acknowledging that people often have direct conscious access to the mental causes of their behavior, and that at such times these states typically cause accurate reports about themselves. Only where access is limited, where the events are not conscious, are our subsequent utterances caused by inferences about likely mental causes rather than by the mental events themselves.¹⁰

Wilson goes on to suggest that it will typically be events that are results of considerable processing which will be relatively inaccessible to the agent, and that “more immediate states” (such as pre-cognitive states) may be much more accessible (p. 39). Moreover,

¹⁰Still, one can understand why Stich would be led to attribute the radical dual-control thesis to Wilson, even though Wilson evidently does not actually hold this view. Stich quotes from what evidently was an earlier version of the above-quoted passage, wherein Wilson said that the function of the verbal system “is to attempt to verbalize, explain and communicate what is occurring in the unconscious system.” Admittedly, this earlier wording suggests that in verbalizing our mental states we never have conscious access to those states. But the present passage, with its explicit acknowledgment of frequent conscious access, evidently cancels this suggestion, along with any implicit commitment to the radical dual-control thesis.
there are many cases which do seem to involve complex processing in which people exhibit integrated verbal and non-verbal behavior in a way that seems difficult to understand if the systems controlling verbal and non-verbal behavior are entirely independent. Consider engaging in some complicated task while explaining to someone else what you are doing—as in working logic problems on the blackboard as one lectures. It is hard to see how such an integrated performance is possible if the actor has no access to the beliefs which cause the non-verbal portion of his behavior (other than via after-the-fact inferences).

We conclude, then, that neither conclusion (1) nor conclusion (2) is warranted by the kinds of psychological studies Stich cites, and hence that his “dual-control” argument against FP is not successful.

III

The “dual-control” argument does not presuppose any particular conception of how FP must be related to CS in order for the two theories to be compatible. Stich’s second argument for the incompatibility of FP and CS, however, does rest upon such a conception. In particular, he requires that beliefs, desires, and the like should be identical with “naturally isolable” parts of the cognitive system; he calls this the modularity principle.

Stich does not attempt to make this principle precise, but instead leaves the notion of natural isolability at the intuitive level. Accordingly, we too shall use this notion without explication; we think the points we shall make are applicable under any reasonable construal.

Stich argues that FP probably fails to satisfy the modularity principle vis-à-vis CS, and hence that there probably are no such events as beliefs, desires, and the like. He focuses on recent trends within CS concerning the modeling of human memory. Some early models of memory organization, he points out, postulate a distinct sentence or sentence-like structure for each memory. These models are clearly modular, he says, because the distinct sentence-like structures can be identified with separate beliefs. Another sort of model, motivated largely by the desire to explain how people are able to locate information relevant to a given task at hand, treats memory as a complex network of nodes and labeled links, with the
nodes representing concepts and the links representing various sorts of relations among concepts. Stich regards network models as "still quite far over to the modular end of the spectrum," however, because in a network model it is generally unproblematic to isolate the part of the network which would play the causal role characteristic of a given belief (1983, p. 239).

But in recent years, he points out, several leading theorists have become quite skeptical about highly modular models, largely because such models do not seem capable of handling the enormous amount of non-deductive inference which is involved in language use and comprehension. Citing Minsky (1981) as an example, Stich writes:

In a . . . recent paper Minsky elaborates what he calls a "Society of Mind" view in which the mechanisms of thought are divided into many separate "specialists that communicate only sparsely" (p. 95). On the picture Minsky suggests, none of the distinct units or parts of the mental model "have meanings in themselves" (p. 100) and thus none can be identified with individual beliefs, desires, etc. Modularity—I borrow the term from Minsky—is violated in a radical way since meaning or content emerges only from "great webs of structure" (p. 100) and no natural part of the system can be correlated with "explicit" or verbally expressible beliefs (1983, p. 241).

If Minsky's "Society of Mind" view is the direction that CS will take in the future, then presumably modularity will indeed be violated in a radical way.

We are quite prepared to acknowledge that CS may well become dramatically non-modular, and hence that the modularity principle may well end up being refuted empirically. Indeed, if one considers the relation between FP and neuroscience—or even the relation between CS and neuroscience, for that matter—one would expect modularity to be violated in an even more dramatic way.

11 Although we think it quite possible that CS will become non-modular at its most fundamental levels, we also believe that certain higher-level branches of theoretical psychology probably not only will remain modular, but will continue to employ the concepts of FP itself. Attribution theory is a case in point. (By a "higher-level" psychological theory we mean one which posits events that are wholes whose parts are the events posited by "lower-level" psychological theories. More on this below.)
There are tens of billions of neurons in the human central nervous system, and thousands of billions of synaptic junctures; so if the "naturally isolable" events of neuroscience are events like neuron-firings and inter-synaptic transfers of electrical energy, then it is entirely likely that the naturally-isolable events of both FP and CS will involve "great webs of structure" neurally—that is, great conglomerations of naturally-isolable neural events.

So if modularity is really needed in order for FP-events to exist and to enter into causal relations, then the failure of modularity would indeed spell big trouble for the proffered compatibility of FP with lower-level theories. In fact, it also would spell big trouble for the proffered compatibility of cognitive science with lower-level theories like neuroscience; thus Stich's style of argument appears to prove more than he, as an advocate of CS, would like it to prove! And indeed, the demand for modularity even spells big trouble for the compatibility of neuroscience with physics-chemistry; for, if the natural-kind predicates of physics-chemistry are predicates like "... is an electron" and "... is a hydrogen atom," then it is most unlikely that entities falling under neuroscientific natural-kind terms like "... is a neuron" will also fall under physico-chemical natural kind terms. Rather, neurons and neuron-firings are entities which, from the physico-chemical point of view, involve "great webs of structure."

We point out these generalizations of Stich's argument because we think they make clear the enormous implausibility of the modularity principle as an inter-theoretic compatibility condition. Surely objects like neurons, or events like neuron-firings, don't have to be "naturally isolable" from the perspective of fundamental physics-chemistry in order to be compatible with it; rather, it is enough that they be fully decomposable into naturally-isolable parts. Similarly, cognitive-psychological events don't have to be naturally isolable from the perspective of neuroscience in order to be compatible with it; again, it is enough that these events are decomposable into naturally-isolable parts.12

12It is worth noting another respect in which Stich's (and Churchland's) arguments seem to lead to sweeping and implausibly strong conclusions. Much formal theory in the social sciences involves ascribing to individual actors states which are recognizably like, or recognizably descended from,
The situation is exactly the same, we submit, for folk-psychological events in relation to the events of CS. Perhaps Minsky is right, and the role of a belief (say) is typically played by a vast, highly gerrymandered, conglomeration of CS-events. This doesn’t show that the belief doesn’t exist. On the contrary, all it shows is that the belief is an enormously complex event, consisting of numerous CS-events as parts. After all, we expect those CS-events, in turn, to consist of numerous neurological events as parts; and we expect those neurological events, in their turn, to consist of numerous physico-chemical events as parts.

Stich never attempts to justify the modularity principle as a compatibility condition, just as Churchland never attempts to justify the demand for reducibility. Thus Stich’s modularity argument suffers the same defect as Churchland’s reducibility argument: viz., it rests upon an unsubstantiated, and implausibly strong, conception of how FP must mesh with more comprehensive lower-level theories in order to be compatible with them. (It is important to note, incidentally, that even though Stich does not demand reducibility, still

the FP notions of belief and desire. Within economic and game theory, for example, individual actors are thought of as having indifference curves, utility schedules, or preference orderings over various possible outcomes, and beliefs about the subjective probabilities of these outcomes. Within economic theories of voting or political party behavior, similar assumptions are made. Even among theorists of voting behavior who reject the “economic” approach, typically there are appeals to voters’ beliefs and attitudes to explain behavior. (See, for example, Campbell et al. (1960).) Clearly, if Stich’s modularity requirement and Churchland’s smoothness of reduction requirement are not satisfied by the FP notions of belief and desire, then they are unlikely to be satisfied by the notions of utility, degree of belief, and so forth employed by such theories. Thus Stich and Churchland seem to have produced general arguments which, if cogent, would show—quite independently of any detailed empirical investigation of the actual behavior of markets, voters, etc.—that all these theories must be false, at least on their most natural interpretation.

13A complex event of the relevant kind might be a mereological sum, or fusion, of simpler events; alternatively, it might be an entity distinct from this event-fusion. We shall take no stand on this matter here. (The issue is closely related to the question whether an entity like a ship is identical with the fusion of its physical parts, or is instead an entity distinct from this fusion, with different intra-world and trans-world identity conditions.) To our knowledge, the most explicit and well-developed theory of parts and wholes for events is that of Thomson (1977); event-fusions are the only kinds of complex events she explicitly countenances.
in a certain way his notion of inter-theoretic fit is actually stronger than Churchland's notion. For, even a reductionist need not require that entities falling under higher-level natural-kind sortals should be naturally isolable from the lower-level perspective. A reductionist does require that there should be biconditional bridge laws correlating the higher-level sortals with open sentences of the lower-level theory, but these lower-level open sentences can be quite complex, rather than being (say) simple natural-kind sortal predicates.

Although Stich offers no explicit rationale for the modularity principle, perhaps he is influenced by the following line of thought:

The propositional attitudes of FP involve a relation between a cognizer and a sentence-like “internal representation” (Fodor 1975, 1978; Field 1978; Lycan 1982). If FP is true, then part of the task of CS is to explain the nature of these internal representations. But CS cannot do this unless internal representations fall under its natural-kind predicates, or at any rate are somehow “naturally isolable” within the cognitive system. And if Minsky’s “Society of Mind” approach is the direction CS will take in the future, then this requirement will not be met. Hence if the events of FP do not obey the modularity principle vis-à-vis CS, then FP must be radically false.

One reason we have for rejecting this line of reasoning is that we doubt whether propositional attitudes really involve internal representations—or whether they have “objects” at all. (Cf. footnote 1 above.) Furthermore, if Minsky’s approach did become the general trend in CS, then presumably this fact too would tend to undermine the claim that sentence-like representations are involved in the propositional attitudes—just as his approach already tends to undermine the claim that such representations are involved in the non-deductive inference that underlies the use and comprehension of language.

Moreover, even if the internal-representation view is correct, and even if part of the task of CS is to give an account of these representations, approaches like Minsky’s would not necessarily render CS incapable of accomplishing this task. For it might turn out that the “atoms” of CS are the components of Minsky’s Society of Minds, and that CS also posits complex, sentence-like “mole-
"molecules" constructed from these "atoms." The molecules might be very complex, and highly gerrymandered. If so, then they won't count as naturally isolable components of the cognitive system when that system is viewed from the atomic perspective; however, they will count as naturally isolable from the higher, molecular perspective. (We think it more likely, however, that if the "Society of Minds" approach proves generalizable within CS, then the result will be a widespread rejection of the mental-representation view of propositional attitudes—a view which, as we said, we think is mistaken anyway.)

Another way one might try to defend the modularity principle is by appeal to Davidsonian considerations involving the role of laws in causality. One might argue (i) that FP contains no strict laws, but only so-called "heteronomic" generalizations (Davidson 1970, 1974); (ii) that two events are related as cause and effect only if they have descriptions which instantiate a strict law (Davidson 1967); and (iii) that event-descriptions which instantiate a strict law of a given theory must pick out events that are naturally isolable from the perspective of that theory. From these three claims, plus the assumption that folk-psychological events enter into causal relations, the modularity principle seems to follow.14

But suppose an event c causes an event e, where c and e both are naturally isolable from the perspective of FP. Suppose that c is fully decomposable into events which respectively satisfy the sortal predicates $F_1 \ldots F_m$ of an underlying homonomic theory T, and hence that these component-events all are naturally isolable from the perspective of T; suppose also that these events jointly satisfy a (possibly quite complex) description $D_1$ of T which specifies their structural interconnection. Likewise, suppose that e is fully decomposable into events which respectively satisfy the sortal predicates $G_1 \ldots G_n$ of T, and hence that these component events all are naturally isolable from the perspective of T; suppose also that these component events jointly satisfy a description $D_2$ of T which specifies their structural interconnection. Now even if c and e do not have natural-kind descriptions under which they themselves instantiate a strict law of T, nevertheless the strict laws of T might jointly entail an assertion of the following form:

14This Davidsonian argument was suggested to us by Stich himself, in conversation.
For any event \( x \), if \( x \) is fully decomposable into events \( x_1 \ldots x_m \) such that \( D_1(x_1 \ldots x_m) \) and \( F_1(x_1), F_2(x_2), \ldots, \text{and } F_m(x_m) \), then \( x \) will be followed by an event \( y \) that is fully decomposable into events \( y_1 \ldots y_n \) such that \( D_2(y_1 \ldots y_n) \) and \( G_1(y_1), G_2(y_2), \ldots, \text{and } G_n(y_n) \).

We see no reason why the causal relation between \( c \) and \( e \) cannot rest upon a regularity of this form. One either can call such regularities strict laws, in which case claim (iii) above will be false; or else one can reserve the term 'strict law' for the relatively simple postulates of a homonomic theory, rather than the set of logical consequences of those postulates—in which case claim (ii) above will be false. Either way, the Davidson-inspired argument for the modularity principle has a false premise. (Incidentally, we do not mean to attribute the argument to Davidson himself, since we doubt whether he would accept claim (iii).)

IV

We have been arguing that FP-events might well be identical with arbitrarily complex, highly gerrymandered, CS-events which themselves are not naturally-isolable relative to CS, but instead are fully decomposable into parts which have this feature. Of course, if FP-events really do exist, then they will have to accord with the causal architecture of FP; that is, they will have to be causally related to each other, to sensation, and to behavior in the ways that FP says they are. Indeed, as functionalists in philosophy of mind have so often stressed, the causal or functional principles of FP are crucial to the very individuation of FP-events; what makes a given event count (say) as a token belief-that-p is, to a considerable extent, the fact that it occupies the causal role which FP assigns to tokens of that belief-type.\(^{15}\)

So if our non-modular picture of the relation between FP and CS is to be plausible, it is essential that complex, gerrymandered events

\(^{15}\)But as the famous case of Twin Earth (Putnam, 1975) seems to show, an event's causal role is not the only factor relevant to its folk-psychological individuation. Our doppelgangers on Twin Earth don't undergo tokens of the type believing that water is good to drink, even though they do undergo events that are functionally indistinguishable from our own token beliefs that water is good to drink. The trouble is that the stuff they call "water" isn't water at all. Cf. Burge (1979).
can properly be considered causes, even if they involve "great webs of structure" relative to lower-level theory. While a detailed discussion must be beyond the scope of this paper, a brief consideration of the causal status of complex events will help to clarify our argument.

Let us say that an event $e$ minimally causes an event $f$ just in case $e$ causes $f$ and no proper part of $e$ causes $f$. We want to advance two claims about minimal causation, each of which will receive some support below. First, even if an event $e$ is a genuine cause of an event $f$, nevertheless $f$ also might be caused by some event which is a proper part of $e$; thus $e$ might be a genuine cause of $f$ without being a minimal cause of $f$. Second, if $e$ causes not only $f$ but also some other event $g$, then it might be that the part of $e$ which minimally causes $f$ is different from the part of $e$ which minimally causes $g$.\(^{16}\)

These two facts are important because they make it relatively

\(^{16}\)While a full defense of these claims must be beyond the scope of this paper, we think they are required for the truth of many causal statements in contexts where highly developed and precise formal theories are not available. Consider the claims (a) that application of a certain fertilizer causes plants to increase in mean height, and also causes them to increase in leaf width; (b) that following a certain study routine $R$ causes an increase in SAT verbal scores, and also causes an increase in SAT mathematical scores; or (c) that certain child-rearing practices cause an increase in the incidence of juvenile delinquency in certain populations. There is an enormous literature detailing complex and ingenious statistical techniques for testing such claims. (Fischer (1935) is an early classic, inspired largely by problems connected with testing claims like (a); and many books on "causal modeling," like Blalock (1971), discuss procedures that are relevant to (b) and (c).) These techniques might well establish that the three claims are true. Yet the cases described in (a), (b), and (c) can easily fail to be minimal causes: the fertilizer will commonly be a mixture, containing compounds which are inert, or which have other effects on the plant besides those mentioned in (a); and it seems implausible to suppose that every feature or detail of study routine $R$ or child rearing practice $C$ is causally necessary for the above effects. (Typically, we have no practical way of determining what the minimal causes in such cases are.) Thus (a) can be true even though the fertilizer is a mixture of several distinct compounds, one of which causes increase in height (but not increase in leaf width) while the other causes increase in leaf width (but not height). Similarly, (b) can be true even though different aspects of study routine $R$ are responsible for the increases in math and in verbal scores. (See Thomson (1977) for further defense of the claim that genuine causes don't have to be minimal causes.)
FOLK PSYCHOLOGY IS HERE TO STAY

easy for events to exist which satisfy the causal principles of FP. If FP attributes both the event f and the event g to a single cause e at time t, and in fact there are distinct (though perhaps partially overlapping) events e₁ and e₂ such that e₁ minimally causes f (at t) and e₂ minimally causes g (at t), this does not necessarily falsify FP. For, e might well have both e₁ and e₂ as parts; indeed, it might well have as parts all those events which minimally cause (at t) one or another of the various events which FP says are effects (at t) of e. As long as this complex event is itself the effect of whatever prior events FP says are e’s causes, the event will be (identical with) e.

The upshot is that FP could very easily turn out to be true, even if modularity is dramatically violated. Not only can FP-events be complex and highly gerrymandered, with numerous naturally-isolable CS-events as parts, but any given FP-event e can cause its effects in a conglomerative manner, with different effects having different parts of e as their respective minimal causes.¹⁷

V

Perhaps it will be objected that our analysis is too permissive; that unless we adopt Stich’s modularity condition, over and above the requirement that FP-events conform to the causal architecture which FP assigns to them, we impose no non-trivial constraints on

¹⁷The point about conglomerative causation is also relevant to Stich’s dual-control argument against FP. Even if verbal and non-verbal behavior should turn out to have largely separate minimal causes, FP could be true anyway; for, FP-events might be complexes of the minimal causes, and these complex events might be genuine causes (albeit non-minimal causes) of both the verbal and the non-verbal behavior. (We should stress, however, that we are not claiming that if the dual-control thesis is true, then whenever a subject is in some state B of his nonverbal behavioral system and some state V of his verbal system, it will always be possible, consistently with FP, to ascribe to him some single folk-psychological cause of both his verbal and non-verbal behavior. Whether this will be possible depends upon the specific states V and B and upon the behavior they cause. In the Storms-Wilson insomnia experiment, for example, the state B (subjects’ attribution of their symptoms to pills) which causes the arousal group to fall asleep is not merely distinct from the state V which causes their verbal behavior (denial that the above attribution had anything to do with their falling asleep); but in addition, these two states cannot, consistently with the causal principles embodied in FP, be treated as components of some single belief.)
the truth conditions of upper-level causal claims; that is, we allow such claims to come out true regardless of the character of the theory that underlies them. We shall conclude by considering this objection.

It is clear that some underlying theories are inconsistent with the truth of some upper-level causal claims. For example, if the world is anything like the way our current chemistry and physics describe it, then possession by the devil cannot be a cause of any psychological disorders, and loss of phlogiston cannot be a cause of the chemical changes undergone by metals when they oxidize. To consider a case which is closer to home, it seems clear that if we are Skinnerian creatures—that is, creatures whose behavior is fully described and explained by the basic principles of Skinnerian psychology—then folk-psychological claims postulating beliefs, desires, and the like as among the causes of our behavior cannot be true.

The worry under consideration is that our non-modular approach to inter-theoretic compatibility is so liberal that it would allow claims of the above sort to come out true even though they seem clearly inconsistent with underlying theory. We shall argue that this worry is ill-founded.

It will be helpful to distinguish two different conceptions or expectations regarding the epistemic role of a radical failure of fit or integration between an upper-level theory and an underlying theory. On the first conception one thinks of this failure of fit as an important epistemic route to the falsity of the upper-level theory, where that falsity may not be obvious otherwise. The idea is that even if direct evidence at the upper level does not clearly point to the falsity of an upper-level theory (and indeed may even seem to support this theory), nonetheless we can detect the falsity of the upper-level theory by noting its failure to fit in some appropriate way with some underlying theory which we have strong reason to believe is true. Clearly, both Stich and Churchland argue in accordance with this conception.

We find more plausible an importantly different conception of the epistemic significance of failure of fit between an upper-level and a lower-level theory. We do not deny, of course, that lower-level theories can be incompatible with upper-level theories. We do doubt, however, whether it is common or typical that one can know that an upper-level theory is false only by noting its failure to fit
with a true underlying theory. More typically, when an upper-level theory is false there is direct evidence for this fact, independently of the failure of fit. The incompatibility arises not because of a failure of modularity, but rather because there simply are no events—either simple or complex—which have all the features which the upper-level theory attributes to the events it posits. Crudely put, the idea is that while various theories of juvenile delinquency or learning behavior can be inconsistent with neurophysiological theories or with physical theories, the former are likely to be confirmable or disconfirmable by the sorts of evidence available to sociologists and psychologists. It will be rare for a theory to be supported by a very wide range of evidence available to the sociologist or the psychologist and yet turn out to be radically false (because its ontology fails to mesh properly with that of some underlying theory). So our conception suggests a greater epistemological autonomy for upper-level disciplines like psychology than does a conception of inter-theoretic compatibility which incorporates a modularity condition.

We have emphasized this epistemological point because it bears directly on worries about the permissiveness of our non-modular conception. While our approach is by no means trivial in the sense that it allows every upper-level theory to be compatible with every underlying theory, it is permissive and deflationary in that, at least for a wide variety of cases, considerations of fit will not play the sort of independent normative role which they would play under a modularity requirement.

With this in mind, let us return to the examples with which we began this section. Consider first the case of possession by the devil. Like other causally explanatory notions, the notion of possession by the devil is to be understood, in large measure, in terms of the role it plays in a network of causal relations. Possession by the devil causes or may cause various kinds of pathological behavior. Such effects may be diminished or eliminated by the use of appropriate religious ceremonies (e.g., prayers or excorcism). When behavior is due to possession by the devil, there is no reason to suppose that it will be affected by other forms of treatment (drugs, nutritional changes, psychotherapy, etc.). The state of possession is itself the effect of the activities of a being who has many other extraordinary powers.
Now if an event of possession by the devil (call it d) is to be a cause of a certain bit of behavior (e.g., jabbering incoherently), then d must, on our analysis, be identifiable with some event (call it e) describable in terms of the predicates of our underlying theory; and it must be the case that, given this identification, at least most of the other causal generalizations in which d is held to figure, according to the theory of devil-possession, should come out true. (Although our conception of inter-theoretic fit countenances failures of modularity, it does insist that the identifications we make preserve the "causal architecture" of the upper-level theory.) We submit that no matter how large and complex one makes the event e with which one proposes to identify d, and no matter how willing one may be to regard proper parts of e as causally efficacious, there is simply no plausible candidate for e which, given our present physical and chemical theory, will make the network of causal claims associated with possession by the devil come out mainly true. That is, there is simply no event e, however complex, which is linked by law to various forms of behavior associated with possession, which is inefficacious in producing such behavior when exorcism is used, which is shown by law to be produced by an agency having the properties of the devil, and so forth.

This example illustrates the general epistemological claim made above. In effect, we have argued that causal claims about possession by the devil are false not because of sophisticated considerations having to do with modularity (or with "smoothness of reduction"), but because the causal architecture associated with possession by the devil is radically mistaken; nothing stands in the network of causal relations with various other events in the way that possession by the devil is supposed to. We can see this immediately by noting that the falsity of claims attributing causal efficacy to devil-possession is, so to speak, directly discoverable without considerations having to do with chemistry, physics, or biology. If one were to run suitably controlled experiments, then presumably one would quickly discover that exorcism does not affect devil-possession type behavior, that certain other therapies do, and so forth.18

18Of course, it might be that some cases of exorcism appear to be efficacious, but this is only because they involve certain features which are also cited by other, more secular, theories (e.g., reassuring the "possessed"
A similar set of observations seems relevant in connection with the allegation that our approach would permit causal claims about beliefs to be true even if we are Skinnerian creatures. FP asserts that beliefs, desires, and other propositional attitudes are related to one another in many and various ways, over and above their causal relations to sensation and behavior. Skinnerian theory, on the other hand, denies that we need to postulate such richly-interacting internal events in order to explain behavior, and it also denies that such events exist at all. Rather, the Skinnerian claims that the causal chains leading from environmental “stimulus” to behavioral “response” are largely isolated from one another, rather like the various parallel non-interacting communication-channels in a fiber-optics communications line; thus, whatever internal events are involved in any particular stimulus-response pairing will not bear very many significant causal relations to the internal events that are involved in other stimulus-response pairings; that is, the Skinnerian claims that as a matter of empirical fact, the generalizations linking stimuli and behavior are so simple and straightforward that they are incompatible with the existence of internal events which interact in the rich way which folk-psychological events are supposed to interact with one another. So if the Skinnerian is right, then there simply are no internal events, in humans or in other organisms, which bear all the causal relations to sensation, to behavior, and to one another which FP assigns to beliefs, desires, and the like. Thus our non-modular conception of inter-theoretic fit would indeed by violated if humans should turn out to be mere Skinnerian creatures. Accordingly, this conception is not unduly permissive after all.

This example also illustrates the epistemological claims made above. It is satisfaction of the “causal architecture” of FP, by some set of (possibly complex) events in the central nervous system, which is crucial to the truth of FP. Hence if we are Skinnerian creatures, so that the causal architecture assumed by FP is not instantiated in us by any events either simple or complex, then
presumably this fact will show up at the level of a relatively coarse-grained analysis of our molar behavior. Stimulus-response laws that are incompatible with the causal architecture of FP will be discoverable, and will be usable to explain and predict the full range of human behavior. Hence it will not be the case that FP seems to be largely true, according to the best available coarse-grained evidence, and yet turns out to be false merely because of failure to fit properly with some underlying theory.

The upshot, then, is that our approach seems exactly as permissive as it should be, and this fact speaks in its favor; by contrast, a modular conception of inter-theoretic fit seems excessively strict, since it is unmotivated and it denies higher-level theories an adequate degree of epistemological autonomy. So, given (i) the notable failure, to date, of behaviorist-inspired psychology's efforts to unearth stimulus-response laws which are applicable to human behavior generally and which undercut the causal architecture of FP, (ii) the fact that folk-psychological notions seem to lie at the very heart of cognitivist theories like attribution theory and cognitive dissonance theory, and (iii) the fact that FP serves us very well in the everyday explanation and prediction of behavior, it seems very hard to deny that in all probability, folk psychology is here to stay.19,20,21

Memphis State University (T.H.)
California Institute of Technology (J.W.)

19Although we have assumed throughout that folk-psychological events are complex events consisting of lower-level events as their parts, we want to acknowledge that it may be possible to defend the compatibility of FP and CS without this assumption. Jaegwon Kim (1966, 1969, 1973) holds that an event is an entity consisting in the instantiation of a property by an object at a time, and that mental events consist in the instantiation of mental properties by individuals at times. Under this approach, it is unclear whether lower-level events can sensibly be treated as parts of FP-events. Nevertheless, an advocate of Kim's theory of events still might be able to argue that FP-events exist and bear all the causal relations to one another that FP says they do. For he might be able to argue that these events are supervenient upon groups of lower-level events, and that supervenience transmits causal efficacy. Cf. Kim (1979, 1982, 1984).

20Throughout this paper we have assumed, as is usual, that if everyday folk-psychological statements are indeed true, then there really exist folk-psychological mental events—that is, token desires, token beliefs, and so forth. In fact, however, one of us (Horgan) thinks there are good reasons
FOLK PSYCHOLOGY IS HERE TO STAY

References


for denying the existence of events in general; cf. Horgan (1978, 1981a, 1982a). Horgan also thinks that if physico-chemical events exist, then normally there will be numerous classes of physico-chemical events from within someone’s head which jointly meet all the causal conditions which would qualify a given class for identification with the class consisting of that person’s folk-psychological mental events; and he takes this to indicate that even if physico-chemical events exist, and even if garden-variety folk-psychological statements (including statements about mental causation) are often true, nevertheless there really are no such entities as mental events; cf. Horgan and Tye (1985). We believe that the essential points of the present paper can be reformulated in a way which does not require the existence of mental events (even if physico-chemical events are assumed to exist), and also in a way which does not require the existence of any events at all. But our objective here has been the more limited one of defending FP within the framework of the ontology of events which is widely taken for granted in contemporary philosophy of mind.

21We thank Stephen Stich, William Tolhurst, and Michael Tye for helpful comments on an earlier version of this paper.
Wilson, T. (forthcoming). “Strangers to Ourselves: The Origins and Accuracy of Beliefs About One’s Own Mental States.”